Deanonymizing Tor Hidden Service Users Through Bitcoin Transactions Analysis

Husam Al Jawaheri, Mashael Al Sabah, Yazan Boshmaf, Aiman Erbad

PII: S0167-4048(18)30990-8
DOI: https://doi.org/10.1016/j.cose.2019.101684
Reference: COSE 101684

To appear in: *Computers & Security*

Received date: 7 September 2018
Revised date: 29 November 2019
Accepted date: 30 November 2019

Please cite this article as: Husam Al Jawaheri, Mashael Al Sabah, Yazan Boshmaf, Aiman Erbad, Deanonymizing Tor Hidden Service Users Through Bitcoin Transactions Analysis, *Computers & Security* (2019), doi: https://doi.org/10.1016/j.cose.2019.101684

Deanonymizing Tor Hidden Service Users Through Bitcoin Transactions Analysis

Husam Al Jawaheri*
University of Luxembourg

Mashael Al Sabah
Qatar Computing Research Institute, HBKU

Yazan Boshmaf
Qatar Computing Research Institute, HBKU

Aiman Erbad
Qatar University

ABSTRACT

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. By exploiting publicly available information, we show how relying on Bitcoin for payments on Tor hidden services could lead to deanonymization of these services' users. Such linking is possible by finding at least one past transaction in the Blockchain that involves their publicly declared Bitcoin addresses.

To demonstrate the consequences of this deanonymization approach, we carried out a real-world experiment simulating a passive, limited adversary. We crawled 1.5K hidden services and collected 88 unique and active Bitcoin addresses. We then crawled 5B tweets and 1M BitcoinTalk forum pages and collected 4.2K and 41K unique Bitcoin addresses, respectively. Each user address was associated with an online identity along with its public profile information. By analyzing the transactions in the Blockchain, we were able to link 125 unique users to 20 hidden services, including sensitive ones, such as The Pirate Bay and Silk Road. We also analyzed two case studies in detail to demonstrate the implications of the information leakage on users anonymity. In particular, we confirm that Bitcoin addresses should be considered exploitable, as they can be used to deanonymize users retroactively. This is especially important for Tor hidden service users who actively seek and expect privacy and anonymity.

KEYWORDS

Bitcoin, Tor Hidden Services, Privacy, Deanonymization, Attack

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 making online transaction or browsing the web. The most famous of these solutions are the decentralized cryptocurrencies and Tor anonymity network. One of the early examples is the Bitcoin network [28], which provides users with the ability to perform online transactions "pseudonymously". Due to its popularity, more than 100K merchants worldwide accept Bitcoin payments [8]. Tor [11] is the most widely used anonymous communication network with millions of daily active users [30]. 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 [9]. Indeed, for those reasons, hidden services are being operated by whistleblowing websites such as WikiLeaks, search engines such as DuckDuckGo, and online social networks such as Facebook. Hidden services have also become a breeding ground for Dark Web vendors, such as Silk Road [6] and Agora [39], which offer illicit merchandise and services [3, 26].

Vincent and Johan [25] discuss that Tor and Bitcoin represent the main components required to achieve 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 user. Unfortunately, Bitcoin lacks retroactive operational security due to its pseudonymity model [28]. This model is limited because of the linkability of Bitcoin transactions that are stored in the Blockchain and their public availability.

Problem. Using Bitcoin as a payment method is a serious threat to the anonymity of Tor hidden services and their users. Yet, Bitcoin is the most popular choice for these services for accepting donations or selling merchandise [3]. Moore and Rid [26] studied how hidden services are used in practice, and noted that Bitcoin was the dominant choice for accepting payments for these services. Although multiple studies [12, 13, 24] demonstrated that Bitcoin transactions are not as anonymous as previously thought, Bitcoin remains the most popular digital currency on the Dark Web [4], and many users still choose to use it despite its false sense of anonymity. Biryukov et al. [2] showed that even if users use Bitcoin over an anonymity network such as Tor, they are still vulnerable to deanonymization and man-in-the-middle attacks at the network level. While previous studies analyze the vulnerabilities that result from using Bitcoin over Tor [2], mostly at the network level, we provide the first study that focuses on hidden services merely using application-level information, shedding light on an exploitable information leakage resulting from correlating public profiles of online social network users with Bitcoin transactions and onion websites. Such correlation is possible due to the transparency of Bitcoin's blockchain design.

Hidden service users are one class of Bitcoin users whose anonymity is particularly important. Hidden service operators and users are actively seeking to maintain their anonymity. However, those users are under the risk of deanonymization when they reveal their Bitcoin addresses. By studying the transactions associated with these addresses, a significant amount of information can be leaked and used to collect sensitive information about hidden services and their customers, where a user can be linked to a hidden service.

[*] Research was conducted while at Qatar University.

In this paper, we seek to understand the privacy risk associated with using Bitcoin as a payment method by Tor hidden services. That is, we show how using Bitcoin leaks public information that can be exploited to deanonymize Tor hidden service users. In particular, we demonstrate the feasibility of linking a user with @alice social network profile to a Tor hidden service with private.onion address. We note that this research has been carefully revised and approved by our institution's IRB board, and it does not put users at any risk other than what they are currently exposed to (Section 3.2). We also discuss a number of countermeasures to improve user privacy and anonymity (Section 4.4).

Approach. By browsing onion landing pages of various hidden services, we observed that it is possible to extract their payment Bitcoin addresses from static HTML content. Accordingly, we crawled 1.5K hidden service pages and created a dataset of 88 Bitcoin addresses operated by those hidden services, including two ransomware addresses. We also crawled online social networks for public Bitcoin addresses, namely, Twitter and the BitcoinTalk forum.[1] Out of 5B tweets and 1M forum pages, we created two datasets of 4.1K and 41K Bitcoin addresses, respectively. Each address in these user datasets is associated with an online identity and its corresponding public profile information (Section 3.3).

Using a clustering heuristic proposed by Meiklejohn et al. [24], we performed a wallet-closure analysis that allowed us to expand the collected Bitcoin addresses per user. In other words, for each address in the user datasets, we identified other addresses belonging to the same user who owns the address. This analysis approximates a user's wallet, which is the set of addresses that are controlled by the user, and thereby might increase the number of identified links between users and hidden services. One problem with closure analysis is that it can over-approximate the size of a wallet, as a consequence of mixing [22] and CoinJoin [40] services. Therefore, we excluded closures that have common addresses. This ensures that users are not double-counted and reported results are lower-bound estimate, as each remaining closure represents a subset of a wallet whose addresses are controlled by a unique user (Section 3.4).

After wallet-closure analysis, we used the datasets to analyze the transactions in the Blockchain. In particular, we searched for transactions between user and hidden service addresses to identify links between them. This enabled us to associate users, or online identities, with hidden services and access their transaction history. To demonstrate the impact of linking, we described two case studies showing that Bitcoin addresses can be used to deanonymize users retroactively. It is important to highlight that deanonymization depends solely on information leaked from public data sources. Finally, to gain insights into the economic activity of the linked hidden services, we analyzed the corresponding transaction history, focusing on number of transactions, the amount of money being exchanged, and the lifetime of these hidden services (Section 3.5).

Results. With wallet-closure analysis, we were able to expand the datasets from 45.2K Bitcoin addresses to more than 19.1M, with an average of 425 addresses per user. Using transaction analysis, we were able to link 125 unique users to 20 Tor sensitive hidden services, such as WikiLeaks, Silk Road, and The Pirate Bay. The case studies unmasked multiple users of The Pirate Bay hidden service, along with their personally identifiable information (PII), such as name, gender, age, and location.[2] We also found that users from multiple countries and different ages had links with the Silk Road address in our hidden service dataset. One of the users, for example, is a teenager who appears to have many social network accounts showing his real identity.

The economic activity analysis of the linked hidden services shows that WikiLeaks is the highest receiver of payments in terms of volume, with 25.6K transactions. In terms of the amount of incoming payments, however, the Silk Road tops the list with more than 29.6K Bitcoins received on its address. We observed that the money flowing in and out of hidden service addresses is nearly the same. This suggests that these services do not keep their Bitcoins on the address they use for receiving payments, but rather distribute the coins to other addresses instead. Finally, from the last transaction dates on the addresses, we found that only eight out of the 20 linked hidden services are inactive in 2017. This, however, does not mean the inactive services stopped making Bitcoin transactions, as they might have used different addresses that we do not know.

Contributions. This paper shows 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 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 mixed wallets.
(2) The first real-world experiment that shows the feasibility of deanonymizing Tor hidden service users by exploiting information leakage resulting from correlating public data sources, namely, online social networks, the Bitcoin Blockchain, and Tor hidden services.
(3) An economic activity analysis of 20 hidden services that were used by linked users. This includes statistics on their transaction volume, flow of money, and lifetime.

2 BACKGROUND

2.1 Bitcoin

Bitcoin [28] 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. One of the reasons of Bitcoin's popularity is its presumed anonymity. The identity of Bitcoin users is hidden using pseudonyms, which are used as addresses to perform transactions. A Bitcoin address is a 160-bit hash of a public key generated by a digital signature algorithm. The set of public/private keys owned by a user is called a wallet. Private keys are used to sign inputs of transactions as a proof of ownership.

2.1.1 Transactions. 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,

[1] https://bitcointalk.org

[2] In accordance with our IRB board's guidelines, we have removed the PII of these users, as the linked hidden services are considered illegal in their countries.

and chooses Bob's address(es) as a transaction output. To protect the transaction, she signs it using her private key(s), and then broadcasts it to the 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 over the Internet, Bitcoin's client, or explored using Block exploring services. 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 Blockchain [3] include more information, such as the relay IP address and the transaction time-stamp that records the time at which the transaction was made.

2.1.2 Anonymity. While transactions in Bitcoin are presumed to be anonymous, linkability between addresses is possible due the nature of the Blockchain [28]. 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.

Meiklejohn et al. [24] 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 B receives, as an output, the unspent change of a transaction where A 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 exceed 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 for Bitcoin transaction analysis.

It is important to note that wallet-closure techniques are noisy and can result in addresses that belong to multiple users. One reason for this is the use of mixing [22] and CoinJoin [40] 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. In this work, we modify the wallet-closure technique to handle Bitcoin mixing for transaction analysis, as described in Section 3.4.

2.2 Tor Hidden Services

Tor [11] 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. Each hidden service designates a number of Tor relays, known as *Introduction Points*, that receive user connection requests specifying (to the hidden service) the *Rendezvous Point*, a user-designated relay where the user is waiting for the connection. Hidden services have also been subjected to active attacks in the wild [10, 23]. For these reasons, the Tor project is actively working on addressing the security weaknesses of hidden services [29].

The design of hidden services in Tor allows users to serve content while concealing the server IP address and location. Hidden services are critical to support human rights activists and whistle-blowers to safely upload and view sensitive documents (i.e., WikiLeaks). Hidden services also attract various stores selling illegal products (e.g., drugs, weapons, and illegal content). Moreover, to ensure transaction anonymity, Bitcoin has become the most popular choice by Tor hidden services for accepting donations or selling merchandise [3]. Unfortunately, this has contributed to the rise of illegal hidden services, such as Silk Road and Agora, which offer illicit merchandises and services [3, 6, 26, 39]. Hidden services are a thriving market for selling and distributing malware, especially Ransomware. Ransomware is a malware category that limits the access of users to their files by encrypting them [18]. Ransomware requires victims to pay in order to get access to the decryption keys. To remain anonymous, Ransomware requires victims to pay through Bitcoin. Ransomware lockers are known to use hidden services as a place to hide their malicious activities [19].

3 APPROACH AND EXPERIMENT

While the goal of using Bitcoin for Tor hidden services is to provide transaction and browsing anonymity, we show that this usage 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.

3.1 Adversary Model

We assume a passive, limited adversary. The adversary has access to or is capable of collecting Bitcoin addresses of Tor hidden services and their users. This attacker does not need to control network resources, but can extract publicly accessible information from online social networks, the Blockchain, and onion pages. Obtaining Bitcoin addresses of users can be either targeted or non-targeted, depending on the attack scenario. For the earlier scenario, the adversary can use social engineering or exploit contextual metadata.

[3] www.blockchain.com/api

For example, if Trudy knows that Alice booked a ticket on Expedia at a certain time with a certain amount of coins, Trudy can easily deduce Alice's Bitcoin address from the Blockchain. For the latter scenario, the adversary can crawl and parse public data sources for Bitcoin addresses and associated identities on a large scale.

We focus on the second, non-targeted attack scenario and show that an adversary can deanonymize hidden service users by correlating public data from online social networks, the Blockchain, and Tor hidden services.

3.2 Ethical Considerations

The deanonymization presented in this work depends on correlating public Bitcoin addresses of users with the transactions stored in the Blockchain. Many prior studies performed similar analyses based on crawled public Bitcoin addresses [13, 24, 32]. While our study narrows down this analysis to the scope of hidden services and their users, we stress that even the Bitcoin addresses of hidden services were readily available on their onion landing pages. We did not try to obtain Bitcoin addresses of hidden services that require authentication, payment, or exchange of emails. However, a web search engine, such as Google, or any other organization that has access to a larger amount of data could perform the analysis on a larger scale, and potentially exploit a significantly larger amount of leaked information about users.

We believe the data collected and used herein is easily available to adversaries. In this research, in addition to the Blockchain and onion landing pages, we used data available from two online social networks, namely, Twitter and the BitcoinTalk forum. For Bitcointalk forums, the data was collected through polite and passive crawlers that respect robots.txt instructions. We made sure that no copyright issues were encountered by going through end-user license agreements of the website. Furthermore, we collect no data from sources requiring authentication, payment, or exchange of emails. All data collected are stored securely in our private, strictly controlled infrastructure that is available only to authorized researchers.

Concerning the Twitter dataset, it is provided by Twitter through their DecaHose API[4] (a paid service). The Decahose delivers a 10% random sample of the realtime Twitter Firehose through a streaming connection delivered in bulk. We don't share the dataset and we don't distribute it openly as per the Twitter re-distribution policy. We also follow all the rules of the developer policy and agreements associated with using DecaHose API. We also follow the developer policy and agreements in the use of Twitter materials and content.

It is important to mention that researchers have to always be careful in data collection and comply with GDPR when required. In general, we collect and process the data in the interest of the public in order to uncover security and privacy issues. Ignoring the existence of the data, or the security implications of using Bitcoin as a payment method for hidden services, can leave both the users and the security community unaware of the involved privacy leaks.

To this end, we have consulted and received the approval of our institution's IRB board to conduct our experiment. We would like to highlight that our research does not put users at any additional risk, but rather expose the existing one. This is important because once users become vulnerable to this deanonymization attack, they stay vulnerable even after they switch to a more secure payment method or stop using the service. As part of our personal and institutional code of ethics, we have reached out to vulnerable users in our datasets and informed them about this threat and possible remedies. We also posted an anonymous notice on BitcoinTalk forum.[5] In Section 4.4, we discuss a number of countermeasures to improve user privacy and anonymity.

3.3 Data Collection

We now describe how we collected public Bitcoin addresses and online identities of Tor hidden services and Bitcoin users. A summary of addresses collected during this phase is in Table 1. The table shows the number of Bitcoin addresses crawled originally and after datasets expansion as discussed in Section 3.4

3.3.1 Hidden Services. Tor hidden services are not indexed by normal search engines, but can be found using indexing services such as Ahmia, which is accessible from the normal Web. Other search engines are available but require a Tor browser in order to access them. These search engines are used to access the onion landing pages, or the websites, of many hidden services. Typically, hidden services publish their Bitcoin addresses on their landing pages for receiving payments. These addresses can be collected by simply downloading these pages and searching for Bitcoin addresses using regular expressions. As a Bitcoin address is a base-58 encoded identifier of 26–35 alphanumeric characters, beginning with the number 1 or 3, we used the following regex:

```
*[13][a-km-zA-HJ-NP-Z1-9]{25,34}
```

With the goal of long-term collection of hidden service addresses, we started expanding the dataset in mid 2015. Over time, however, we found that fewer hidden services publicly exposed their Bitcoin addresses on their onion pages, resorting to online wallets or other crypto-currencies, possibly due to the increasing awareness of Bitcoin's privacy issues (Section 4). Therefore, our analysis focuses on the time window when publishing long-term Bitcoin addresses was a common practice. We note that the deanonymization attack is feasible using historic data that is publicly available since the inception of Bitcoin in 2010 up until now.

In our experiment, we first compiled a list of onion addresses from Ahmia. We then downloaded the landing pages of more than 1.5K hidden services. While our goal was to automate the process of collecting Bitcoin addresses, many of the onion addresses listed by Ahmia were unavailable or offline when we ran the scripts on Jan 27, 2016. A simple regex search on the landing pages allowed us to extract a small number of Bitcoin addresses, less than 20 addresses.

Furthermore, by browsing various hidden services, we were able to extract more addresses. We also observed that many services did not expose their Bitcoin addresses on their landing pages, and would require users to attempt purchasing items before a Bitcoin address is shown to the user. Services we manually visited offered variety of different content ranging from dark markets (e.g. drug, stolen card, and arms) to services such as WikiLeaks. In addition, we included couple of known Ransomware addresses that are published on the Web and the Blockchain.

[4] https://developer.twitter.com/en/docs/tweets/sample-realtime/overview/decahose

[5] https://bitcointalk.org/index.php?topic=2602885

Label	Date collected (dd/mm/yyyy)	# users	# addresses Original	Expanded
hiddenServices	27/01/2016	88	88	–
twitterUsers	30/12/2014	4,183	4,183	623,189
forumUsers	26/10/2016	40,970	40,970	19,213,141

Table 1: Dataset summary of collected Bitcoin addresses

Our search resulted in a total of 105 Bitcoin addresses. We verified that those addresses were active by downloading their transactions. We removed addresses that contained no transactions or had very low amount of Bitcoins, less than 0.00001Ƀ, and are likely to be inactive. This resulted in 88 unique Bitcoin addresses which represent the hiddenServices dataset, as summarized in Table 1.

While the number we ended up with seem relatively small compared to the total number of hidden services, our goal is to show the feasibility of linking users to hidden services using only public information. As described in Section 3.1, an adversary that possess wider access to resources, or actively interacts with hidden services, is expected to collect significantly larger number of addresses.

3.3.2 Users. Bitcoin users often post their addresses on online social networks for different purposes, such as receiving donations, offering services, or showing that 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 personal information representing their online identities, such as contact information, gender, age, and location, depending on the social network used.

Bitcoin addresses and the associated online identities of users can be collected by crawling and parsing their user profiles or by using the native APIs provided by the social network itself. In our experiment, we collected the addresses and identities of users of two online social networks, specifically, Twitter and the BitcoinTalk forum, as summarized in Table 1. We note that there is a one-one mapping between an address and its associated online identity. We assume that the address found on the profile of a specific user belongs to that user and in our case only one Bitcoin address was found in the "address" field of their BitcoinTalk profile. For Twitter, we also maintain a one address to one Twitter account handle.

Twitter. We used Twitter Decahose stream data [38] that we previously collected from Dec 11, 2013 to Dec 30, 2014. Decahose provides a 10% realtime random sampling of all public tweets through a streaming connection. The reason we chose this dataset is because we wanted to find Bitcoin addresses for Twitter users, which coincide with our hidden services' Bitcoin addresses. Recall that prior to 2016, it was more common for users and hidden services to share their long-term addresses. Overall, data collection resulted in 10TB of JSON-formatted files representing 5 billion tweets. In addition to its textual content, each tweet has the public profile information of its author, which sometimes contains the author's Bitcoin address. To extract tweets that contain Bitcoin addresses, we scanned the whole dataset and kept the tweets that matched the regex described in Section 3.3.1, resulting in 509,173 tweets. Next, we ran another pass on the matched tweets to group them into unique Bitcoin addresses. From 509,173 matched tweets, we found 4,183 unique addresses and identities, where an address of an identity appeared in 165 different tweets, on average. We refer to this list of addresses as the twitterUsers dataset.

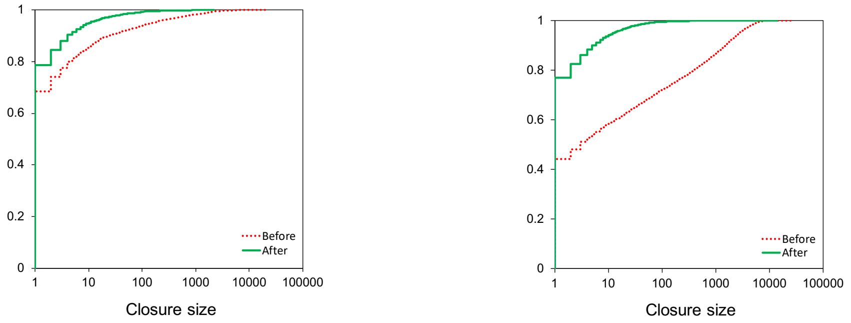
Forum. BitcoinTalk is one of the most popular Bitcoin forums with more than 900K users who exchange interests, technical expertise, and experiences in the development of the Bitcoin software. The forum also has several different sections for coin mining, technical support, and the economy of Bitcoin. It is the first forum of its kind that discusses topics related to Bitcoin and has reached its billionth post in Jul, 2012. As of Nov 2017, the forum contained around 2.5 billion posts. Based on its popularity, we sought to use it as a resource to extract addresses of Bitcoin users.

In our experiment, we crawled and parsed 900K user profiles by retrieving each profile page using its URL, where each page is indexed by a unique user identifier that starts from 1. Overall, the crawling resulted in 22 GB of unparsed user profiles. Having the profiles downloaded, we parsed them searching for Bitcoin addresses using the regex described in Section 3.3.1, resulting in 40,970 unique addresses and identities. We refer to this list as forumUsers dataset.

3.4 Wallet-Closure Analysis

The goal of wallet-closure analysis is to expand the set of Bitcoin addresses that are controlled by a user in order to establish a unique many-one mapping between addresses and an identity. Increasing the number of Bitcoin addresses per user allows us to identify more links between the user and hidden services. Using the first wallet-closure heuristic originally proposed by Meiklejohn et al. [24], 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 the mixing process results in closures that include addresses belonging to multiple identities. Accordingly, one might end up with closures that have a large number of addresses that are not mutually exclusive. In other words, there will be some Bitcoin address that appear in multiple closures. Below we explain how we reduce the noise.

Mixing services are third party services that receive coins from one user, mixes the coins with those received from other users, and then sends back the same amount of coins, albeit shuffled with



(a) twitterUsers

(b) forumUsers

Figure 1: CDFs of closure size before and after cleaning

different addresses using a number of transactions, to the original user. CoinJoin, on the other hand, is a peer-to-peer mixing protocol that achieves a similar goal, but uses a more sophisticated approach. These services are used to improve the anonymity of transactions and reduce linkability. In order to eliminate the effect of mixing services, we perform the following cleaning process: We find all closures that share at least one address and consequently merge them, after which we remove all closures that have been merged, ending up with unique closures that have no intersections and are mutually-exclusive. While doing so ensures that wallet-closures which belong to multiple users are not double counted, it also means that the resulting number of users that are linked to hidden services represents a lower-bound estimate of the actual number of user that can be linked and deanonymized.

In our experiment, after applying wallet-closure, we were able to expand the twitterUsers dataset by 619,006 additional addresses for 1,322 users out of the 4,183. The closures were more significant for the forumUsers dataset, where we were able to add 18,508,012 addresses for 22,843 users out of the 40,970. In total, for the two datasets, we ended up with 19,172,171 addresses for 45,153 users, with an average of 425 addresses per user representing the average closure size. After wallet-closure cleaning, we ended up with 3,640 closures for the twitterUsers dataset, and 23,567 closures for the forumUsers dataset. These closures under-approximate user wallets, where each consists of at least one Bitcoin address and is uniquely mapped to its owner who is a user with an online identity. We use these closures to link users to hidden services in the next section.

Figure 1 shows the closure size CDFs for both datasets, before and after the cleaning process. As illustrated in the figure, there is a significant drop in the size of closures after cleaning; the average size of a wallet decreased from 75 addresses to 7 for the twitterUsers dataset, and from 452 addresses to 6 for the forumUsers dataset. The standard deviation also decreased from 606 to 67 and from 1194 to 114, respectively. Which suggests that wallet sizes are getting closer to the mean. In fact, more than 90% of the users in both datasets have 10 addresses or less in their wallets after cleaning. The figure also suggests that a larger number of BitcoinTalk users had larger wallet size than Twitter users, as shown by the difference in their before/after distributions.

As part of our closure analysis, we looked for possible ground-truth information to support our results. For that, we used a well-known web service called WalletExplorer[6], which uses similar approach to calculate wallets and tag them based on aggregated information from the web. We crawled wallets information for both of our datasets. Figure 2 shows the top 10 wallet tags for each dataset. As seen in the figure, most of these wallets belong to Bitcoin exchanges, which means some of these users have their wallets on online exchanges. We observed that the results from our analysis corresponds to those obtained from WalletExplorer. We found that after cleaning, the number of tagged wallets decreased from 120 to 67 for forumUsers dataset and from 50 to 37 tags for twitterUsers dataset. What we noticed from WalletExplorerer results, is that that from our datasets, all of the addresses, before cleaning, that were in wallets with an address count < 700 were untagged. After cleaning we noticed that the number of wallets with address count < 700 increased from 83% and 97.63% to 99.95% and 99.75% for both forumUsers and twitterUsers datasets, respectively. Untagged wallets highly suggest that these wallets are personal wallets and not a part of an exchange or a service. This shows how cleaning reduced the likelihood of noise and improved our unique wallets mapping.

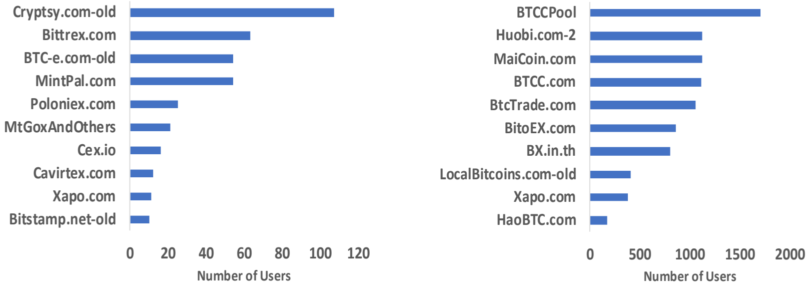
3.5 Bitcoin Transaction Analysis

We now describe how we linked users to hidden services, give two deanonymization case studies, and analyze the economic activities of the linked services.

3.5.1 Linking. In order to establish a link between a user wallet and a hidden service, we need to search the Blockchain for a transaction whose input is any of the addresses in the wallet and whose output is a hidden service address.

In our experiment, we first downloaded the whole Blockchain using the Bitcoin Core client software, which is also responsible for managing the client's runtime and transactions. As of March 2018, the time of the analysis, the size of the Blockchain was over 230GB. The Bitcoin Core client does not provide an easy, native way to access Blockchain transactions. For that, we used BlockSci [17], which is a high performance, open-source platform for Blockchain analytics.

[6] www.walletexplorer.com



(a) Twitter Wallets

(b) Bitcointalk Wallets

Figure 2: Top 10 tagged wallets by number of users from forumUsers and twitterUsers datasets

We performed the linking process after wallet-closure as follows: For each address in the hiddenServices dataset, we queried the Blockchain for the transactions history of that address. This query returns a set of transactions in which the address appears as either an input or an output. We then issued the same query for each address in the twitterUsers and forumUsers datasets. This resulted in two sets of transactions; one for hidden services and one for users. Finally, we cross-matched the two sets of transactions; if any address of a user is found as an input in any transaction where a hidden service address appears as an output, and vice-versa, then the user has a relationship with that hidden service, and thus, a link is established. For the twitterUsers dataset, we were able to link unique 28 users to 14 hidden services using 167 transactions. Similarly, for the forumUsers dataset, we were able to link unique 97 users to 20 hidden services using 115 transactions. Some of these users were linked to multiple hidden services, and a total of 20 unique hidden services were linked to users from the two datasets. As suggested by the results, although Twitter users were smaller in number compared to BitcoinTalk users, they were more active and had a larger number of transactions with hidden services. In fact, some of these users are "returning customers," as they have performed multiple transactions with the same hidden services.

Table 2 lists the hidden services sorted by how many users were linked to them. The list is topped by WikiLeaks, a service that publishes secret information provided by anonymous sources, with 46 linked users. This is followed by the Silk Road, a famous black market on the Dark Web, with transactions from 22 users whose input coins have been seized by the FBI. Although the wallet address of Silk Road was seized, it is still receiving transactions until recently. However, from our analysis, we observed that a number of transactions were performed prior to the seizure. Ranked fifth, The Pirate Bay, which is known for infringing IP and copyright laws by facilitating the distribution of protected digital content, was linked to 10 users. Other services listed in Table 2 include hacking services, such as Liberty Hackers, mixing services, such as Darknet Mixer, and various secure mailing providers, such as ProtonMail.

3.5.2 Deanonymization. As linked users have associated online identities, they could be deanonymized with different levels of certainty, depending on how much personally identifiable information they have shared on their social network's user profiles. We next focus on two case studies that illustrate this threat in more detail. It is important to note that we found a number of sensitive details about these users including location, gender, age and email addresses. However, due to ethical considerations, we only disclosed the information that demonstrate the privacy implications of this type of analysis.

The first case study is of a Pirate Bay hidden service user. The associated online identity indicates that the user is a middle aged man from Sweden. This user is particularly interesting because 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 [33]. Since then, the website has been changing its domain constantly, and eventually operated as a Tor hidden service. Therefore, the existing link between this user and The Pirate Bay through recent transactions could be incriminating.

In the second case, we focus on the Silk Road hidden service. As shown in Table 2, there are 22 users that had a link to Silk Road through transactions with seized Bitcoin addresses. These users are located across the world in countries such as India, Canada, and the USA. They include 4 males and 6 females of different ages that range between 13 and 42 years. The 18 users from the forumUsers dataset were active on BitcoinTalk between 2013 and 2015, while 3 of them were active in 2018. As for the 4 users from the twitterUsers dataset, they posted an average of 45 tweets in 2014. One particularly interesting user is a young teenager from the USA. This user has been a registered BitcoinTalk member since 2013, and had a transaction with Silk Road in 2013, the takedown year, during which he was even younger than what his current age shows on his profile. The associated profile also includes his personal website, which contains links to his Facebook, Twitter, and Youtube profiles. His activities on these social media accounts further confirmed the teenager's age and identity.

3.5.3 Economic Activity. In order to gain insights about the economic activities of the linked hidden services, we analyzed all of their transactions in the Blockchain. In our experiment, for each service, we collected information about its number of transactions (i.e., volume), the amount of coins the service has received or sent (i.e., flow of money), and the timestamps of its first and last transactions (i.e., lifetime). The results are also summarized in Table 2.

	# linked users			Volume	Flow of money (฿)		Lifetime (dd/mm/yyyy)		
Name	twitterUsers	forumUsers	Total	(# txs)	Incoming	Outgoing	First tx	Last tx	# days
WikiLeaks	11	35	46	26,399	4,043.00	4,040.74	15/06/2011	21/03/2018	2,470
Silk Road	4	18	22	1,242	29,676.99	29,658.80	02/10/2013	19/03/2018	1,628
Internet Archives	3	13	16	2,957	775.86	746.89	06/09/2013	21/03/2018	1,656
Snowden Defense Fund	3	8	11	1,722	218.95	218.95	11/08/2013	18/03/2018	1,680
The Pirate Bay	3	7	10	1,214	76.80	76.80	29/05/2013	21/08/2017	1,544
DarkWallet	9	1	10	983	114.62	97.40	16/04/2014	02/11/2016	931
ProtonMail	1	7	8	3,096	208.40	208.36	17/06/2014	18/03/2018	1,369
OpenStreetMap Donations	0	5	5	440	24.01	24	13/05/2013	13/07/2017	1,522
Darknet Mixer	1	2	3	22,110	306.16	341.48	21/01/2014	24/07/2017	1,645
Liberty Hackers	0	2	2	85	2.79	2.79	10/04/2013	11/07/2017	1,553
Onion Mail	1	0	1	226	84.92	84.92	15/09/2014	30/03/2015	196
Bitcoin Fog	1	0	1	121	10.46	10.46	12/08/2014	09/01/2015	150
Bitmessage Mail	0	1	1	106	2.78	1.37	28/04/2014	16/06/2017	1,145
Secure Tor Messaging	0	1	1	105	12.5	12.5	02/01/2013	05/01/2016	1,098
Ransomware	1	0	1	72	41.15	41.15	28/02/2014	15/08/2014	168
Bitcoin Lottery	0	1	1	33	0.22	0.22	28/02/2014	02/05/2015	428
Libertarian Nuts	0	1	1	23	0.31	0	28/02/2014	24/10/2015	603
Unknown1	1	1	2	42	8.32	8.31	23/06/2014	23/06/2017	1,096
Unknown2	1	0	1	132	4.19	4.18	22/06/2014	01/01/2015	193
Unknown3	0	1	1	39	7.95	7.94	01/05/2014	18/05/2017	1,113

Table 2: Linked hidden services. The three unknown services belong to onion pages that were taken down before we could manually identify and validate their hidden service provider.

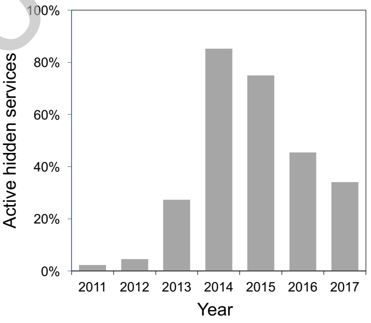


Figure 3: Lifetime of hidden services

Volume. While the list of services is small, our results indicate that they have been involved in a relatively large number of transactions. For example, WikiLeaks tops the list with 25,569 transactions. The Darknet Mixer, on the other hand, has a volume of 22.1K transactions that is greater than the remaining services combined. One explanation for this popularity is that users are actually aware of the possibility of linking, and try to use mixing services in order to make traceability more difficult and improve their anonymity.

We found out that 10% or less hidden services had more than 1,000 transactions. This complies with our previous results showing that most of the transactions from twitterUsers and forumUsers datasets were attributed to the top 4–5 hidden services, in terms of volume.

Money flow. We calculated the total incoming and outgoing Bitcoins for each service in order to determine how much money is flowing in and out of their addresses. One interesting observation is that almost the same amount of coins flow in and out of these addresses. This indicates that the money is being distributed to other addresses, and is not stored on payment-receiving addresses. One explanation for this behavior is that by distributing funds to other addresses, a hidden service can reduce coin traceability. Also, hidden services still need to distribute their revenues among owners, sellers, and other stakeholders.

We also observed that multiple hidden services have a revenue of more than 4K Bitcoins and up to 29.6K, where one Bitcoin was valued at seven thousand USD as of August, 2017. The Silk Road, for example, has received more than 580 million USD on its address.

One interesting observation was that the money flowing in and out of hidden service addresses is almost identical. We found out that only the top 10% of hidden services had received more than 100฿. As shown in Table 2, top-3 services had significantly bigger revenue, with more than 4,000฿.

Lifetime. Tracking the economic activity of hidden services over time allows us to estimate their operational lifetime, at least as seen from their associated addresses. From the transaction history of each hidden service, when filtered by its Bitcoin address, we define the lifetime of the service as the difference between the timestamps of the last and the first transactions involving the address. This allows us to estimate the period of operation of each hidden service, and accordingly, determine if the service is still active.

As summarized in Table 2, hidden services vary in their lifetime, ranging from 2–6 years of operation. We note that the first transaction date does not imply that the service began its operation on that date; it indicates the date on which the service started receiving Bitcoin payments. Looking at last transaction dates, most of the hidden services are still active in 2017.

Figure 3 shows the percentage of hidden services that were active during the period 2011–2017. As denoted by the figure, most of these services were active during the years 2014–2015, which is the same time frame we based our data collection period on. Furthermore, as expected, the activity for their addresses is significantly lower in the following years. This is most likely due to the usage of online wallets or their migration to other cryptocurrencies, as discussed in Section 4.

4 DISCUSSION

We now discuss the deanonymization attack, focusing on its feasibility and implications. We also list the limitations of this work and highlight a number of existing countermeasures for improved privacy and anonymity.

4.1 Feasibility

As discussed in Section 3.1, we consider a passive, limited adversary that falls within the adversary model of Tor, and focus on hidden services that use Bitcoin as a payment method. Through a real-world experiment, we showed that an 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 accomplished by inspecting the transactions involving these two addresses in the Blockchain.

While collecting Bitcoin addresses of hidden services, we found that many services started to hide their payment Bitcoin addresses from onion landing pages. One explanation of this trend is that hidden service operators realized that publicly sharing Bitcoin addresses can leak information which could be used for linking and improving traceability. Instead, these services let users register accounts on their website and use them to perform transactions without exposing the addresses used to receive Bitcoins. The way this works is as follows: If Alice wants to perform a transaction with a hidden service, she starts by creating a personal account on the service. The hidden service creates and controls a new Bitcoin address, from a new public/private key pair, to which Alice makes a transaction from her personal Bitcoin addresses.

An active, more resourceful, local adversary can compile a larger set of hidden service addresses by performing small transactions with such services in order to reveal their addresses. Moreover, adversaries can impersonate a hidden service and receive payments on addresses they control. Deanonymization that exploit such techniques has been practiced in the wild by governments to uncover illegal hidden services [7]. The adversary can also map collected Bitcoin addresses to IP addresses in order to deanonymize users at a more granular level [20].

In this work, due to ethical concerns, we simulated only a passive, limited adversary. However, active, local adversaries that are well-funded are likely to exist in practice. We note that the success rate achieved by a passive, limited adversary represents a lower bound of that of an active, local adversary.

4.2 Implications

The main security implication of our work is that a Bitcoin addresses can be exploited to deanonymize users. The experiment we conducted can be extended to include other online social networks and information sources. It is also likely that an online identity, or a user account, on one social network has links to others networks that provide additional PIIs. Some users explicitly reveal their name, age, nationality, and other information in their bio or through posts. This represents a serious threat to their anonymity, since the hidden services they engage with might be associated with questionable transactions. The linking process can also be used as a tool by third-parties to track users, perform surveillance, and audit financial transactions.

Due to these security and privacy concerns, users have to follow simple yet effective guidelines in order to protect themselves. First, users should never expose their Bitcoin addresses along with their personally identifiable information. Second, as discussed in Satoshi's white paper [28], a new address should be generated for each transaction in order to reduce linkability of transactions, regardless of whether the user is the sender or the receiver of the payment. This is especially true for cases where users expose a donation address on different kinds of online social networks. Unfortunately, a large number of users do not follow this practice, possibly due to poor usability of Bitcoin tools, unfamiliarity of Bitcoin internals, or reliance on third-party wallets and exchange services [21].

4.3 Limitations

Our work has two main limitations. First, in our analysis, we assume that linking a user, represented as an online identity, to a hidden service is sufficient to deanonymize the user. This is not always true. Users can always create fake identities in order to hide their real ones. While doing so improves their anonymity, once the links are established the adversary can perform online surveillance to track down the users and uncover their true identities. Even if the identities are fake, an adversary with more resources can get access to users' IP addresses and deanonymize targets. An example would be the subpoena of Bitcointalk forums in 2014 [5]. The second limitation is the use of mixing services. While the wallet-closure cleaning process we used eliminates the effect of mixing, it is aggressive and could exclude users who did not use mixing services. Accordingly, our results underestimates the prevalence of the deanonymization threat.

It is important to mention that some of the linked users through this analysis are not concerned about their transactions anonymity to services such as Internet Archives. However, some users are unconscious about protecting their privacy, and this type of analysis can be used to incriminate these users. For example, services such as The Pirate Bay are illegal in Sweden, and therefore, establishing a link through a transactions with such a service could potentially be used as an evidence to accuse the user of illegal activities.

One might argue that the number of deanonymized users is small. However, in order to understand the significance of our results, it is useful to put numbers into perspective. According to recent web statistics [30, 36], the number of worldwide Internet users is around 3.58 billion and the number of Tor network users is about 3 million, which are a superset of Tor hidden service users. This means that 0.086% of Internet users are Tor network users, on average. The datasets we collected include 45,153 Twitter and BitcoinTalk users, out of which 0.277% of them were deanonymized. While this percentage is larger than 0.086%, likely due to biased sampling, it is still relatively small, as expected. In other words, because we do not know how many of the users in the datasets are Tor hidden service users, we should expect that only a small percentage can be deanonymized.

4.4 Countermeasures

There are two general ways to achieve improved anonymity for users and hidden services. The first one is operational, and it focuses on following Bitcoin best practices, as discussed in Section 4.2. For those users who can be linked, the best course of action for them is to clean their social network footprint, focusing on removing PII that is publicly shared or deleting their linked online identities, all together. The second way to improve anonymity is technical, and it involves improvements to the current Bitcoin protocol or the introduction of new crypto-currencies that are based on Bitcoin's Blockchain technology.

Second generation anonymization techniques, such as CoinJoin, Fair Exchange [15], CoinSwap, and stealth addresses have been proposed to be implemented as extensions or services for Bitcoin's original protocol. These are discussed in details in [27]. Furthermore, other alternative coins based on different modifications to Bitcoin protocol have been introduced to provide additional anonymity for transactions on the Blockchain. The most prominent of them being Monero, which is based on Crypto Note v2.0 protocol [34], which is privacy focused, and Zcash [35], which also offers privacy and selective transparency. In fact, Monero is already making its way into the hidden services of the Dark Web [37] and many services already began using it as a payment method. These new coins aim to deliver full anonymity of transactions, and have their own advantages and disadvantages compared to Bitcoin.

5 RELATED WORK

There are a number of studies that investigate user anonymity and privacy concerns in Bitcoin [12, 13, 16, 31]. Fergal and Martin [31] demonstrated that using passive analysis of publicly-available Bitcoin information can lead to a serious information leakage. They constructed two networks representing transactions and users from the Blockchain. Integrating these networks with off-network information, such as user profiles from online social networks, and context discovery and flow analysis techniques, it was possible to study the flow of Bitcoins between addresses and investigate thefts. Fleder et al. [13], on the other hand, explored the level of anonymity in the Bitcoin network. The authors annotated the transaction graph by linking user pseudonyms to online identities collected from online social networks. They also developed a graph-analysis framework to summarize and cluster the activity of users. The analysis links identities of users to their transactions. These studies form the base for our approach, as we use some of their techniques in our analysis. The difference in our study is that we target a specific portion of Bitcoin users, which are Tor hidden service users. We also study the economic activities of linked hidden services, which is important to understand the level of threat the users are exposed to.

DuPont and Squicciarini [12] proposed a technique to determine a Bitcoin user's physical location by examining user spending habits and linking it to the user's time zone. Androulaki et al. [1] 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. This method can be used with our technique to increase the deanonymization level to the physical identity.

A recent study by Harrigan and Fretter [14] showed the effectiveness of address clustering using the Blockchain of Bitcoin. These clusters are constructed using different heuristics such as the one we used in our study. The authors performed address clustering on the entire Bitcoin's Blockchain and showed that despite the existence of CoinJoin and mixed transactions, address clustering is still suitable for Blockchain analysis and re-identification attacks. The findings presented in their work strengthen our analysis results and further proves that our linking to hidden services is still valid until the current time.

6 CONCLUSION

We show that using Bitcoin as a payment method for Tor hidden services leaks information that can be used to deanonymize their users. This represents a serious threat to these users, because they actively seek to maintain their anonymity by using Tor. The deanonymization is mainly due to the lack of retroactive operational security present in Bitcoin's pseudonymity model. In particular, by inspecting transactions in the Blockchain, an adversary can retroactively link users, who publicly share their Bitcoin addresses on online social networks, with hidden services that publicly share their Bitcoin addresses on their landing pages.

In a real-world experiment, we were able to link many users of Twitter and the BitcoinTalk forum to various hidden services, including WikiLeaks, Silk Road, and The Pirate Bay. Using information from their public user profiles, we were able to show concrete case studies where the anonymity of the users is broken. Our results has one immediate implication: Bitcoin addresses should always be assumed compromised as they can be used to deanonymize users.

ACKNOWLEDGEMENT

We would like to thank Hasan Al Jawaheri, Ryan Riley, and Jaideep Srivastava for their help and feedback. The first author would like to thank Qatar University for a generous research assistantship. This work was made possible by NPRP grant 7-1469-1-273 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

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

[2] Alex Biryukov and Ivan Pustogarov. 2015. Bitcoin over Tor isn't a Good Idea. In *2015 IEEE Symposium on Security and Privacy.* IEEE, 122–134.

[3] Alex Biryukov, Ivan Pustogarov, Fabrice Thill, and Ralf-Philipp Weinmann. 2014. Content and popularity analysis of Tor hidden services. In *2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW).* IEEE, 188–193.

[4] Michael del Castillo. 2016. Bitcoin Remains Most Popular Digital Currency on Dark Web. CoinDesk. (2016).

[5] Caleb Chen. 2014. Bitcoin Forum Bitcointalk.org Receives and Answers First Subpeona. (Sep 2014). https://www.ccn.com/bitcoin-forum-bitcointalk-org-receives-answers-first-subpoena-doj/

[6] Nicolas Christin. 2013. 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, 213–224.

[7] Amanda Cochran. 2013. What was alleged Silk Road mastermind's "fatal flaw"? Find out how FBI tracked him down. (Oct 2013).

[8] Anthony Cuthbertson. 2015. Bitcoin now accepted by 100,000 merchants worldwide. International Business Times. (2015).

[9] Roger Dingledine. 2012. Using Tor Hidden Services for Good. The Tor Project. (2012).

[10] Roger Dingledine. 2014. Tor security advisory: "relay early" traffic confirmation attack. The Tor Project. (2014).

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

[12] Jules DuPont and Anna Cinzia Squicciarini. 2015. Toward de-anonymizing Bitcoin by mapping users location. In *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy.* ACM, 139–141.

[13] Michael Fleder, Michael S Kester, and Sudeep Pillai. 2015. Bitcoin transaction graph analysis. *arXiv* (2015).

[14] Martin Harrigan and Christoph Fretter. 2016. The Unreasonable Effectiveness of Address Clustering. In *Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences.* IEEE, 368–373.

[15] Danushka Jayasinghe, Konstantinos Markantonakis, and Keith Mayes. 2014. Optimistic Fair-Exchange with Anonymity for Bitcoin Users. In *e-Business Engineering (ICEBE), 2014 IEEE 11th International Conference on.* IEEE, 44–51.

[16] Sarah Meiklejohn Marjori Pomarole Grant Jordan, Kirill Levchenko Damon McCoy, and Geoffrey M Voelker Stefan Savage. 2013. A Fistful of Bitcoins: Characterizing Payments Among Men with No Names. (2013).

[17] Harry Kalodner, Steven Goldfeder, Alishah Chator, Malte Möser, and Arvind Narayanan. 2017. BlockSci: Design and applications of a blockchain analysis platform. *arXiv preprint arXiv:1709.02489* (2017).

[18] Amin Kharraz, William Robertson, Davide Balzarotti, Leyla Bilge, and Engin Kirda. 2015. Cutting the gordian knot: A look under the hood of ransomware attacks. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment.* Springer, 3–24.

[19] Paul Kimayong. 2016. New Family of Ransom Locker Found, Uses TOR Hidden Service. https://goo.gl/a8Erio. (2016).

[20] Philip Koshy, Diana Koshy, and Patrick Mcdaniel. 2014. An Analysis of Anonymity in Bitcoin Using P2P Network Traffic. *Financial Cryptography and Data Security Lecture Notes in Computer Science* (2014), 469–485.

[21] Katharina Krombholz, Aljosha Judmayer, Matthias Gusenbauer, and Edgar Weippl. 2016. The other side of the coin: User experiences with bitcoin security and privacy. In *International Conference on Financial Cryptography and Data Security.* Springer, 555–580.

[22] Moser M. 2013. Anonymity of bitcoin transactions: An analysis of mixing services. *Munster Bitcoin Conference* (2013).

[23] Nick Mathewson. 2014. Some Thoughts on Hidden Services. The Tor Project. (2014).

[24] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. 2013. A Fistful of Bitcoins: Characterizing Payments Among Men with No Names. In *Proceedings of the 2013 Conference on Internet Measurement Conference (IMC '13).* ACM, New York, NY, USA, 127–140. https://doi.org/10.1145/2504730.2504747

[25] Vincent Van Mieghem and Johan Pouwelse. 2015. Anonymous online purchases with exhaustive operational security. *CoRR* abs/1505.07370 (2015). http://arxiv.org/abs/1505.07370

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

[27] Malte Möser and Rainer Böhme. 2017. Anonymous alone? measuring Bitcoin's second-generation anonymization techniques. In *2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW).* IEEE, 32–41.

[28] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. (2008).

[29] The Tor Project. 2015. A Hidden Service Hackfest: The Arlington Accords. Tor. (2015).

[30] The Tor Project. 2017. Tor Metrics Portal. https://metrics.torproject.org. (2017).

[31] Fergal Reid and Martin Harrigan. 2011. An analysis of anonymity in the bitcoin system. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on.* IEEE, 1318–1326.

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

[33] Mikael Ricknäs. 2017. Pirate Bay Appeals Looks Set to Start in September. http://www.pcworld.com/article/191304/article.html. (2017).

[34] Nicolas van Saberhagen. 2013. Crypto Note v 2.0. *Cryp to Note* (2013).

[35] E. B. Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza. 2014. Zerocash: Decentralized Anonymous Payments from Bitcoin. In *2014 IEEE Symposium on Security and Privacy.* 459–474. https://doi.org/10.1109/SP.2014.36

[36] Statista. 2017. Number of internet users worldwide from 2005 to 2017. (July 2017). https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/

[37] Kyle Torpey. 2016. Darknet Customers Are Demanding Bitcoin Alternative Monero. (August 2016). http://www.nasdaq.com/article/darknet-customers-are-demanding-bitcoin-alternative-monero-cm671031

[38] Twitter. 2016. Gnip Decahose: Real-Time Trend Detection and Discovery. (2016). https://gnip.com/realtime/decahose/

[39] Joe Van Buskirk, Sundresan Naicker, Amanda Roxburgh, Raimondo Bruno, and Lucinda Burns. 2016. Who sells what? Country specific differences in substance availability on the Agora cryptomarket. *International Journal of Drug Policy* 35 (2016), 16–23.

[40] Aaron van Wirdum. 2016. CoinJoin: Combining Bitcoin Transactions to Obfuscate Trails and Increase Privacy. https://goo.gl/LwT5u1. (2016).

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

- **Husam Al-Jawaheri**

Husam received his MSc. In Computing from the Computer Science and Engineering department of Qatar University (QU). He was mentored by Dr. Qutaibah Malluhi and Dr. Mashael Al-Sabah. His master’s thesis presented one of the first deanonymization attacks on Tor Hidden Services that rely on Bitcoin as a payment platform. His main research interests revolve around privacy and anonymity, blockchain and cryptocurrencies, and information security.

- **Dr. Yazan Boshmaf**

Dr. Yazan Boshmaf received his PhD. in Electrical and Computer Engineering from the University of British Columbia (UBC) under the mentorship of Prof. Konstantin Beznosov and Prof. Matei Ripeanu. His PhD thesis presented one of the first security analyses of malicious socialbots on the web. Dr. Boshmaf's research centers around the security and privacy of social and information networks, with an emphasis on problems that broadly impact the way people use technology and the Internet.

- **Dr. Mashael Al-Sabah**

Dr. Mashael Al-Sabah is a Scientist at Qatar Computing Research Insitute. Her research interest includes the general area of privacy enhancing technologies with a particular interest in anonymous and secure communication, cryptocurrency, traffic analysis and side-channel attacks. Dr. Mashael’s research papers are assigned readings in a number of top US universities such as UC Berkeley, and Johns Hopkins University. She received many awards including the Andreas Pfitzmann best paper award at Privacy Enhancing Technologies Symposium and received the Platinum award for education excellence by His Highness the Emir. Dr. Mashael spent 15 months at MIT as a visiting scientist during which she identified a critical security vulnerability which was featured on MIT’s front webpage, dozens of news articles, and led to the shutdown of Agora, one of the largest dark web markets. Dr. Mashael obtained her PhD in Computer Science from the University of Waterloo in 2013.

- **Aiman Erbad**

Dr Aiman Mahmood Erbad is the Director of Research Support and an Assistant Professor of Computer Engineering at Qatar University. Dr. Erbad obtained a PhD in Computer Science from the University of British Columbia (Canada), a Master of Computer Science in Embedded Systems and Robotics from the University of Essex (UK), and a Bachelor of Science in Computer Engineering from the University of Washington (USA). Dr. Erbad research interests span cloud computing, multimedia systems and networking, and security. He published his work in reputed international conferences and journals. Dr. Erbad is leading a number of funded grants. Dr. Erbad acts as a technical program member in international conferences and as an expert in information technology strategy and research techniques for various national entities.