Mixed Signals: Analyzing Ground-Truth Data on the Users and Economics of a Bitcoin Mixing Service

Fieke Miedema, Kelvin Lubbertsen, Verena Schrama,
and Rolf van Wegberg, *Delft University of Technology*

https://www.usenix.org/conference/usenixsecurity23/presentation/miedema

This paper is included in the Proceedings of the
32nd USENIX Security Symposium.

August 9–11, 2023 • Anaheim, CA, USA

978-1-939133-37-3

Open access to the Proceedings of the
32nd USENIX Security Symposium
is sponsored by USENIX.

Mixed Signals: Analyzing Ground-Truth Data on the Users and Economics of a Bitcoin Mixing Service

Fieke Miedema, Kelvin Lubbertsen, Verena Schrama, and Rolf van Wegberg

Delft University of Technology

Abstract

Bitcoin mixing is a commodity, mostly offered in the underground economy, selling anonymity in the bitcoin ecosystem. Its popularity is rather remarkable, as transactions initiated by its users run through wallets of a centralized service where personal identifiable information is collected in the mixing process, without any prior knowledge of data retention policies. This leaves us to wonder if users resort to strategies to mitigate these risks – like the usage of IP proxy services – or test the service with smaller transactions to identify scam services at 'low' costs.

In this paper, we explore unique ground-truth data capturing 15,574 mixing transactions, initiated by 8,838 users, totaling US $45M worth of bitcoins mixed through *BestMixer* between July 2018 and June 2019. We find that user adoption of risk mitigation strategies is limited, while transaction volumes users entrust *BestMixer* are high and usage is frequent and recurrent – with 23% of users returning. Our analysis shows that only 61% of all transactions used some form of IP address obfuscation – i.e., VPN or VPS usage. We discuss possible explanations for these findings, including how information asymmetries and the role of mixers in the process of cashing-out criminal proceeds might force users to accept the risks associated with bitcoin mixing. Furthermore, we address the implications of our findings for the broader cryptocurrency security ecosystem.

1 Introduction

While cryptocurrencies offer advantages in terms of decentralization and transparency, their pseudonymous nature creates challenges regarding the traceability of transactions and the privacy of users as a result of that. This traceability is seen as a liability in criminal use-cases when transferring crime proceeds in a non-incriminating manner. As a consequence, criminal entrepreneurs rely on specialized suppliers in the underground economy to fulfill their cash-out needs. As online crime is significantly intertwined with cryptocurrencies – from ransom payments to transacting with bullet proof hosters – so-called bitcoin mixers have become a prominent commodity. Bitcoin mixers advertise to anonymize bitcoin transactions, obstructing follow-the-money efforts by 'mixing' inputs and outputs of users interacting with the service.

Blockchains without privacy-enhancing features allow everyone to – in essence – follow funds across each successive transaction. Besides the risks of financial surveillance, traceability can cause users to become vulnerable to targeted attacks – for example theft (e.g., address poisoning attacks[1]) or deanonymization (e.g., dusting[2]). These users could also use mixing services to achieve their privacy aims. However, given that the goal of bitcoin mixers is to obfuscate ownership, such services do not employ Know-Your-Customer (KYC) or transactional due diligence programs and are thus, while not explicitly illegal, in certain legal systems not compliant with Anti-Money Laundering (AML) regulations [9]. Using such mixing service providers could therefore lead to the mixed funds of all users to become classified as "from an illicit source".

Additionally, a mixing service does not include any contractual safeguards – e.g., mixing services do not facilitate an escrow service. Yet, most bitcoin mixing services do facilitate a review system. Earlier work has indicated that these reviews do allow users to circumvent scam services [46]. However, reviews do not cover the financial position (i.e., reserve) the service has available for mixing transactions or any details on their data retention policy, thus creating an information asymmetry between users and the service. As a result, users run the risk of leaving incriminating evidence – i.e., IP addresses when using a clear web mixer and their bitcoin addresses – thus accepting the chances of attribution of their crime proceeds [3, 25, 36].

[1] In address poisoning attacks, the attacker poisons the transaction list of its victim, by using a vanity address similar to the address of the victim to send coins to the victim, hoping the victim uses that address [44].

[2] A dusting attack is sending very small amounts of cryptocurrency (e.g. 'dust') to multiple addresses, in order to track the coins when they are spent, to deanonymize the addresses' or wallet's owner(s) [37].

Given these intrinsic risks, we ask ourselves: how do users interact with bitcoin mixing services? Do they first test the service before mixing large amounts, or provide multiple output addresses to prevent demixing efforts? In other words, which mitigation strategies – if any – do users employ attempting to prevent attribution of their financial assets? And what can we learn from this?

We capture these mitigation strategies on a single mixing service: BestMixer, who facilitated over 200 transactions per day, until taken down by Dutch law enforcement in 2019. We leverage unique ground-truth data of more than 15,000 transactions facilitated by BestMixer on their clear web site and hidden service between July 2018 and June 2019. We compare BestMixer to other centralized mixers active during its lifetime and attribute the clusters sending to and receiving from BestMixer based on Chainalysis labelling [5]. We contextualize the role of BestMixer in the ecosystem by comparing its market share and fees to other centralized mixers. We find that BestMixer was a top-10 player, competing for a top-3 spot, with service fees similar to other services. Finally, we reflect on our findings to formulate recommendations based on the implications of our findings for the broader cryptocurrency security ecosystem. In short, we make the following contributions:

- We provide the first detailed empirical study on the users of a bitcoin mixing service, leveraging unique ground-truth transaction data of BestMixer. On average $259,951 flowed through BestMixer on a daily basis between July 2018 and June 2019, which can be extrapolated to over 190 million US dollars in just under a year of doing business. In total, the user base (n=8,838) of BestMixer initiated 15,574 transactions, with an average value of $2,887.60. We also find that 23% of user base utilized the service on a recurring basis.

- We uncover that users barely make use of commonly known techniques to safeguard anonymity. We uncover that one-third of transactions is made via a not-obfuscated IP address and the average number of output addresses per transaction is only 1.34. However, we discovered that 48% of returning users consistently accessed BestMixer via a proxy.

- We find that users who accessed the service via the hidden service of BestMixer do use significantly more output addresses per transaction (1.66). At the same time we see that these users entrust BestMixer with significantly larger funds per transaction ($12,455).

- We discuss possible explanations for our findings, including that the information asymmetry between the service and its users consequently leaves users to accept a certain risk of attribution or scam.

The remainder of this paper is structured as follows. First, we introduce the concept of bitcoin mixing in Section 2 and the concepts related to risk attribution strategies in Section 3. In Section 4 we provide an overview of our approach to process the raw, ground truth-data, do internal and external validation, outline external data sources used to enrich this ground-truth data and present descriptives thereof. Section 5 covers user strategies to mitigate the risk of scam. Our analysis on strategies mitigating financial and identity attribution is presented in Section 6 and 7 respectively. We deepen our analysis with a comparison between the use of the clear web instance and the hidden service of BestMixer in Section 8. We contextualize the role of BestMixer in the mixing ecosystem in Section 9, by analyzing its transaction volume, market share, and deposit and output cluster attribution. We discuss these results, including its limitations and implications for the cryptocurrency security ecosystem, in Section 10. We position our findings against related work in Section 11 and Section 12 concludes.

2 Cryptocurrency Mixing

In this section, we examine the concept of bitcoin mixing and describe one provider thereof – BestMixer– in more detail. Next, we compare some of the cryptocurrency mixing services that were available alongside BestMixer during its lifetime.

2.1 Mixing services

Contrary to popular belief, most cryptocurrencies – including bitcoin – are pseudonymous rather than anonymous [25]. For example, the bitcoin blockchain stores all transactions publicly. This ironic transparency has even fueled an entire industry turning this pseudonymity into an analytics product used by financial institutions to do due diligence and law enforcement to do attribution of illicit funds. This level of traceability can be problematic, however, when one for whatever reason in both legitimate or illegitimate use cases, wants to hide the origin or destination of cryptocurrency – e.g., to hide bitcoin payments to a web shop [2].

The goal of a mixing service is to obfuscate money flow, by 'mixing' different transactions, making it difficult if not impossible to connect source with destination and vice versa. Both centralized as decentralized mixing services exist. Decentralized mixing services such as Wasabi Wallet and Samourai Wallet enable users to keep control of their funds – initiating a mixing transaction with other users putting in the same denomination [42]. These mixing services therefore provide mathematical security to its users, but do require coordination with other users who want to mix a similar amount at the same time, which can make this process time consuming. Centralized mixing services on the other hand run their own wallets. Users transact the funds they wanted mixed to the central wallet, and the service pays out the same amount minus a small

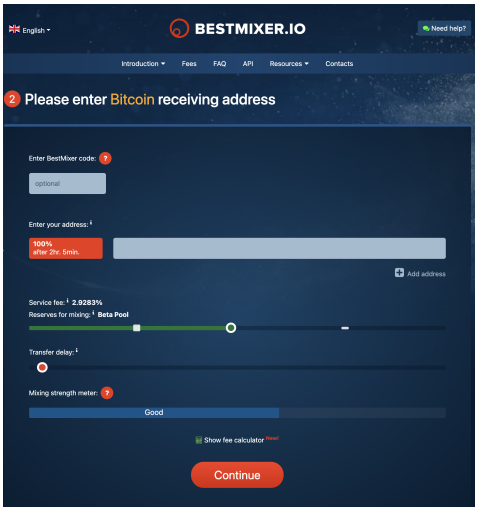

Figure 1: **Screenshot of BestMixer's order form, including the mixing strength meter.**

service fee to a specified address from earlier deposits. This way, the money flow is obfuscated, as source and destination become disconnected. Centralized mixing services are easier to use, as they do not require users to install a special wallet, and provide liquidity for larger transactions in a shorter period of time. But this ease of use does come at the risk of getting scammed [46].

2.2 BestMixer

BestMixer ran a clear web instance next to a hidden service only accessible through Tor[3]. The functionalities of the service on both instances were identical. In short, their functionality comes down to allowing users to specify a specific amount of funds to be mixed and enter one or more bitcoin addresses to which these funds should be transferred after being mixed. In return, the users pay a specific fee. By doing so, users of BestMixer, either by accessing the clear web or hidden service, could mix bitcoins to obscure any traces associated with the fund's original source.

When a user specified more than one output address, it was possible to create a payout distribution to allocate the coins to the different addresses. Moreover, users were encouraged by a mixing strength meter to specify a transfer delay and more output addresses, which would ensure a higher degree of anonymity. This mixing strength meter provided a visual indicator to the strength of the 'mix', similar to password strength indicators on login pages. This value ranged from 'weak' to 'good' up to the maximum strength of 'strong'. The strength indicator increases when the service fee, number of outputs or delay is increased. The default settings resulted in 'good', where the delay and service fee were randomly picked. To decrease the strength value to 'weak' only one output could be specified and the service fee was lowered. Also, each user received a "BestMixer code" after first use of the service. Entering this code in future mixing transactions would prevent the user from ever receiving their own coins from previous transactions. When agreed upon the terms of use, the user could download a letter of guarantee. This letter of guarantee is signed by the private key of the publicly known donation address (which was a vanity address beginning with 1BestMix...). With the letter of guarantee a user could verify that the deposit address was generated by BestMixer and one would not be scammed. Thereafter, the user had to deposit cryptocurrency within 24 hours on the provided address in order to receive it 'mixed' after the specified delay(s) on the specified address(es).

[3] One of the indicators as to why this service was taken down by law enforcement was the content and placement of their advertisements in the underground economy, such as their affiliate program with DeepDotWeb.

2.3 Mixing market ecosystem

Since the first cryptocurrency mixer BitcoinFog started in 2011, a market of competing mixers emerged. When BestMixer was operational, competitors such as ChipMixer, Blender.io, Jambler.io partners, BitcoinFog, CryptoMixer.io, Helix and CoinJoin wallets like Samourai and Wasabi were also active [9]. We will analyze the role of BestMixer in this market in Section 9. Most of these mixers advertised their services in the underground economy [9] and some, such as BestMixer, were present on fora such as BitcoinTalk.org. On that forum, the BestMixer account described its mixing security features such as "unsurpassed protection against blockchain analysis" and provided analyses of why competing mixing services were deemed "not anonymous" [4]. In recent years several mixing services have been targeted by law enforcement operations. First, BestMixer was taken offline by Dutch law enforcement in May 2019 in collaboration with various other countries and coordinated through Europol [35]. Additionally the operators of Helix Mixer and BitcoinFog have been charged by US authorities [31, 32] and in 2022 the US Treasury sanctioned Blender.io [33] and Tornado Cash [34], alongside one arrest in connection to the latter [14].

Notwithstanding these takedowns, the volume of cryptocurrency that mixers transacted has steadily increased from 2017 onward: according to Chainalysis [5], the 30-day moving average value received by mixers reached an all-time high in 2022, totaling 51.8M USD [6]. There are currently multiple centralized mixers active – such as YoMix.io [49], Sinbad.io [41] and Coinomize [8] – that provide users with features similar to BestMixer. For example, these mixers let users specify multiple output addresses, for which the user or the mixer specifies output distributions and a (random) delay between transactions. They all provide users with a code that prevents users from receiving coins from previous transactions, ask fees that range from 0.5% to 5%, and offer both a clear web site and hidden service.

3 User's security practices

We identify expected security practices of users interacting with bitcoin mixing services. Users make use of cryptocurrency mixers to achieve a higher degree of anonymity. Their goal is mainly to have cryptocurrency funds that are less likely to be traced back to their wallets or to their identity. The security strategies to achieve these goals can be characterized as *financial attribution mitigation* and *identity attribution mitigation* respectively. The landscape of mixing services is filled with services that are scams, making it necessary to employ steps to prevent being scammed while using such a mixing service. We name these security strategies *scam mitigation*.

First, for financial attribution mitigation, the users can employ different strategies: choosing a high number of output addresses, choosing delays between outputs and defining an uneven output distribution [19]. These strategies were also promoted by BestMixer themselves through the aforementioned mixing strength meter. Second, to mitigate the attribution of one's identity to the funds being mixed, users can take steps to prevent IP address reuse and IP address deanonymization. Strategies may include using the Tor instance of BestMixer, but also utilizing VPN-services or to set up a VPS to obfuscate their IP address. Last, users can exercise security strategies to mitigate the impact of the service being a scam. Users can test the BestMixer service with a transaction of little monetary value, to verify that the service works as advertised. Next, users can choose to not send large transaction volumes at once, but make multiple, smaller transactions over a longer period of time. This way they do not risk losing a large amount in case the service executes a form of an exit-scam.

4 Methodology

To gain insight into mitigation strategies adopted by bitcoin mixer users, we were granted access to packet capture data from wiretaps on BestMixer servers, placed by law enforcement (LE) during the investigation into the service. In this section, we first define our approach to extract features from the packet capture data. Next, we describe and validate the resulting dataset. We then introduce external datasources we employ to enrich our data and present the high-level descriptives of the resulting dataset. The ethical considerations are described in Section 10.

4.1 Approach

The packet capture data was provided to the research team in the form of 739 .pcap[4] files, containing the network data from and to two hosts of the BestMixer service.

[4] .pcap is the file extension commonly used for network traffic captured by packet capture libraries such as libpcap or Npcap.

HTTP traffic. Manual analysis of at least 40 .pcap files revealed that the interactions between the users and BestMixer were programmed through HTTP requests and responses and that the procedure of mixing funds involved four steps.

First, a user sends a request with the currency and corresponding output addresses, fee percentages and delays to the server. If the user was referred to BestMixer through a referral link from a website that partnered with BestMixer, a "partner id" would also be sent to the service. During this step, it was also possible to fill in the aforementioned "BestMixer code", which effectively operated as a user id. Second, the server returns a temporary session id. Third, the user confirms the terms of service with this temporary id. Fourth, the server responds to this confirmation by creating an order id and a deposit address for the user. All this was provided to the user as a "letter of guarantee" that could be downloaded by the user. The order status – after following these four steps – reads "awaiting" a deposit of the user. After a deposit is made, the transaction follows different states: "unconfirmed", "pending", "sending" and "complete". Additionally, the administrators of BestMixer could cancel orders, resulting in the status "canceled".

From this HTTP network traffic, we extracted the following features of each user-BestMixer interaction using TShark [47]: IP addresses used during the mixing process, the 1 to 10 user-specified output addresses, the date and timestamp of the last recorded connection to the service, the deposit address BestMixer generated, the status of the transaction, the deposit amount, the currency, the BestMixer code, the order id, the service fee and the fee per address. This resulted in n=23,175 rows of data. For the orders that were captured multiple times, we combined their data and kept the timestamp of the latest interaction. This left us with 23,031 unique mixing orders.

MySQL traffic. In addition to HTTP, we observed MySQL connections in the network traffic. From these connections we were able to parse the "letters of guarantee" that BestMixer provided to its customers. These letters contained the generated deposit address and the payout distribution of the user-specified output addresses. Using TShark we were able to extract the following features: the deposit address generated by BestMixer, the date and timestamp of the "letter of guarantee", the deposit amount, the order id, the letter of guarantee and the 1 to 10 user-specified output addresses. This resulted in n=9,281 mixing orders.

Redis traffic. Finally, we could observe traffic related to redis storage [38]. In this traffic, interestingly enough, we saw user interactions with BestMixer with 127.0.0.1 (localhost) as source IP. This turned out to be order data from users that connected to and from the Tor service. From this traffic, we extracted the following features using TShark: the deposit address generated by BestMixer, the date and timestamp of

the last recorded connection to the service, the IP addresses used during the mixing process (127.0.0.1 for the Tor users), the deposit amount, the currency, the language settings of the website (set by the user), the BestMixer code, the user-agent, the order id, the letter of guarantee and the 1 to 10 user-specified output addresses. This resulted in n=2,126 mixing orders. The first user-BestMixer interactions in the wiretap date from mid-July 2018. The wiretap was intermittently active for different periods of time until May 2019. The exact dates data is available for can be seen in Table 1. HTTP traffic was available for 151 days total, MySQL traffic for 37 days and redis traffic for 7 days.

Table 1: **Data availability per traffic type**

HTTP	MySQL	redis
2018/07/18 - 2018/08/13	–	–
2018/11/12 - 2019/01/06	–	–
2019/02/07 - 2019/03/06	–	–
2019/03/21 - 2019/04/13	–	–
2019/05/07 - 2019/05/22	2019/4/15 - 2019/5/21	2019/5/15-2019/5/21

4.2 Validation

We validate the extracted data by performing an external and internal validation. For the internal validation, we first validated the correctness of the parsing through manually analyzing ∼20 randomly selected mixing orders in the wiretap data. For each order, we confirmed that its addresses in the wiretap data existed in the extracted mixing orders and that all other features were available and correctly parsed. We also compared the contents of the three different traffic source types (HTTP, MySQL and redis) based on the order ids when data from similar dates is available. For the 15 days overlap between the HTTP and MySQL data, 98.6% of the orders in the HTTP data are in the MySQL data and 75.12% of the orders vice versa. The overlap between the HTTP data and the redis data is 60% of the total available redis data. Most of the redis data is also captured in the MySQL data: during the periods data was available for both, 98.95% of redis orders was also found in the MySQL data. The differences between the datasets can be attributed to orders made via the hidden service: these orders are not captured in the HTTP but in the redis data, and the MySQL data captures data from both clear web users (HTTP) and hidden service users (redis). We explore these differences in Section 8. By combining the three datasources and removing duplicate orders based on deposit address, the number of unique orders totals 24,073.

For the external validation, we queried the blockchain for transactions to and from all deposit and output addresses. This ensured we only analyze the mixing transactions that actually took place. Our step-by-step approach was as follows. First, we identify if the deposit address existed on the blockchain. Second, we check if an incoming transaction took place (and only 1) within one day of the specified time of the mixing order on BestMixer (the time range specified in the FAQ). Third, for every output address of a mixing order we looked for outgoing transactions of BestMixer's wallet to the specified address and if it exists on the blockchain within a time range of four days (as 72 hours is the maximum delay that could be specified by BestMixer and an additional 12 hours to cope with network confirmation delays).

Other coins. In the data, we found 874 orders related coins other than bitcoin: litecoin (437), bitcoin cash (352) and bitcoin testnet (85). We removed the bitcoin testnet transactions because these were admin-made – this currency had never been offered to customers. After the first and second step, only 139 litecoin and 54 bitcoin cash transactions remained, mainly due to the absence of any incoming transactions. After step three, the amount of validated transactions even decreases to 45 litecoin and 8 bitcoin cash. The 45 orders of litecoin on average mixed 13.41 ltc, amounting to a mean order value of $931.01, while using on average 1.09 output addresses. For bitcoin cash, the average transacted amount is 0.52 bch or $148.69, with on average 1 output address. Because the number of transactions that involved cryptocurrency other than bitcoin is marginal, we decided to focus on bitcoin transactions for the remainder of this paper.

Bitcoin. Excluding the non-bitcoin transactions, we are left with 23,199 orders. We first excluded 2,723 orders based on the absence of their deposit address on the bitcoin blockchain. Second, we excluded 199 orders as no deposit was made in the 24 hours after the address was presented to the user. Third, we excluded 3,661 orders as we found more than one incoming transactions to the output addresses within four days of the transaction to the deposit address. This resulted in 15,574 orders for which both the deposit address and all the output addresses have been validated through the bitcoin blockchain.

4.3 External datasources

Besides our ground-truth data, we make use of four external datasources which we discuss below.

MaxMind databases. MaxMind is a commercial organization that bundles the information associated with all IP addresses and sells this information package publicly. We had access to commercial MaxMind GeoIP databases (GeoIP2-Anonymous-IP) that register the geolocation, ISP, organisation, connection type, and anonymous status of an IP address. Because we had access to 33 databases published between 2018-04-02 and 2019-06-03, we could perform historical look-ups on the IP addresses based on the transaction date. These look ups resulted in the complete IP attribution (geolocation, ISP, organisation, connection type, and anonymous status) for the 14,085 transactions that

had an IP address. The reliability of this database at the country-level as estimated by the provider is 99.8% [24], but external research has showed this to be an upper rather than lower bound. While there are no recent validations of the country-level accuracy of the commercial MaxMind GeoIP2 database, we point the reader to the works of Gharaibeh et al. [15] on the accuracy of various IP-to-geolocation databases when geolocating routers, and Schopman [40] on the city-level accuracy of the open-source MaxMind GeoIPLite database.

Bitcoin blockchain. The data from the wiretap was enriched with the data from the bitcoin blockchain since not all data relevant for user behavior analysis was captured directly from the wiretap. For instance it only included the deposit addresses whereas we are interested in the transaction hashes of the deposits. We did this by recreating the wallet using the ground-truth data in Bitcoin Core which we then fully synchronized with the bitcoin network for the latest information.

Coincap. To be able to express transaction volumes in US dollars, we scraped the historical exchange rates from the Coincap API [7]. These exchange rates were available with an increment of a minute, but did not provide an exchange rate for every timestamp we requested. For the timestamps that data was missing, we averaged the surrounding four exchange rates – two minutes before and two minutes after – to fill the missing data. Multiplying the deposit amount in bitcoin with the historic exchange rate provided us with the US dollar value of each transaction.

Chainalysis. Chainalysis [5] is a blockchain analytics company that enables users of their services to attribute addresses to entities or services (e.g., exchanges or online anonymous markets). The attribution is based on clustering: they create clusters of one or multiple addresses, for which they try to find the controlling entity (e.g., BestMixer). This entity is then categorized into one of their entity categories (e.g., "Mixing"). When no controlling entity can be found, the cluster is labeled as "unidentified cluster". A subset of these clusters is labeled "unnamed service", if Chainalysis – based on certain heuristics – suspects a collection of addresses to belong to a service. In this paper we use Chainalysis Reactor to obtain the clusters and transaction volume of all mixing services as attributed by Chainalysis in the time period that BestMixer was active. For each cluster, we used Chainalysis Reactor to retrieve the clusters that sent to or received from such a cluster and, in the case of an identified cluster, its entity category.

4.4 Data descriptives

The final dataset used for our analysis consists of the 15,574 validated transactions parsed from the HTTP, MySQL and redis traffic. The distribution of unique and validated transactions to their origins is 12,847 from HTTP, 554 from MySQL, 6 from redis and 2,167 from two or three sources. As we will explain in more detail in the ethics section in Section 10, before we could use this dataset for our analysis, the features that contained personally identifiable information were anonymized by Law Enforcement. In practice, it meant that we did not have access to the IP addresses, but rather a string representation by which we could assess whether two addresses were the same or different.

BestMixer allowed returning users to enter a user id to make sure that the mixer did not return bitcoins from a user's previous transaction, which it called the "BestMixer code". For the remainder of this paper we will refer to it as the user id to make it mixer independent. With this feature we could link transactions that were most likely carried out by the same user or by users who are part of a group sharing a user id. Using the user id, we constructed chains of consecutive transactions for returning users. Based on the amount of unique user ids, there were 8,450 users that used a user id in the 173 days for which we have data. There were 388 transactions for which there is no user id registered. If we assume all these transactions to be from individual users, there was a maximum of 8,838 users. This is an upper bound, since BestMixer allowed – and not forced – users who already carried out one or more mixing transactions via BestMixer to specify an id. As a result, the actual number of users can be lower due to transactions from the same user having different user ids (because the user had not specified its previously received id).

For all 15,574 transactions, we have data on the deposit address (15,574 unique), date and time (15,574 unique), the order id (15,574 unique), user id (8,838 unique), the deposit amount in bitcoin (μ=0.608102, σ=4.983137, `median`=0.049760, `min`=0.000011, `max`=156.511178) and dollar (μ=2,887.60, σ=24,790.57, `median`=225.24, `min`=0.04, `max`=1,001,963.16), the number of output addresses (μ=1.34, σ=0.99, `median`=1.00, `min`=1.00, `max`=10.00) and IP addresses (n=14,085, 8,482 unique).

An overview of the number of transactions across output addresses is presented in Figure 2. A majority of 12,644 transactions only included one output address and 1,813 transactions just include two output addresses. A few transactions were found with three (n=631) and four (n=225) outputs.

The empirical cumulative distribution function and (logged) histogram of the deposited amount in USD in Figure 3 show that while 99.3% of all transactions deposit an amount below $50,000 dollar, the highest 0.7% transacts anywhere between that and $1,000,000 dollar in a single transaction.

We will discuss the findings related to the number of output addresses as well as the transaction volume (deposit amount) in the following sections. Because the users of the hidden service already took a measure against identity attribution by using Tor and we have few datapoints (n =491) for this subset of users, we exclude them (408 users with in total 575

transactions) from our general analyses of scam, financial attribution and identity attribution mitigation. Instead, we compare the orders of the Tor users to clear web users for the week of data for which we have both in Section 8. For computing the total transacted volume in USD in Section 9, we will use the combined, deduplicated data from all sources, since this only relies on the date and deposit amount features.

5 Scam mitigation

This section investigates efforts to mitigate scamming by BestMixer. We map mixing intensity to see how much funds users entrusted BestMixer: if the average mixed amount can be considered 'small', perhaps users expected the service to be a scam. Next, we identify whether users will trial run the service with test transactions.

5.1 Mixing intensity

We started by analyzing the number of users who used the service of BestMixer and the volumes these users mixed via BestMixer. Given the total number of validated wiretapped transactions from users of the clear web instance of BestMixer (14,999), we found that each user on average performed 1.78 mixing transactions. For returning users, the average number of transactions is 4.28. We found a maximum of 194 transactions with one user id. A difference in transaction volume is present between users who only mixed bitcoins via BestMixer once and users who did so multiple times. One-time users (n=6,430) transacted an average of $2,096.11 per transaction (σ=18,193.46, `median`=178.56, `min`=0.09, `max`=578,933.61). Returning users (n=2,000) mixed on average an almost equal amount, $2,571.74 (σ=16,905.73, `median`=238.60, `min`=2.26, `max`=423,414.43). They mixed in total significantly more: $12,484.30 per transaction (σ=94,247.55, `median`=706.38, `min`=4.53, `max`=2,823,316.86) .

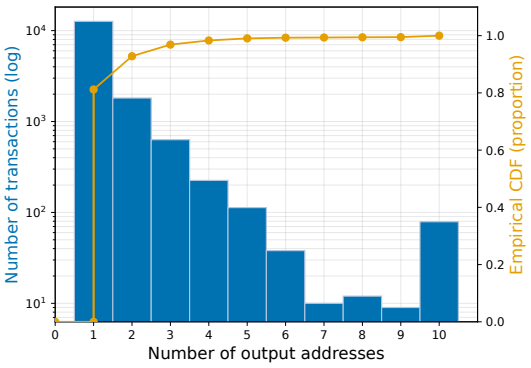

Figure 2: **Histogram (logged) and empirical cumulative distribution function of the number of output addresses used in that transaction.**

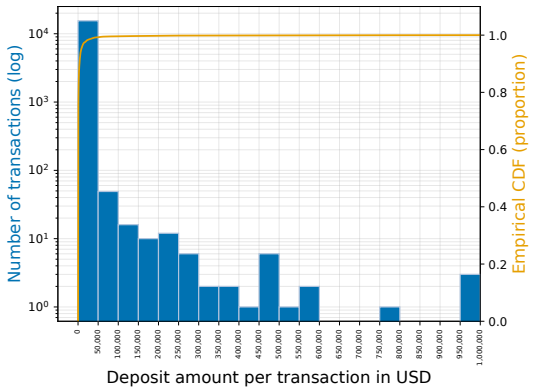

Figure 3: **Histogram (logged) and empirical cumulative distribution function of the deposited amount in USD.**

However, there is a large spread in the total transaction value in the user population. This can be seen in the violin-plot based on the Gaussian kernel density in Figure 4. The standard deviation and the maximum values show that the last 25% of values is causing long-tailed distributions for both one-time and returning users. Both distributions are also very right-skewed, with a mean that is a multiple of the median. Returning users do not seem to transact more on average, but simply do more transactions to generate a larger total transaction volume. This total transaction value is significantly larger than any single transaction made by one-time users. Although users are generally hesitant to trust new bitcoin mixing services [9], the users of BestMixer were willing to mix large volumes via the clear web instance in a single transaction.

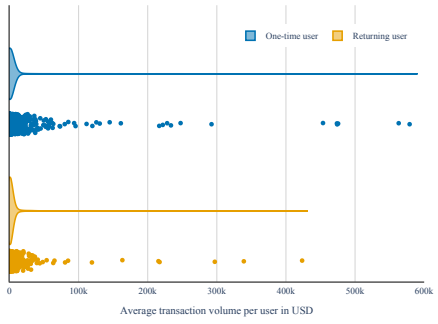

Figure 4: **Kernel density plot of average transaction value in USD of one-time and returning users.**

5.2 Testing the mixer

To investigate if users mitigated the risk of scam by performing a test transaction, we looked into the transactions of returning users (n=2,000); users that had multiple transactions

linked by a user id. Because BestMixer was active longer than the wiretap was active, we cannot say with certainty that the first observed transaction of a user id was the actual first transaction performed by that id. We calculated the average time period between the first and final transaction for each user: 24.87 days. To limit the effect of including users that possibly made a first transaction before the wiretap started, we excluded the users that made a transaction in those first 25 days. This decreased the set quite drastically to 875 returning users.

Next, we can analyze whether there are test transactions in the dataset. We first calculated the number of users of which the first transaction is the lowest transaction of all transactions. This is a conservative definition of a test transactions since it does not specify how much lower the first transaction needs to be. Of the total, we found 233 users that possibly started with a test transactions. We are, however, interested in whether this amount of returning users with their first transaction being the lowest can be explained by random probabilities, or whether this is an indication that our dataset contains a larger than expected amount of deviating first transactions.

For this, we use the Poisson binomial distribution. We view each returning user to be a trial, in which the p is the random chance that the first transaction it made is the lowest. This means that when a user has 8 transactions, the chance of the first transaction randomly being the smallest transaction is $\frac{1}{8}$. With 875 users, there are 875 trials with the probability of $1/\texttt{numTransactions}$ of success. To compute the right-tailed probability of the number of occurrences of test transactions being greater than or equal to 233 based on random probabilities, we used the Python module `poibin` [45]. Using this implementation based on the characteristic function (CF) of the Poisson binomial distribution using discrete Fourier transforms [18], we calculated that $\Pr(X \geq 233) = 0.999 > 0.05 = \alpha$. This means that based on our observation of 233 users that have a first transaction that is the smallest, we cannot reject the null hypothesis that the probabilities of the first transaction being the smallest are smaller or equal than the random probabilities. In order words, by observing just 233 users, we cannot claim that the probabilities for observing the first transaction being the smallest are in any way different from the probability that any of the transaction a user made is the smallest.

As users who aim to mix large amounts have more reason to test the service, we separately analyzed the group of users who mixed more than the average total transaction value for returning customers ($12,484.30). Of those 177 users, 37 (20.90%) had a first transaction that was the smallest. When we repeat the same analysis based on the random probabilities of the first transaction being the smallest, we find that $\Pr(X \geq 37) = 0.917 > 0.05 = \alpha$. This again means that our observation of these users can be explained by random probabilities of a first transactions being the smallest. Although there is no proof for a general trend of test transactions, we do believe that it is very likely that there are examples in the dataset. Of these 37 users, the first transactions differed on average $14,109.44 or 36.90% from the mean of their following transaction(s). Additionally, it could be that users performed a test transaction with a different user id. We believe this to be unlikely, since transactions performed by one-time users were not significantly lower on average than those of returning users (see 5.1). These findings, therefore, do not prove that users commonly perform a test transaction encompassing a strategy to mitigate the risk of scam.

6 Financial attribution mitigation

This section investigates user efforts to mitigate the risk of financial attribution. We look at strategies that mitigate the risk of incriminating, financial information becoming available due to for example a breach, confiscation by LE or a misconfiguration of BestMixer. We look at three features that a user can directly influence: the number of outputs, the delay per output and the distribution of amounts between outputs.

6.1 Outputs

First, we interpret the extent to which users spread the output of mixing transactions over different output addresses. When the output address to which a mixing service returns the mixed bitcoins remains the same for all mixing transactions, the user's transactions will be easier to trace [36]. Besides sending the mixed funds to different wallets based on operational needs, specifying multiple output addresses is mostly related to a conscious decision to improve the quality and protection a mixing service offers. Therefore, we state that if users specify multiple output addresses for one or more transactions, this indicates that the user implements additional security to increase the chance of a successful mixing process – i.e., foolproof obfuscation of their funds.

Based on Figure 2 we know that most users used only one output. Again, we compared the behavior of one-time and returning users and found they used, on average, 1.29 and 1.26 output addresses respectively. We show their distributions in Figure 5 and Figure 6. The T-test showed no significant difference between the number of output addresses one-time and returning users specified (T-statistic = -1.494, p-value = 0.135). For the returning users we could also investigate how many unique output addresses were used in all transactions, which turned out to be 5.80 addresses. These findings do not provide conclusive evidence for the use of multiple output addresses to mitigate the risk of financial attribution.

Our findings do show a significant positive relationship between the transaction value and the number of output addresses used in a single transaction (Spearman's ρ= 0.231, p-value = 0.00). This relationship implies that users who transact relatively large funds are more eager to spread their mixed bitcoins over different output addresses. This could

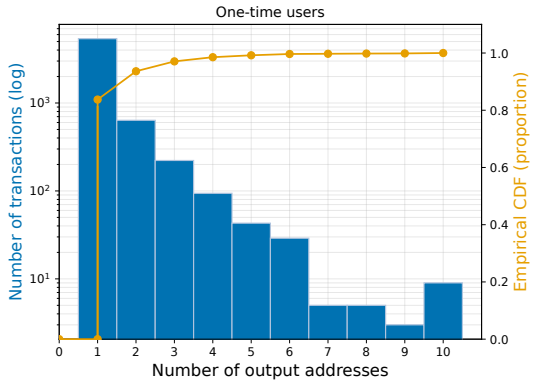

Figure 5: **Histogram (logged) and empirical cumulative distribution function of the number of output addresses used by one-time users.**

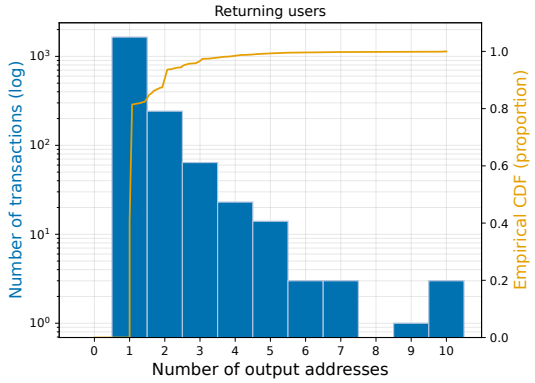

Figure 6: **Histogram (logged) and empirical cumulative distribution function of the average number of output addresses used by returning users.**

indicate that users are in fact careful with higher transaction volumes. In other words, this relationship shows that users who want to mix higher volumes are more likely to mitigate the risk of financial attribution.

6.2 Delays and distributions

Given the apparent strong bias towards ease of use, we also analyzed the specified delay per output and the distribution of amounts between outputs. Based on these features we analyzed how users of multi-output transactions used these inputs (2,930 transactions in total). Because 99% of all users relied on 5 output addresses per transaction or less, we include the transactions having 2 to 5 different output addresses (see Figure 7).

It turns out that users prefer to get most of their money out early, as can be seen that the first quartile of the percentage boxplots of the later outputs tends to get lower. Additionally, the maximum time delay in seconds that we observe is around 100,000, which turned out to be far shorter than the advertised delay of at most 72 hours (i.e., 273,600 seconds). We note that these timestamps were taken from the wallet itself and therefore represent the time the transaction was created and (most likely) published to the network. We therefore assume that the mixing service, in order for it to ensure that users get their money back within the specified delay ensured that they send it out earlier such that even with a highly populated mempool the users still get their money back in time. Based on this we conclude that users do have the tendency, no matter how many outputs they use, to get their money back relatively quickly. This shows that users rather opt for a strategy wherein speed is key, than a strategy where delays and spreading over multiple outputs decrease the chances of financial attribution.

7 Identity attribution mitigation

This section analyzes user strategies to mitigate the risk of identity attribution. We first observe the proxy usage at the transaction level. Then we shift our focus to proxy usage at the user level and pinpoint the persistence of proxy usage strategies for returning users. In this section, we include the transactions from users of the hidden service in 7.1, but leave them out of our analysis in 7.2 and 7.3, so we can analyze their unique characteristics in Section 8.

7.1 Proxy usage at transaction level

For all transactions we retrieved the features *connection type*, *anonymous status*, and the *organization* based on the IP address and transaction date from MaxMind [24] (see Section 4.3 for a description of the MaxMind database).

We label transactions as anonymous or non-anonymous based on the following reasoning. If the anonymous status MaxMind attributes equals `unknown` and connection type is `cable/DSL` (4,410 transactions), `cellular` (1,004 transactions), `unknown` (53) or `dial_up` (12), we classified the transaction as not-anonymous. These connection types, in general, mean that it is possible to identify the owner of the device or endpoint the transaction was made with, because these are not corporate connections. There is often a subscription – e.g. with an ISP or telecom provider – that links the IP address to an entity. It differs per legal framework how long these records are kept and in which circumstances of probable cause that data can be subpoenaed. This is for example governed by the U.S.C. Title 18 §2703 "Required disclosure of customer communications or records" [16] in the United States, which can also be used for retrieving information from European telecom providers if a mutual legal assistance treaty (MLAT) between the US and that country exists. European examples are “Section 100j - Subscriber data request” of the German Code of Criminal Procedure [12] and Article 76A of the Austrian Code of Criminal Procedure [13]. Our rationale here is that LE agencies are able to subpoena these IP addresses to

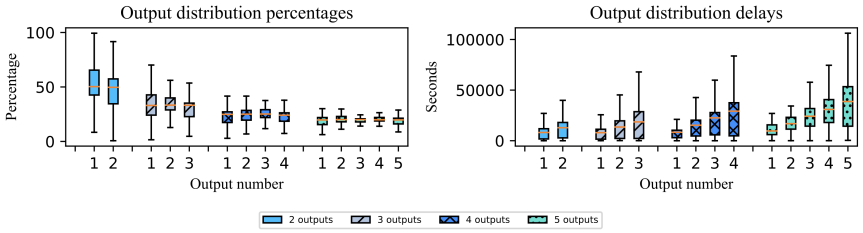

Figure 7: **Effects of the usage of security features: spreading money over multiple addresses and delaying payouts.**

attribute it to an identity. Whether that identity matches the user making the mixing transaction depends of course on the (additional) security strategies applied.

All other values of the anonymous status that MaxMind provides – such as `hosting provider`, `public proxy`, `Tor exit node` and `anonymous VPN` – we classify as anonymous, regardless of their connection type. The logic being that these IP addresses are most likely not related to a private identity, if they are related to an identifiable identity at all. Additionally, we classify all transactions made via the Tor instance of BestMixer as anonymous. Following this classification, we find 8,606 transactions to be anonymous – 61.10% of all transactions with an IP address (n=14,085). This means that 5,479 transactions (38.90% of all transactions) were performed via the clear web instance without the usage of a proxy.

If we look at the transactions without IP obfuscation, we see that $\approx$ 75% originates from the following ten countries: Germany (1,241), United States (1,137), United Kingdom (481), Poland (346), Russia (241), Ukraine (227), Australia (175), Canada (131), France (119) and The Netherlands (96). As most these are either Commonwealth or European countries, LE is most likely allowed to subpoena the subscriber information for these transactions.

To gain insight into the exact strategy users applied, we look at the share of the anonymous IP addresses that belonged to either the category `VPN`, `hosting provider/VPS`, or `Tor`. To this end, we classify the potential *Anonymous Status* values into these three categories according to the mapping in Table 2). For each IP address that originates from a hosting provider, we assume it is a rented server/VPS unless a VPN-service, Tor node or public proxy was explicitly named. Our classification demonstrates that of all anonymous transactions, 5,822 transactions (67.65%) were performed through a hosting provider/VPS, 826 transactions (9.60%) were performed with a VPN, and 23 transactions (0.27%) originated from a Tor exit node. Additionally, there were 491 transactions (5.71%) performed through the hidden service of BestMixer (see their separate analysis in Section 8).

Table 2: **Classification of values for MaxMind's anonymous status into VPN, Hosting provider/VPS or Tor.**

Values for Anonymous Status	Classified anonymous status	n
Hosting Provider	Hosting provider/VPS	5,822
Anonymous VPN	VPN	70
Anonymous VPN, Hosting Provider	VPN	619
Anonymous VPN, Hosting Provider, Public proxy	VPN	23
Anonymous VPN, Public proxy	VPN	2
Public proxy	VPN	2
Hosting Provider, Public proxy	VPN	110
Hosting Provider, Tor Exit Node	Tor	1
Anonymous VPN, Hosting Provider, Tor Exit Node	Tor	18
Hosting Provider, Tor Exit Node	Tor	4
Unknown	Unknown	1,444
Tor	Hidden service	491
Total amount of anonymous transactions:		**8,606**

It becomes clear that using a server at a hosting provider/VPS as a proxy is the most popular obfuscation technique. We have to note that MaxMind's classification into a VPN 'exit point' or 'routing point' is not perfect, especially for the smaller VPNs. Therefore, it could be that a part of the IPs classified as a VPS, actually belong to the category VPN but MaxMind failed to recognize this properly. Additionally, servers at a hosting provider could also have been compromised machines rather than rented machines.

7.2 Proxy usage consistency

Of the 14,999 transactions made by clear web users, we were able to find information on the IP address for 13,547 transactions. These transactions are associated with 7,583 unique user ids. From the subset of users that made multiple transactions, 949 users made at least one transaction with a non-anonymous IP address (51.72% of the returning users). Next, 1,048 returning customers made at least one transaction through a proxy (57.11% of the returning users). The latter two findings are not mutually exclusive, because it could be the case a user made two transactions: an anonymous and a non-anonymous transaction. Of the 1,835 users who performed multiple transactions, 886 users (48.28%) obfuscated their IP address for every transaction. The most consistent user made sure to obfuscate every IP address of 50 transactions in total. When we look at the returning users who performed at least one anonymous transaction, we observe that on average 57.40% of all used IP addresses is obfuscated. These findings show that users on average were not consistent with their IP address obfuscation.

We also found patterns that look like mistakes of users – indicating sloppiness or a lack of consistency in obfuscation strategies. First, we found 149 users (8.08% of all return-

ing users) who use an obfuscated IP address for their first transaction, but later make one or more transaction(s) with a not-obfuscated IP address. Second, we find 139 users (7.54% of all returning users) who start with a transaction without IP obfuscation, but obfuscate the IP address for one or more transaction(s) later on. Third, we could find 174 users (9.44% of all returning users) who used an obfuscated IP address for each transaction except for one.

7.3 IP address reuse

The reuse of an IP address by a user does not seem to be related to a user's efforts to obfuscate an IP address. We performed a Spearman ρ correlation test between the percentage of obfuscated IP addresses and the percentage of unique IP addresses for each user with multiple transactions. We found a non-significant correlation of 0.001 (p = 0.898), which indicates no correlation exists. In addition, the number of transactions a user made was not correlated with the percentage of unique IP addresses (Spearman's ρ=0.038, p=0.001). The percentage of obfuscated IP addresses, however, appeared to be correlated with the amount of transactions a user made (Spearman's ρ=-0.687, p=0.001). Unsurprisingly, but with potential grave implications, this finding implies that users are more likely to reuse IP addresses when they execute a larger number of transactions.

8 Comparison with hidden service users

To see if users of the hidden service – who made the conscious choice to access the service of BestMixer via the Tor browser – include additional mitigation strategies, this section aims to identify the differences between the transactions performed via the clear web instance and the hidden service of BestMixer. To perform this comparative analysis, we utilize the week from the 15th of May 2019 up and until the 21nd of May 2019, in which data is available for both clear web and hidden service users. Since the original client IP addresses of the transactions performed via the hidden service remain unknown, we cannot say anything about the use of IP obfuscation of these transactions. In addition, the time frame of available data is too short to look at returning users. Therefore, we focus on testing for differences in mixing intensity and the use of multiple output addresses.

8.1 Differences in mixing intensity

From all transactions captured in these seven days (1,266), 491 transactions were executed via the hidden service. On average, the hidden service was used for ∼70 transaction per day (ranging from 53 to 79), while there were ∼111 transactions from the clear web service (ranging from 83 to 138). The hidden service was used by 408 different users. Paradoxically, of these users, 31 had also used the clear web instance with the same user identifier. Despite that the daily average number of transactions made via the hidden service was much lower than the number of transactions performed via the clear web, we uncover that the total volume mixed via the hidden service is significantly higher. When we compare the volume of the transactions initiated via the clear web instance and the hidden service, we find that $2,988,828 was mixed in transactions performed via the clear web, whilst more than $6,115,522 – so more than double – was mixed via the hidden service. This shows that users of the hidden service trusted BestMixer with significantly higher transaction volumes (T-statistic=1.96, p=0.05). This was not due to some outliers: the average transaction volume was three times higher for the hidden service versus clear web transactions. An average transaction via the hidden service mixed $12,455.24 while an average transaction via the clear web service mixed $3,856.55.

A potential explanation for these relatively high transaction volumes may well be that users who aim to mix larger transaction volumes feel a greater urgency to mitigate the risk of identity attribution and access the hidden service to do so. Another potential explanation would be that users who accessed the hidden service are more comfortable with mixing higher volumes as they have mitigated the risk of identity attribution. Ironically, the use of the hidden service does not affect the risks of scam or financial attribution. In that line of thought, these findings might demonstrate that users already feel more confident with performing mixing transactions when having mitigated the risk of identity attribution.

8.2 Differences in number of output addresses

To see if users interacting with BestMixer through their hidden service also take steps to prevent financial attribution, we look into the number of output addresses they specify when mixing. The transactions performed via the hidden service of BestMixer included significantly more output addresses than the transactions initiated via the hidden service (T-statistic=3.920, p=0.00). On average, a user that accessed BestMixer via the hidden service sent bitcoins to 1.66 output addresses in a single transaction. The group of users who accessed the service via the clear web only used 1.38 output addresses per transaction in these seven days. The specification of a higher number of output addresses by hidden service users indicates that a hidden service user is more likely to mitigate the risk of financial attribution.

9 Economics

In the previous sections, we described the characteristics and risk mitigation strategies of the users of BestMixer. In this section, we examine BestMixer itself, though a delineation of its role as a service provider in the mixing ecosystem. First, we show the transaction volume of BestMixer based

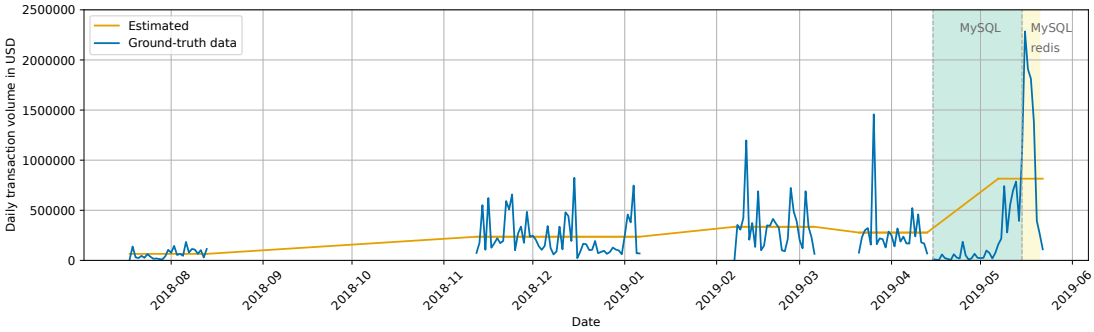

Figure 8: **Observed daily transaction volumes of BestMixer along with the estimated transaction volume over time ($).**

on our ground-truth data and a conservative extrapolation. Second, we utilize Chainalysis data to plot the market shares of BestMixer and other centralized mixers active during its lifetime. Third, we analyze Chainalysis' entity category labels of the clusters of direct counterparts of the deposit and output addresses. The ground-truth data used in this section consists of the combined, deduplicated and validated transactions from HTTP, MySQL and redis data.

9.1 Transaction volume

Collectively, the user base of BestMixer on average mixed US $259,950.94 per day in ∼90 transactions in the periods that ground-truth data was available. The range of the total transaction value that was mixed on a daily basis ranged from the 25th percentile $71,593.51 to the 75th percentile $323,803.29 — with outliers up to $2,284,533.56 per day. To get an idea of the total amount of money that was mixed between July 2018 and May 2019, the ground-truth data set was extrapolated to approximate the daily transacted value for when no data was available. We took the average daily transaction amount for the periods data was available and assumed a linear increase between time periods. By doing so, we found that if data would have been collected consistently, a transaction volume of around 26,000 bitcoins would have been mixed. This equals to approximately $192,000,000. The total transaction value per day from the ground-truth data and the estimation is visualized in Figure 8. The two highlighted areas indicate at which time periods the MySQL and redis data sources were available (see Table 1 for the exact dates).

9.2 Market share

For calculating BestMixer's market share, we looked at the transaction volume of all mixing services attributed by Chainalysis in the period of March 2018 until May 2019 – the time BestMixer was operational. Chainalysis identifies 21 centralized mixers as active in BestMixer's lifetime [5]. For each mixer, we queried its daily deposits in USD (the price at the time the deposit was made) to calculate monthly transaction volumes. Next, we calculated the total transaction volume and each mixer's monthly market share. We plot the market share for nine of the biggest mixers and group the remaining mixers in the category *Other* in Figure 9. It shows that BestMixer grew to be a top-3 player and even was market leader for a brief period of time around October 2018. In general, it seems as though the mixing market was quite competitive, with seven different mixers obtaining more than 20% of market share during one month of this time period.

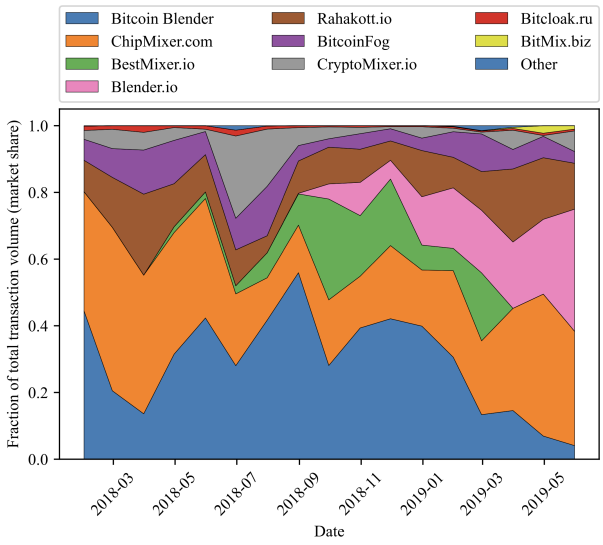

Figure 9: **Market share per centralized mixer.**

[5] BitcoinFog, Bitmixer.io, Helix Mixer, Bitcoin Blender, Chipmixer.com, Blender.io, Cryptomixer.io, Bitmix.biz, Rahakott.io, Sinbad.io, Jambler.io, Bitcoin-laundry.com, Bitcloak.ru, Coinomize, Coinmixer.se, FoxMixer.com, Privcoin.io, Bitcoin-mixer.com, Coinmix.to, Brave Bunny Mixer and Bitsmix.biz

To see whether there was an economic reason for the popularity of a certain mixer, we gathered the (observed) mixing fees of the six largest mixers from open source data – mostly from published third-party reviews or affidavits. Most mixers offered a range of fees, based on transaction size, type of liquidity pool used or number of output addresses: 1%-3% for Bitcoin Blender [39], ∼2.44% for ChipMixer.com [30], 0.5%-2.5% +0.0005 BTC/address for Blender.io [11], 1.00%-2.195% + 0.0005 BTC/address for Rahakott.io [23], ∼2.32% for BitcoinFog [29] and 0.5% +0.0005 BTC/address for CryptoMixer.io [22]. If we compare them to BestMixer's fees, that ranged between 0.5% and 3.5%, we see no apparent differences.

9.3 Attribution of interacting clusters

Based on the cluster of BestMixer that Chainalysis identifies, we will show the attribution of deposit and output addresses to direct counterparts. For each deposit address, we retrieve the entity category for the cluster(s) that deposited bitcoin. Since the output address is specified by the user, we obtain the entity categories for all the output addresses. We divide these entity categories into *high risk cluster* and *low risk cluster*, with the former being one of the following: `high risk exchange`, `gambling`, `mixing`, `scam`, `sanctions`, `darknet market`, `stolen funds`, `special measures`, `fraud shop`, `ransomware`, `high risk jurisdiction`, `child abuse material`, `illicit actor-org` or `terrorist financing`. We plot this in Figure 10.

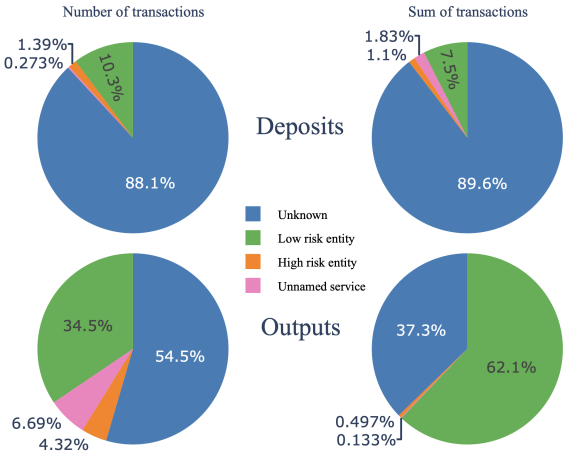

Figure 10: **Chainalysis labels of clusters that made deposits to or received output transactions from the BestMixer cluster, based on number of transactions or sum of transactions.**

The total number of clusters that made a deposit to the BestMixer cluster is n=27,472, with a total sum of $74,492,727.14. Only 11.7% of transactions came from an identified and categorized cluster, with 1.39% of transactions originating from a high risk entity. This pattern is almost similar for the sum of deposited transactions. For the outgoing transactions, however, the percentage of attributed clusters increases to 38.81% of transactions or 62.57% of transaction volume. This could indicate that users mainly specified addresses from low risk entities, such as known low risk exchanges, as their output addresses. In total, 78,113 outgoing transactions, totaling to a sum of $74,404,453.06, were identified by Chainalysis.

10 Discussion

In this section, we discuss the ethics of using seized data, touch upon inherent limitations that arise from our methodology and the data sources we used, embed our findings into the broader cryptocurrency security ecosystem by providing recommendations for security practitioners, and share lessons learned from partnering with LE. First however, we discuss possible explanations of our findings.

Possible explanations for our findings. We analyzed mixing transaction data and uncovered that the user base of BestMixer entrusts the service with large funds, whilst putting limited effort in mitigating the risks of mixing. This trust could be explained from BestMixer's reputation arising from its affiliation with DeepDotWeb, its active presence on BitcoinTalk.org and its communication on its 'mixing security features'. However, its mixing fees, transaction volumes and that it advertised on darknet forums do not significantly differ from other mixers active at that time. BestMixer neither seems to have been a smaller player with a well-known underground customer base, nor a large firm that strove to obtain a mainstream, legitimate reputation. This leads us to believe that this willingness to trust BestMixer is not inherently related to BestMixer itself, but rather the result of the information asymmetry between users and service providers [20]. While service providers know their own data retention policies and implementations, as well as their mixing algorithm and security features, users do not have access to this information, nor can they validate any of the security or privacy claims made. As a result, users 'just' need to entrust mixing services like BestMixer in that the service properly returns their bitcoins and processes their transaction data securely.

An alternative explanation is that there is a dominant type of user that does not simply trust mixing services, but accepts the risk that mixing services bring along. Criminals using traditional cash-out techniques are willing to accept a significant loss of their criminal proceeds – the illicit funds they initially aim to cash-out. The fee BestMixer asks users to pay for a mixing transaction is somewhere between 0.5% and 3.5% of the transaction volume. In that sense, mixing services like BestMixer provide plenty of room to take risks and potentially lose some funds in the process. For users who possess large sums of bitcoins and are rather desperate to anonymize these in a short time span, it may be worth to take the risk.

This explanation, however, does not cover the trust users seem to have with respect to the transactional data they provide to BestMixer when making a transaction. The higher the transaction and the less effort a user does to obfuscate, the more the resulting privacy of the user is dependent on how the transactional data is safeguarded. Although users may accept the risk of losing money, they still face the risk of attribution. Here, an explanation may be that users trust the mixer's guarantee of generating anonymous transactions more than the proper mixing and return of bitcoins.

Ethics. In line with applicable laws and regulations, Dutch authorities were able to first wiretap the web server of BestMixer and later seize their infrastructure. Regardless, using this legally seized data for research purposes raises some ethical issues, which we discuss below. While we use data from a legal seizure, one should not assume that users were engaged in illegal behavior or that this was a factor in deciding to use this data for our research. Note that providing evidence of any kind for continued law enforcement efforts is not the purpose of this study.

Before back-end data was made accessible to us for academic research purposes, public prosecutors weighed, among other things, the impact of the work on the rights and privacy of all (involved and third) parties. A Dutch law enforcement privacy officer vetted that our data subset was limited, only contained data vital to our research and contained no personally identifiable information. Similar to earlier work presented at USENIX Security in recent years [10, 21, 28], all of our analyses were conducted on-site at Dutch law enforcement agencies, where the data was stored and protected under their safety and security guidelines. We conferred with our IRB beforehand and they viewed this work as outside of their jurisdiction, yet were satisfied with the assessments and applied procedures outlined above stemming from the public prosecutors and law enforcement privacy officer. In order to protect the privacy of BestMixer users, we took great care not to analyze personally identifiable information (PII) – i.e., the data was stripped of all PII. This anonimization process was initiated by the involved privacy officer following strict regulations which go beyond GDPR or IRB institutional frameworks, implemented by law enforcement and as a result we only had access to a set of strings/hashes representing IPs. When our analysis did involve PII we asked law enforcement to run our code and return the output.

With this approach, the data was cleared by law enforcement authorities for the purpose of this research in accordance with Dutch privacy law. We believe that our analysis does not create further harm as we did not partake in or stimulate any criminal business model – by using criminal services, or in any other way contribute to its ecosystem. The authors and involved law enforcement professionals believe the benefits of a comprehensive understanding of bitcoin mixing, outweigh the potential cost of making this kind of knowledge more widely known. More so, as the anatomy and economics of bitcoin mixing services are already well-documented in earlier work [26,43,46].

Limitations. First, our research focuses on a single bitcoin mixing service: BestMixer. Naturally, this is a limiting factor in our ability to generalize our findings. After the takedown of BestMixer, other measures such as regulatory interventions, sanctions and law enforcement takedowns might have resulted in an increased awareness of financial attribution risks among those mixing bitcoins via online mixing services, as well as an increased security-awareness among those providing a mixing services in the underground economy. This would mean that we observe a time frame wherein users were perhaps operating less securely, compared to today.

However, when we look at current centralized mixing services as described in Section 2.3, we see that they offer the same features that we have studied: they allow the user to specify multiple output addresses, set payout delays and sometimes choose a specific pool to mix money with. These mixers also have the same usage purpose that BestMixer had, are centralized services that might store data, and provide their services on both the clear web or as a hidden service. Because we measure the responses of users to mixer features and these features have hardly changed, we expect our insights still to be relevant today. Future research should try to replicate our analysis and see if security practices involving bitcoin mixing have evolved.

Second, as there were periods that the wiretap was offline, parts of our analyses are hampered by missing data. The transaction volumes and values we describe thus represent the lower-bound of the business that was conducted via the service. So, both the number of returning users and the number of times they visited the service are expected to be higher. Next, the available time period with data from both the clear web and hidden service that BestMixer operated, only covered transactions over a week. This restricts generalizing the differences we found between the transactions performed via the clear web and the hidden service as well as to make claims concerning returning users of the hidden service.

Finally, our research is focused on a centralized mixer. Recently, decentralized mixing services have grown more popular, overcoming the limitations of centralized mixing services. Decentralized mixers are peer-to-peer services that are available on advanced blockchain platforms. Although centralized and decentralized mixing services apply different processes to mix bitcoins, they generate the same result. As such, the users of both types of mixing services will pursue equivalent goals. Therefore, we can reasonably presume that the users who approach centralized mixing services do not fundamentally differ from those who approach decentralized mixing services. On this account, the results on user patterns found in this study will remain applicable when a shift towards more decentralized mixing services happens.

Recommendations for the cryptocurrency security ecosystem. Our findings have implications for the broader cryptocurrency security ecosystem. While the risks of financial attribution when using for example bitcoin are relatively well-known, the risks of identity attribution are considered less often. The currently available options for enhancing privacy effectively, surrender users to unregulated and often not-compliant service providers such as cryptocurrency mixers. We believe that more transparency – e.g., in the form of audits or certifications – could benefit users immensely in assessing the security of providers. Such audits or assessments can either be performed by the services themselves through a form of self-regulation, or initiated by security practitioners externally. Moreover, security practitioners could turn to creating or vetting open-source tools that offer privacy-enhancing features, that are not dependent on the practices of a centralized entity.

We showed that even though centralized mixing services offer more secure options such as multiple output addresses and ways to hide your IP address – e.g., by offering a hidden service on Tor – a large part of the user base of BestMixer.io did not use appropriate mitigation strategies. In the light of more decentralized forms of mixing such as CoinJoin, other users' lack of mitigation strategies becomes even more of a problem as their mistakes greatly influence other users that are in the same round, decreasing the anonymity set. Because of this, we propose security practitioners to take the behavior of users into account when either evaluating or designing tools and services that aim to enhance the privacy of users making cryptocurrency transactions. Additionally, we believe that our results show the importance of a defense-in-depth strategy, where security practitioners offer their users multiple layers of measures to safeguard their security and privacy, instead of relying on one service (such as a mixer) to do so.

Finally, this research is a first exploration of the risks that users take when using a bitcoin mixing service. We believe future research is needed to for example evaluate de-mixing risks, to provide further insights into the attribution of cryptocurrency addresses to entities and to provide security practitioners with best-practices when aiming for more privacy in the cryptocurrency ecosystem.

Lessons learned for future LE collaboration. There are three take-aways we share here, in the hopes of encouraging others to pursue a collaboration with law enforcement for this type of research. First, we want to stress that such collaborations rely on strong individual partnerships and that building those takes time. Procedures relating to the ethical considerations and obtaining the correct permissions cannot be rushed, so it is wise to allocate time and expectations accordingly. Second, because the data is not collected by the researchers themselves, we advise to always start with a broad exploratory analysis to get accustomed to the data. Additionally, since the data is collected in an adversarial setting, it is of great importance to invest ample time and energy into the internal and external validation of the data. Third, such a collaboration is beneficial for both parties, since law enforcement has operational intelligence but often no time to perform high-level analyses and researchers lack access to back-end data but bring the academic rigor to make such research possible.

11 Related work

Our paper builds on and benefits from recent advancements into three topics. First, our work relates to other bitcoin mixing research. Second, we can identify similar analyses compared to our investigation of criminal services using ground-truth data. Third and last, we contribute to the research body on trust in the underground economy. In this section, we discuss related work on these three topics.

Bitcoin Mixing. Our work benefited from previous studies that analyzed the structure and effectiveness of bitcoin mixing services. An early study on bitcoin mixing services was performed by Moser et al. [26], who tested the inner workings of three bitcoin mixers that were available in the underground economy back then. By reverse-engineering the mixers, they were able to understand how to link input and output transactions that were mixed in one of the three mixers. By doing so, their study provides a first overview of the inner workings of bitcoin mixers. Their work also states that mixing services have implications for law enforcement in the context of anti-money laundering by demonstrating that it is unlikely that mixing services apply a successful know-your-customer (KYC) structure.

Several heuristics and methods that aim to break the anonymity that mixing services claim to provide, have been introduced and subsequently evaluated. The work of Tironsakkul et al. [43] continues in this line of work by putting forward a novel method that is focused on tainting at the address level rather than at transaction level. In their study, they applied this address taint analysis to investigate the possibility to track mixed bitcoins from the deposited bitcoins. Their results suggests that the taint analysis method is the first method that enables the linking of deposited bitcoins with mixed bitcoins.

At the same time, new approaches to enhance the power of mixing services to anonymize bitcoin transactions appear. Recently, Wu et al. [48] performed a study to understand and illustrate state-of-the-art bitcoin mixing services, dividing their strategies into swapping and obfuscating mechanisms. Their study provides a method that is able to identify 92% of the mixing transactions that were mixed with the obfuscating mechanism. Having these mixing transactions identified, the study provides an estimation of the profit of mixing services – indicating the monetary impact of state-of-the art mixing services.

Analyzing criminal services with ground-truth data. Similar to our work, the work of Noroozian et al. [28] presents an unique empirical study using ground-truth data revealing an inside perspective of a seized criminal service. Ground-truth data on a criminal service was also used in the work of Van de Laarschot & Van Wegberg [21]. In their study, the authors investigate the security practices of vendors on Hansa Market and measure the prevalence and patterns of poor security across the vendor population. Recently, Aliapoulios et al. [1] presented the first empirical study of ground-truth data containing transactions from an underground shop selling stolen credit/debit cards. In contrast to that work, we use ground-truth data to uncover the actions of users of a service rather than to learn more about the actions of the service provider itself.

Trust in the underground economy. Our work builds on studies defining the development and presence of trust in the underground economy. Most of this work on trust focuses on transactions via marketplaces. For instance, the study of Holt et al. [17] focused on identifying signals vendors use to convince buyers of their trustworthiness via their advertisements for stolen data on Russian and English language web forums. The analysis was motivated by the idea that the information asymmetry between vendors and buyers at online marketplaces forces vendors to put effort into creating such signals - regardless of whether they are actually trustworthy. Our work takes a similar approach by starting from the existence of this information asymmetry between the provider of a mixing service and its users.

Other work on the trust in vendors is the study of Norbutas et al. [27], in which the authors argue that the development of the required buyers' trust is primarily explained by the sociological concept "dyadic embeddedness" rather than on the reputation of the seller according to others in the network. This finding implies that the experiences in previous exchanges determine development of trust. A buyer is not only unlikely to start or complete an exchange after negative experiences, but also unlikely to engage in new transactions with different sellers on the same market. The authors specify that this effect is especially present in the case of first-time buyers. Although our work analyzes mixing transactions performed by users of a single service, both studies assume that there is some need for trust in vendors or service providers for the completion of transactions, and try to better understand how this trust is developed and nurtured.

Finally, we contribute to the body of work identifying reasons to distrust bitcoin mixing services. For instance, the study of Meiklejohn et al. [25] proves that mixing transactions can be linked to addresses. Biryukov et al. [3] show how mixing transactions can also be linked to IP addresses, which forms a significant step in attributing ones identity. The work of Van Wegberg et al. [46] reveals that, although some services provide an excellent, professional and well-reviewed service at competitive cost, others turned out to be scams, accepting bitcoin but sending nothing in return. Likewise, Pakki [36] argues that bitcoin mixers are continuously accused of scams, lack implementation of academically proposed techniques, and display poor resistance to common mixer-related threats. He concludes that mixing services focus on presenting users with a false sense of control to gain their trust, rather than building truly secure mixing technologies.

12 Conclusion

In this paper, we looked at the presence of strategies to mitigate the risk of financial attribution, identity attribution or scam in ground-truth transaction data of the BestMixer service. Although the risks involved when using BestMixer were significant, our findings barely provide any evidence for the adoption of mitigation strategies by users.

We first analyzed mixing intensity and the presence of test transactions. We found that users of BestMixer entrusted the service with, on average, a high transaction volume ($2,888). At the same time, we did not find any evidence for test transactions. Therefore, the effort clear web users put into mitigating any risks of scam seems to be limited. Next, we looked at the presence of strategies to mitigate the risk of financial attribution. Most of the transactions were not associated with multiple output addresses. Additionally, users could set different delay timings for their transactions, as well as change the payout distribution over their output addresses. We found that very few users utilized these options, and those who did, often chose the same quick-and-dirty pattern: little to no delay, only one output address or, in the case of two outputs, the largest payout to the first address. In short, we could not find evidence for strategies to mitigate the risk of financial attribution. Third, we looked at the attempts users made to hide their identity. The results show that users turned to a VPN, accessed the hidden service or connected to the service with different IP address for different transactions. However, still a great part of users did not make use of any of these obfuscation strategies. Around 52% of returning users made at least one transaction with a non-anonymous IP address, leaving them vulnerable to de-anonymization of all transactions through that one transaction. This is remarkable because these strategies are low in effort and are very effective for the mitigation of the risk of identity attribution.

Overall, our findings show that the adoption of strategies to mitigate the risks associated with using BestMixer is limited, while the transaction volumes users entrust BestMixer with are high. The effort that users do invest concerns the mitigation of identity attribution risks. Our analyses indicate that users do seem to have a reasonable amount of trust in a centralized mixing service concerning the compromise of financial attribution and the proper mixing and return of bitcoins. Yet, users seem more careful about their identity – which they try, however not always successfully, to obfuscate.

Acknowledgments

We are grateful for our collaboration with Dutch law enforcement that allowed us to study BestMixer. We thank the anonymous reviewers and especially our shepherd for the constructive feedback that improved the paper. This research was partially supported by the Ministry of Finance and Ministry of Justice and Security of The Netherlands.

References

[1] M. Aliapoulios, C. Ballard, R. Bhalerao, T. Lauinger, and D. McCoy. Swiped: Analyzing ground-truth data of a marketplace for stolen debit and credit cards. In *(USENIX Sec. 21)*, 2021.

[2] A. Beganski. Ethereum cofounder says he used now-blacklisted tornado cash to donate to ukraine, 2022. https://decrypt.co/107075/ethereum-cofounder-used-blacklisted-tornado-cash-donate-ukraine.

[3] A. Biryukov and I. Pustogarov. Bitcoin over tor isn't a good idea. In *2015 IEEE Symp. on Sec. and Privacy*, pages 122–134. IEEE, 2015.

[4] BitcoinTalk.org. Post by bestmixer account. https://bitcointalk.org/index.php?topic=3140140.0.

[5] Chainalysis. Chainalysis reactor. https://www.chainalysis.com/chainalysis-reactor/.

[6] Chainalysis. Crypto mixer usage reaches all-time highs in 2022, with nation state actors and cybercriminals contributing significant volume. https://blog.chainalysis.com/reports/crypto-mixer-criminal-volume-2022/.

[7] CoinCap. CoinCap API. https://docs.coincap.io.

[8] Coinomize. Faqs. https://coinomize.biz/faq.

[9] J. Crawford and Y. Guan. Knowing your bitcoin customer: money laundering in the bitcoin economy. In *2020 13th Int. Conf. on Systematic Approaches to Digital Forensic Engineering (SADFE)*, pages 38–45. IEEE, 2020.

[10] Alejandro Cuevas, Fieke Miedema, Kyle Soska, Nicolas Christin, and Rolf van Wegberg. Measurement by proxy: On the accuracy of online marketplace measurements. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 2153–2170, 2022.

[11] "Deepwebsiteslinks.com". Blender.io review. https://www.deepwebsiteslinks.com/blender-io-review/.

[12] Bundesministerium der Justiz. German code of criminal procedure - 100j - subscriber data request. https://www.gesetze-im-internet.de/englisch_stpo/englisch_stpo.html#p0806.

[13] EuroJust. Cybercrime judicial monitor, 2022.

[14] FIOD. Arrest of suspected developer of tornado cash. https://www.fiod.nl/arrest-of-suspected-developer-of-tornado-cash/.

[15] M. Gharaibeh, A. Shah, B. Huffaker, H. Zhang, R. Ensafi, and C. Papadopoulos. A look at router geolocation in public and commercial databases. In *Proceedings of the 2017 Internet Measurement Conference*, pages 463–469, 2017.

[16] US Government. United states code, 2010 edition; title 18 - crimes and criminal procedure. https://www.govinfo.gov/content/pkg/USCODE-2010-title18/html/USCODE-2010-title18-partI-chap121.htm.

[17] T. Holt, O. Smirnova, and A. Hutchings. Examining signals of trust in criminal markets online. *Journal of Cybersecurity*, 2(2):137–145, 2016.

[18] Y. Hong. On computing the distribution function for the poisson binomial distribution. *Computational Statistics & Data Analysis*, 59:41–51, 2013.

[19] Y. Hong, H. Kwon, J. Lee, and J. Hur. A practical de-mixing algorithm for bitcoin mixing services. In *Proceedings of the 2nd ACM Workshop on Blockchains, Cryptocurrencies, and Contracts*, BCC '18, page 15–20, New York, NY, USA, 2018.

[20] M. Korczyński and A. Noroozian. Security reputation metrics. In *Encyclopedia of Cryptography, Security and Privacy*, pages 1–5. Springer Berlin Heidelberg, 2021.

[21] J. van de Laarschot and R. van Wegberg. Risky business? Investigating the security practices of vendors on an online anonymous market using ground-truth data. In *Proc. USENIX Sec. Symp.*, pages 4079–4095, 2021.

[22] Livedarknet.com. CryptoMixer.io review. https://livedarknet.com/p/Crypto/cryptomixer-io/.

[23] MasterTheCrypto.com. Rahakott wallet: Masterthecrypto user reviuw guide. https://masterthecrypto.com/rahakott/.

[24] MaxMind. MaxMind GeoIP2 Anonymous IP Database. https://www.maxmind.com/en/solutions/geoip2-enterprise-product-suite/anonymous-ip-database.

[25] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G.M. Voelker, and S. Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proc. of the 2013 conf. on Internet measurement conf.*, pages 127–140, 2013.

[26] M. Möser, R. Böhme, and D. Breuker. An inquiry into money laundering tools in the bitcoin ecosystem. In *2013 APWG eCrime researchers summit*, pages 1–14. Ieee, 2013.

[27] L. Norbutas, S. Ruiter, and R. Corten. Believe it when you see it: Dyadic embeddedness and reputation effects on trust in cryptomarkets for illegal drugs. *Social Networks*, 63:150–161, 2020.

[28] A. Noroozian, J. Koenders, E. van Veldhuizen, C.H. Ganan, S. Alrwais, D. McCoy, and M. van Eeten. Platforms in everything: analyzing ground-truth data on the anatomy and economics of bullet-proof hosting. In *USENIX Security 19)*, pages 1341–1356, 2019.

[29] U.S. Department of Justice. Affidavit in support of criminal complaint and arrest warrant - case 1:21-mj-00400-rmm. https://storage.courtlistener.com/recap/gov.uscourts.dcd.230456/gov.uscourts.dcd.230456.1.1_1.pdf.

[30] U.S. Department of Justice. Affidavit in support of criminal complaint and arrest warrant - case 2:23-mj-00528. https://www.justice.gov/opa/press-release/file/1574581/download.

[31] U.S. Department of Justice. Individual arrested and charged with operating notorious darknet cryptocurrency "mixer". https://www.justice.gov/opa/pr/individual-arrested-and-charged-operating-notorious-darknet-cryptocurrency-mixer.

[32] U.S. Department of Justice. Ohio resident pleads guilty to operating darknet-based bitcoin 'mixer' that laundered over $300 million. https://www.justice.gov/opa/pr/individual-arrested-and-charged-operating-notorious-darknet-cryptocurrency-mixer.

[33] U.S. Department of Treasury. U.s. treasury issues first-ever sanctions on a virtual currency mixer, targets dprk cyber threats. https://home.treasury.gov/news/press-releases/jy0768.

[34] U.S. Department of Treasury. U.s. treasury sanctions notorious virtual currency mixer tornado cash. https://home.treasury.gov/news/press-releases/jy0916.

[35] C. Osborne. Bestmixer seized by police for washing $200 million in tainted cryptocurrency clean. https://www.zdnet.com/article/bestmixer-seized-by-eu-police-over-laundering-of-200-million-in-cryptocurrency/.

[36] J. Pakki. *Everything You Ever Wanted to Know About Bitcoin Mixers (But Were Afraid to Ask)*. PhD thesis, Arizona State University, 2020.

[37] S. Ramos, F. Pianese, T. Leach, and E. Oliveras. A great disturbance in the crypto: Understanding cryptocurrency returns under attacks. *Blockchain: Research and Applications*, 2(3):100021, 2021.

[38] Redis.io. Introduction to redis - documentation. https://redis.io/docs/about/.

[39] J. Redman. Mixing service bitcoin blender quits after bestmixer takedown. https://www.bitcoininsider.org/article/69440/mixing-service-bitcoin-blender-quits-after-bestmixer-takedown.

[40] M. Schopman, H.P.E. Vranken, and K. Kohls. Validating the accuracy of the maxmind geolite2 city database. Technical report, Radboud University, 2021.

[41] Sinbad.io. Faq. https://sinbad.io/en/faq.

[42] J. Stockinger, B. Haslhofer, P. Moreno-Sanchez, and M. Maffei. Pinpointing and measuring wasabi and samourai coinjoins in the bitcoin ecosystem. *arXiv preprint arXiv:2109.10229*, 2021.

[43] T. Tironsakkul, M. Maarek, A. Eross, and M. Just. Tracking mixed bitcoins. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, pages 447–457. Springer, 2020.

[44] Trezor. Address poisoning attacks. https://trezor.io/support/a/address-poisoning-attacks#.

[45] Tsakim. Poisson binomial distribution for python. https://github.com/tsakim/poibin.

[46] R. van Wegberg, J.J. Oerlemans, and O. van Deventer. Bitcoin money laundering: mixed results? *Journal of Financial Crime*, 2018.

[47] Wireshark.org. tshark(1) manual page, 2023. https://www.wireshark.org/docs/man-pages/tshark.html.

[48] L. Wu, Y. Hu, Y. Zhou, H. Wang, X. Luo, Z. Wang, F. Zhang, and K. Ren. Towards understanding and demystifying bitcoin mixing services. In *Proc. of the Web Conf. 2021*, pages 33–44, 2021.

[49] YoMix. Frequently asked question. https://yomix.io/en/pages/faq.