Dirty Digital Washing Machines

Identifying money laundering in mixing services

by

S.F. Beudeker

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday May 25, 2022 at 9:00 AM.

Student number:	5145600	
Project duration:	June 14, 2021 – May 25, 2022	
Thesis committee:	Prof. dr. M. J. G. van Eeten,	TU Delft, chair
	Dr. R. S. van Wegberg,	TU Delft, daily supervisor
	Prof. dr. P. H. Hartel,	TU Delft, second supervisor
	K. J. M. Lubbertsen Msc.,	FIOD, external supervisor

TU Delft

Abstract

Money laundering has been in existence for a long time, but with recent developments in cryptocurrency technologies, using them as a new method of laundering criminal proceeds has become available to a wider audience of criminals due to mixer services. Bitcoin can be considered pseudonymous, and Mixer services attempt to obfuscate the money trail further by creating transactional chaos. Little is known about the exact functioning of these mixers and how criminals interact with these services. In this thesis, a literature study on existing anti- money laundering (AML) technologies and pattern analysis are applied. Unique to this research is the use of insider mixer administration data combined with the use of public blockchain data. The goal of this research is to gain insight into the degree with which mixer transactions can be identified as money laundering. This will increase the foundation for AML regulation and Law Enforcement procedures. The results show that a small share of the transactions can strongly be identified as money laundering, while for a large share of transactions the identification as money laundering is less strong. In addition, some unexpected phenomena such as reverse money laundering were identified.

Contents

List of Figures

List of Tables

1 Introduction

Cybercrime is a growing problem in today's society. With digital systems becoming more integrated in our daily lives, vital infrastructures (e.g., the energy grid, hospitals, or personal- and work computers) have a new vulnerability path through these digital interconnections. In addition to that, this digital infrastructure provides new pathways for financing of illicit materials (e.g. drugs) or illicit activities (e.g. terrorism). However, Criminals require a way to obtain their funds, while remaining anonymous (Levi, 2015). Traditionally this is done with cash, however, the introduction of bitcoin (among others) as a decentralized and anonymous currency has provided a new way for criminals to obtain their profits. On so called "dark markets" criminal goods such as drugs or stolen credentials are exchanged for bitcoin (Motoyama et al., 2011), or ransomware applications require a victim to pay in bitcoin. In addition to this, there is very little regulation regarding bitcoin, and the digital aspect makes it easy to operate across borders. While bitcoin can be seen as an enabler for (digital) criminal activity (Burden & Palmer, 2003), the transparency of the transaction history poses a problem for criminals. In order to combat this problem, criminals try to anonymize their transactions further by making of a service called Bitcoin mixers. In short, a bitcoin mixer receives bitcoin from various customers, and "mixes" the bitcoins before handing them back. In this way, the history of the bitcoin becomes more obfuscated. This poses a big obstacle for law enforcement agencies in tracking down criminals and their activity (European Police Office., 2015). However, such mixing services are not used only for criminal activity. For example, one could decide to use the service due to privacy concerns. This puts the use of such mixers in an ethically grey area. The structure of this thesis is as follows: In chapter 2 existing literature is reviewed and a knowledge gap is identified before the main research question is posed. In chapter 3 the research approach is described by means of created subquestions. Chapter 4 describes the background information necessary to understand some concepts in the research. Then, in chapter 5 indicators of money laundering are derived. Chapters 6 and 7 describe the dataset that was used and the method to prepare the data for analysis. After this, chapter 8 shows the characteristics of the mixer transactions. Following that, in chapter 9 the results of the analysis of input-output relations and money flows are shown, including methods to further characterise the dataset. The final chapters 10 and 11 discuss the social relevance, future work and recommendations for law enforcement, before concluding the research with the answers to the research questions.

2

Literature Review

For this section, a literature review was performed in an attempt to identify a knowledge gap in the current literature. This section consists of several sub sections. First, some core concepts were defined to focus the search. Second, the actual search was performed and literature was selected. Third, the selected literature is discussed. Fourth and last, the main research question is derived from the review.

2.1. Core Concepts

The first core concept for this research is Bitcoin and more generally cryptocurrency. This is an essential concept, since the decentralized nature of these cryptocurrencies makes efforts to regulate it extremely difficult. It can be argued that this decentral system structure is what makes it attractive to criminal actors. Very simply put, Bitcoin is a type of digital currency which has certain characteristics due to the technology it is based upon. This technology is named blockchain and it is the second concept which is highly related to Bitcoin. The Blockchain is a mechanism for processing transactions of cryptocurrency coins which is entirely decentralized, posing as a replacement system for the traditional bank. Since there are various types of blockchains, in this research the blockchain refers to the bitcoin blockchain or similar blockchains. Some essential characteristics of this mechanism are that coins are stored in digital "wallets" which are not linked to personal details, and that all transactions made are stored in a public ledger, which is available for everyone to inspect. (Nakamoto, 2008) Money laundering is the third core concept. One definition of money laundering, found after a simple google search, is as follows: "the concealment of the origins of illegally obtained money, typically by means of transfers involving foreign banks or legitimate businesses." The reason for this concealment is generally so that criminal actors are able to use their proceeds as normal funds. While the previously discussed concepts are, at a very basic level of detail, widely known, the fourth concept is not widespread knowledge as such. This concept is the crypto mixer. A crypto mixer is an essential component in crypto money laundering (Moser et al., 2013). The purpose of these mixers is to obfuscate the origin of certain crypto-coins such as the bitcoin.

2.2. Method

For the literature review, several searches were performed with certain keywords on titles, abstracts and keywords, shown in table 2.2. From the most specific search, all seven articles were included. From the other two searches the five most cited results were selected from each search. Resulting in a selection of 17 items, shown in table 2.1.

Item Type	Author	Title	Search	Focus of Study
journalArticle	(van Wegberg et al., 2018)	Bitcoin money laundering: mixed results?: An explorative study on money laundering of cybercrime proceeds using bitcoin	3	Cash out experiment
journalArticle	(Custers et al., 2019)	Banking malware and the laundering of its profits	3	Cash out model
conferencePaper	(Maksutov et al., 2019)	Detection of blockchain transactions used in blockchain mixer of coin join type	3	Detection for AML/CFT on mixers
book	(Prado-Romero et al., 2018)	Discovering bitcoin mixing using anomaly detection	3	Mixing detection with social network model
journalArticle	(Liu et al., 2020)	A game-theoretic approach of mixing different qualities of coins	3	Coin blacklist avoidance strategy
conferencePaper	(Seo et al., 2018)	Money Laundering in the Bitcoin Network: Perspective of Mixing Services	3	Detection for AML/CFT on mixers
conferencePaper	(Crawford & Guan, 2020)	Knowing your bitcoin customer: Money laundering in the bitcoin economy	3	Mixer identification and understanding
conferencePaper	(Moser et al., 2013)	An inquiry into money laundering tools in the Bitcoin ecosystem	1	Deanonymization/Detection of ML
conferencePaper	(Juels et al., 2016)	The ring of gyges: Investigating the future of criminal smart contracts	1	Criminal Smart Contracts
journalArticle	(Bryans, 2014)	Bitcoin and Money Laundering: Mining for an Effective Solution	1	Cryptocurrencies effects on AML
journalArticle	(McGinn et al., 2016)	Visualizing Dynamic Bitcoin Transaction Patterns	1	Data visualisation
journalArticle	(Stokes, 2012)	Virtual money laundering: the case of Bitcoin and the Linden dollar	1	ML utility and limitations of crypto
book	(Bonneau et al., 2014)	Mixcoin: Anonymity for bitcoin with accountable mixes	2	Accountable mixing
book	(Ruffing et al., 2014)	CoinShuffle: Practical decentralized coin mixing for bitcoin	2	Anonymous mixing
book	(Valenta & Rowan, 2015)	Blindcoin: Blinded, accountable mixes for bitcoin	2	Accountable mixing
conferencePaper	(Ziegeldorf et al., 2015)	CoinParty: Secure multi-party mixing of bitcoins	2	Anonymous mixing
conferencePaper	(Bissias et al., 2014)	Sybil-resistant mixing for Bitcoin	2	Robust mixing

Table 2.1: Literature Selection

#	Keywords	SCOPUS	Web of Science
1	(”bitcoin*” OR ”cryptocurrenc*”) AND (”money laundering” OR ”AML”)	139	78
2	(”bitcoin*” OR ”cryptocurrenc*”) AND (”mixer” OR ”mixing”)	63	30
3	(”bitcoin*” OR ”cryptocurrenc*”) AND (”money laundering” OR ”AML”) AND (”mixer” OR ”mixing”)	7	2

Table 2.2: Literature Search Keywords

2.3. Literature Review

It can immediately be seen that there is a limited amount of focused research with mixers as a main element, as the broadest search produced under 150 articles. This makes sense, since cryptocurrency and mixing, are still quite new technologies. Furthermore, the third search focusing on money laundering AND mixers only produced 7 results. This shows the novelty of the research subject. Now it is worth noting, that the searches were only performed on the title-abstract-keyword fields of articles, so articles that did not specifically focus on mixers but did study them were not included in those results. Nevertheless, even though there is limited literature available on the subject, the literature that is present does vary significantly in their focus of study. From the reviewed literature, three distinct groups can be derived from the studies based on their focus. These groups are discussed in the following sections.

2.3.1. Mixing Protocol Proposals

This group of literature is mainly focused on the proposal of mixer algorithms, and consists of the literature selected from the second search. These focus on more the privacy benefits of mixers and less on the money laundering aspects, since the use of a mixer is not inherently malicious. Still, within this group we see a differentiation within this group based on the main goal of the proposed mixers. Two of these are books which propose a mixer scheme which focuses on the accountability of mixers, which shows that they are also concerned about the possible negative behaviour that they enable, such as money laundering or theft (Bonneau et al., 2014; Valenta & Rowan, 2015). A different focus is taken by (Ruffing et al., 2014; Ziegeldorf et al., 2015) with mixer proposals which are oriented towards total anonymity in an attempt to tackle certain weaknesses that other mixers suffer from. Lastly, one of the mixer proposals also attempts to deal with apparent weaknesses of current mixer protocols by focusing on the robustness of the mixer. This is to protect against attacks and other interference (Bissias et al., 2014).

2.3.2. Money Laundering Detection

When viewing the situation from a different perspective, the second group of literature is oriented towards detection of illicit behaviour such as money laundering, and is therefore, in contrast with the other two specified groups. Each of these apply different methods, such as social network theory (Prado-Romero et al., 2018) and we see that several do manage to deanonymize certain mixers (Maksutov et al., 2019).

2.3.3. Money Laundering Strategies

With the first group’s perspective focusing on anonymization, and the second on deanonymization, here a different perspective is taken. The literature grouped under money laundering strategies focuses on the perspective of the malicious actor in an attempt to identify behaviour of these actors and possible motivations for decision-making. One of the papers focuses on criminal smart contracts, facilitated by a different crypto technology (Juels et al., 2016), while in another a game theoretic approach is taken (Liu et al., 2020). Others focus on more general cash-out models (Custers et al., 2019; van Wegberg et al., 2018) Aside from the distinct groups, there is some literature regarding more general money laundering utility and effects of cryptocurrencies and mixers (Bryans, 2014; Crawford & Guan, 2020; Stokes, 2012)or a presentation of a data visualisation method (McGinn et al., 2016).

In conclusion, since bitcoin mixers are relatively new, not much research has yet been conducted into how these mixers operate, nor on how they are being used for money-laundering purposes. Nevertheless, there are several approaches and perspectives that have been researched, as described.

In some of the literature reviewed in the groups money laundering detection and strategies, obtaining address and transaction information on the mixer is done by interacting with the mixer, and by scraping publicly available data (Prado-Romero et al., 2018; van Wegberg et al., 2018). A limitation of such methods is the fact that it is uncertain where the exact boundaries of the mixers are, and uncertain what share of the mixer is analysed.

2.4. Main Research Question

This thesis aims to fill the knowledge gap of incomplete data in money laundering detection by doing research on the identification of money laundering transactions in bitcoin mixer services. The Fiscal Information and Investigation service (FIOD) seized the administration of a mixer, and took it offline. This provides the opportunity to perform novel research by working with a unique data set. By analysing patterns in the transaction data, the FIOD will be better prepared to intervene. From this, the following research question has been developed:

RQ: To what extent can mixer transactions be labelled as money laundering?

3

Research Approach

The approach for this research is a case study. Since the administration of a bitcoin mixer can be provided by the FIOD (Dutch Fiscal Information and Investigation Service) there is a unique opportunity to do research on this case, since they will be providing insider administration data of the mixer operations. The case in question is a specific bitcoin mixer, named Bestmixer, of which internal administration and transaction data has been gathered. To structure the case study in such a way that the research question can be answered, the research is divided up in sub-questions which are explained in the rest of this section.

Since the main research question aims to label mixer transactions as money laundering, the first step is to determine the definition of money laundering behaviour in the context of the case, and how it can be recognised. Corresponding to this is the first sub-question:

SQ1: What are indicators of money laundering?

In a similar fashion, before an analysis can be performed, transaction characteristics should be described and defined. Using the results of the previous sub-question in combination with the seized data and publicly available data, characteristics can be defined by developing data metrics. The data features and developed data metrics together form the transaction characteristics. The second sub-question:

SQ2: How to characterize bitcoin mixer transactions?

addresses this step of the research. This information provides the starting point for an analysis of the data set of the bitcoin mixer. Using this enriched data set, an observation can be made regarding the relation between input and output transactions, which brings the third sub-question:

SQ3: How do the input transactions and output transactions of the mixer relate?

With the results of this final sub question, the main research question can be answered.

4

Background

4.1. Blockchain

In order to use bitcoin as a currency, it is necessary to be able to perform transactions on the blockchain. For this, one needs a bitcoin address, and an accompanying private key to use the address to perform transactions. These keys and addresses can be generated in an unlimited amount, giving the user the freedom to use as many addresses as they desire.

4.1.1. Clusters

In an attempt to reduce the level of anonymity, clustering techniques are used to link together addresses to a single owner. There are two clustering techniques that are widely known and used effectively: Common Input Ownership Heuristic and One Time Change heuristic (Ermilov et al., 2017; Harrigan & Fretter, 2016; Zhang et al., 2020). The first heuristic is common input ownership, or co-spending, which assumes that addresses used as input in the same transaction belong to the same owner. Each transaction needs to be signed by the private key belonging to the addresses used as input, which is difficult to do without sharing the private key with another party. However, the effectiveness of this heuristic is diminished due to the use of CoinJoin transactions. This is a type of method which allows multiple users to engage in a single transaction (Maurer et al., 2017) The second heuristic is the one time change address. Since all funds need to be spent in a single transaction, a change address is used to receive the remainder of the transaction back. Generally, this address is newly generated for each transaction, and not re-used. Using this pattern, the change address can be linked to the input addresses.

4.2. BestMixer

The case study in this research is performed on the mixer known as BestMixer. In this section the process of using BestMixer is explained from a user perspective. This is done to gain an understanding of how a user can make use of the service, and what decisions a user can make. Overall the process can be summarized into five steps:

1. Browse to mixer interface
2. Choose currency
3. Choose order details
4. Deposit funds
5. Receive funds

For the first step, browsing to the interface, there are two options. The BestMixer interface was hosted both on the clear web and the dark web. Then, after arriving at the interface, a user can choose the currency they wish to mix. BestMixer offered their services for Bitcoin, Bitcoin Cash and LiteCoin,

with future plans to include Ethereum. The third step for a user is to specify the order details. This consists of several elements:

BestMixer Code: Using a BestMixer code is an optional function that a user can use. It functions as a User ID (UID) which can be used in repeated transactions. The effect of using a UID is twofold. The first and main effect of using an UID is to ensure that a user never receives their own coins back. If the UID is not specified, it is possible that a user receives money they deposited into the mixer when operating the mixer for a second time or more. The second effect of using the UID is a loyalty discount. This UID is provided to the user after their first mix. Payout Addresses: The payout address is provided by the user to let the mixer know where to transfer the mixed funds to. A single address is mandatory to make the mixing possible. Optionally, more addresses can be provided to distribute the funds over. Each additional output address increases the service fee. Service Fee: The service fee of BestMixer ranges from 0.5% to 3% of the order value. By varying the service fee, the type of mixing pool could be chosen by the user. Mixer pool selection: Three types of mixing pools were available to the user: Alpha pool, Beta pool and gamma pool. The alpha pool is a pool containing funds originating from other users of the mixer. The Beta pool is a pool containing private system assets, investors' funds and large transactions of users. The gamma pool is similar to the Beta pool, with the difference being that no funds of other users are present in this pool. Payout distribution: If the user has chosen multiple output addresses, they can specify the distribution with which the funds are transferred to the output addresses. Transfer delay: The delay between the deposit and the payout can be determined by the user, as well as the delay between payouts. The maximum delay is 72 hours.After placing the order, the user receives a deposit address on which to deposit the funds which will enter the mixer. To complete the mix request, the user has to transfer the funds to this address. Lastly, after the specified time delay, the user receives funds distributed over their specified payout addresses

5

Money Laundering Indicators

This chapter reflects on the relation between mixing services and money laundering. The exact definition of money laundering can vary slightly between legal systems, but one of the definitions is as follows: ”Money laundering is hiding and/or giving an apparently legal status to an object (usually money or goods) that originates from a crime, so that it can be spent and invested in the upper world.” (AMLC, 2022) From this, two main concepts can be identified, obfuscation and illegal origin. Obfuscating the origin of funds is not inherently illegal, except when this origin is a criminal act. The purpose of a mixing service is to obfuscate the origin of funds such as bitcoin.(Seo et al., 2018) This makes the use of a mixer very attractive to use in money laundering activities. However, this does not make the use of the mixer money laundering by default. Without a criminal origin it cannot be considered money laundering. Therefore, the use of a mixer service can be seen here as an indicator of money laundering, and will be treated as such in this report. Aside from mixing services, there are other services in the bitcoin network which can be considered as money laundering indicators due to their potential illegal nature. In section 5.2.1 these services are categorised in different types and levels. There is another concept closely related to money laundering which is named *reverse money laundering* or *money dirtying*. This concept is similar to money laundering, but has an opposite goal. Where regular money laundering attempts to conceal the origin of the funds, the goal of reverse money laundering is to conceal the destination of the funds. In general, this type of money laundering is related to terrorist financing (Cassella, 2003; Zabyelina, 2015) but it is not limited to terrorist financing as it can also be used for other criminal activities(Compin, 2008).

5.1. Security

For the continuation of illegal activities it is important for malicious entities to be able to conduct their business in a safe way. Both the illegal acts and the laundering of the proceeds need to be performed securely in order for the business to survive. Security behaviours are defined by van de Laarschot as ”any attempts to compromise the availability or usefulness of evidence to the investigative process” (van de Laarschot, 2020; van de Laarschot & van Wegberg, 2021). He then further divides it into five distinct subclasses of data hiding, trail obfuscation, data destruction, data minimisation and choice of online service provider. The focus of this thesis is on the laundering process, and does not include the practices of the illegal origin of the funds. Taking into account the main concepts derived from the money laundering definition, only two of the five subclasses, trail obfuscation and choice of service take focus in this analysis. One of the subclasses, data minimisation, is indirectly related, but is not focused on as an individual subclass. While the use of a mixer in itself can be considered trail obfuscation, there are other steps that can be taken to obfuscate the trail. The choice of service is related to money laundering in two ways. The first relation is with the illicit origin, as some services have inherent illegal nature. The second is related to obfuscation and data minimisation, as there are other services than mixers which can obstruct the tracing process, either by obfuscation, or by minimizing personal data.

5.2. Chainalysis

There are several companies which provide services to analyse blockchain transactions, and one of them is chainalysis. They developed a tool which uses several information sources in order to identify services in the bitcoin blockchain. This tool uses several clustering heuristics, such as the ones described in 4.1.1 to link addresses to services. (Ermilov et al., 2017)

5.2.1. Labels

In order to gain an indication of the source and destination of the coins, the chainalysis service is used to identify known services. For this identification, data from other sources than the blockchain, or off-chain data, is combined with the clustering. For example, if a service publicly announces a bitcoin address to be theirs, this address and the corresponding cluster can then be linked to the service. The type of service is used to label a specific cluster (e.g. and exchange or a gambling service). In this section, the labels and what they signify is explained. Chainalysis has defined 30 categories, of which 17 are present in the data set. Depending on the practices of the services, the categories are divided into levels of severity: Severe, High, High/Medium, Medium and Low. All severity levels except severe are present in this data set. The 18 categories which are present in the data set are explained in table 5.1. As explained in the previous section, the use of certain services can be considered a money laundering indicator.

5.2.2. Service Categories

In this section a description of each category is given. The descriptions for each of the categories in this section are retrieved from the knowledge base of chainalysis. Each service is labelled by chainalysis based on these definitions, which makes them essential to determine what the labels signify. (Chainalysis, Reactor knowledge base, 21 December 2021)

Fraud shop
Financially motivated shops selling different types of data including, PII (Personally Identifiable Information), credit card data, stolen accounts, and more. Unlike Darknet Markets, Fraud Shops are normally operated by a single actor/team and are the sole merchant within the service. Fraud shops also tend to have behavioral differences from darknet markets such as top-up depositing of funds (incremental increases to the total amount), as well as no customer withdrawals. Therefore, most outgoing transactions can be linked to the operators of the Fraud Shop.

Illicit actor organization
Individuals and/or organizations that operate directly or indirectly in various forms of illicit activities. These entities are directly or indirectly involved with risky entities such as darknet markets, fraud shops, extremist financing, hacking, etc.

Darknet Markets
Darknet markets are commercial websites that operate on the dark web, which can be accessed via anonymizing browsers or software such as Tor or I2P. These sites function as black markets by selling or advertising illicit goods and services such as drugs, fraud materials, and weapons, among others. Darknet markets use cryptocurrency payment systems, often with escrow services and feedback systems to help develop trust between the vendor and customer. Darknet markets have become more security conscious over the past few years due to multiple law enforcement shutdowns.

Stolen Funds
Stolen funds comprise instances of hacked exchanges and services. Attackers engage in sophisticated and persistent social engineering, and exploit pre-existing vulnerabilities to transfer funds from exchange hot wallets to their control. The payoff for actors can be enormous with single incidents often resulting in tens of millions of dollars in losses.

Ransomware
Ransomware is special malware designed to encrypt a victim's computer data and automatically request a ransom to be paid in order to decrypt the data. Attackers employ social engineering and phishing schemes that trick people and organizations into downloading the malicious software.

Scam
Scams can impersonate a variety of services, including exchanges, mixers, ICOs, and gambling sites. This category also encompasses scam emails, extortion emails, and fake investment services. They usually offer unrealistic returns on investment, many times trying to mask a pyramid scheme, or pretend to have incriminating personal data on the victim and ask for money in order to not disclose it.

Cryptocurrency ATMs
Cryptocurrency ATMs facilitate the conversion of physical cash into cryptocurrency, or cryptocurrency into physical cash. They operate similar to normal fiat ATMs and typically have a KYC requirement (with smaller amounts requiring less KYC info and larger amounts requiring more KYC info). ATMs typically charge a premium for their service, which allows convenience and speed in buying and selling cryptocurrency compared to online exchanges.

The possibility for exploitation is often dependent on the ATM's KYC requirements. Without KYC, individuals with influxes of physical cash from drug sales and other illicit activity are able to convert funds into cryptocurrency with relative ease. Besides money laundering, attackers who want to receive cryptocurrency by exploiting those who are not technically savvy will often direct their victims to send the funds via ATMs because they're easy to understand.

Infrastructure as a Service
The infrastructure as a service category comprise of all infrastructure surrounding computing and information services, including but not limited to VPN, VPS, Domain Registrar and other popular types of cyber infrastructure. The sending of funds to infrastructure as a service entities could be payment for bulletproof hosts or other infrastructure that could be used for illicit purposes. Conversely, receipt of funds from this category could indicate a cyber infrastructure business account.

High risk exchange
A high risk exchange is an exchange that meets one of the following criteria:

- No KYC: The exchange requires absolutely no customer information before allowing any level of deposit or withdrawal. Or they require a name, phone number, or email address but make no attempt to verify this information.
- Criminal ties: The exchange has criminal convictions of the corporate entity in relation to AML/CFT violations.
- High risky exposure: The exchange has high amounts of exposure to risky services such as darknet markets, other high risk exchanges, or mixing. We examine if the exchange's exposure to illicit activity is an outlier compared to other exchanges. A service with direct high risk exposure one standard deviation away from the average across all exchanges identified by Chainalysis over a 12 month period is considered a high risk exchange.

P2P exchange
P2P exchanges are online sites that facilitate the buying, selling, and trading of cryptocurrency between two individuals while, usually, not being directly in possession of the funds. Some P2P exchanges will not require any KYC (Know Your Customer), making them attractive for money laundering activities.

Mixing services
Mixers are websites or software used to create a disconnection between a user's deposit and withdrawal. Mixing is done either as a general privacy measure or for covering up the movement of funds obtained from theft, darknet markets, or other illicit sources.

Mixers typically pool incoming funds from many users and re-distribute those funds with no direct connection back to the original source.

Gambling
Online gambling can take many forms from resembling a typical casino where you can play card games like blackjack and poker, slot games and the like, to sites for wagering bets on sports or eSports outcomes.

The industry has been an early adopter of cryptocurrency. Users will send cryptocurrency as a convenient alternative to fiat, and get started betting. Gambling is treated differently depending on the jurisdiction, and many sites have lax KYC requirements. Because of this, there's potential for these sites to be used for laundering money. Many of these companies are located in/operating out of island nation-states (such as Curaçao, Cyprus, or Malta).

Exchanges
Exchanges allow users to buy, sell, and trade cryptocurrency. They represent the most important and widely-used service category in the cryptocurrency industry, accounting for 90% of all funds sent by services.

Hosted Wallets
Hosted wallets are an alternative to core wallets (full node wallets). Wallet software allows users to store their public and private keys, and connects to blockchain nodes to transfer funds and check balances. Wallets that control the user's private keys are considered custodial, or hosted, while software that allows users to retain full control of private keys is considered non-custodial.

Hosted wallets can be risky because the user doesn't actually hold their funds, thus opening the possibility of being scammed. It's also possible the service does not implement sufficient security measures, and is vulnerable to attack. However, a reputable hosted wallet service that takes advanced security measures is likely more reliable and convenient than a non-technical or careless individual.

Merchant Services
Merchant services are authorized financial services that enable businesses to accept payments on their customer's behalf. They are also known as payment gateways or payment processors. These services allow merchants to accept cryptocurrency for invoicing and online or in-person payments. This often includes conversion to local fiat currency and settling funds to the merchant's bank account.

Merchant services is generally a low-risk category. Users mostly comprise mainstream, traditional businesses on one end and their customers on another. However, it's worth noting that scammers sometimes integrate merchant services with a malicious website to accept cryptocurrency payments from their victims.

Mining and Mining Pools
Mining is the process by which cryptocurrency is generated. Mining pools are special services where miners can pool their resources - typically GPU or specialized ASIC mining hardware - together towards mining cryptocurrency. By pooling mining resources the pool has a bigger chance of mining a block and the returns are divided among all the miners according to how much mining power each contributed.

Mining pools typically only receive funds from direct mining activity, and as such are typically low risk. However, a pool that accepts deposits from sources other than mining can be exploited for money laundering.

Mining is used for coin generation, when new coins are minted from the mining process.

Other
This category is used when the entity does not represent a widely popular field of operation or is a particular type of operation or entity such as donation addresses, social network bots, seized funds, among others. This category does not have any inherent risk but may contain risky entities.

Unnamed Service
Clusters we identify as behaving as services fall into this category. These are services that have not yet been identified but show the behavior expected of a service. There isn't a standard risk for this category, but once any entity in this category is identified, it is labeled and moved to an appropriate category.

5.2.3. Severity levels

How strong the relation is between money laundering and a service can be coupled with the severity level of a service. Any service with a severity level of high can be directly linked to illegal activity. Services with medium severity are not directly linked to illegal activity, but do allow for either further obfuscation, or data minimisation.

Category	Severity	Description
Fraud shop	High	Shops selling data including PII, credit card data, stolen accounts, etc.
Illicit actor organization	High	Entities (in)directly involved with risky entities such as darknet markets, fraud shops, extremist financing, hacking, etc.
Darknet Markets	High/Medium	Commercial websites operating on the dark web selling illicit goods such as drugs, weapons and others.
Stolen Funds	High/Medium	Instances of hacked exchanges/services
Ransomware/Medium	High	Ransomed funds for encrypted systems
Scam/Medium	High	Instances of entities impersonating individuals or services
Cryptocurrency ATMs	Medium	Devices which allow for converting cash into cryptocurrency or vice versa.
Infrastructure as a Service	Medium	Services regarding any type of digital infrastructure (e.g. servers, VPN, bullet-proof hosting)
High risk exchange	Medium	An exchange with either: No KYC, Criminal ties, or high exposure to risky entities.
P2P exchange	Medium	Exchanges facilitating exchange of cryptocurrency funds without being in possession of the funds.
Mixing services	Medium	Services used to create a disconnect between a user's deposit and withdrawal
Gambling	Medium	Various types of gambling such as casino's or sports betting services
Exchanges	Low	Places for users to buy, sell and trade cryptocurrency.
Hosted Wallets	Low	Services which control wallets for users.
Merchant Services	Low	Payment processing services which allow businesses to accept cryptocurrency
Mining and Mining Pools	Low	Services which operate to support the blockchain. Typically, they only receive funds from mining activity.
Other	-	Identified services which do not belong to a broad category of services.
Unnamed Service	-	Not identified as a service but shows behavior like a service

Table 5.1: Chainalysis service categories

5.3. Trust

Considering the behaviour of users can be beneficial in interpreting the presence of certain patterns. One element of behaviour is trust. The degree of trust has an influence on the choices of the user. This trust can be divided into two categories: intentions and capability. The intentions relate to the premise that the mixer will operate honestly with respect to their promises. If the mixer operators have malicious intent, they could steal the funds which were deposited, or they could leak information to third parties, such as authorities. The capability relates to the quality of the mixing process. Is the mixing process of high enough quality to be able to provide the anonymity that is promised? This is a question that a user of the mixer might consider before making use of the service. In either case, there is a risk of de-anonymisation. The level of additional security measures a user takes can be related to this trust. For example, a darknet vendor wants to cash out his revenues. If the vendor has a high level of trust towards the mixer, they might decide to deposit the revenues directly into the mixer, and use an address of an exchange as pay-out address to convert the revenue to fiat currency. If the level of trust in the mixer is lower, a user could decide to use additional security measures to increase the level of anonymity.

6

Data Description

There are two main sources of data used for this thesis. The first data source is ground truth data retrieved by law enforcement, which again consists of three different data sets: A list of orders, addresses controlled by the wallet and the wallet transactions. The second data source is the public Bitcoin blockchain, containing the history of all transactions performed worldwide.

6.1. Wallets

The wallets contain bitcoin addresses related to the specific wallets and transaction history of the wallets. The transaction history includes the following information of the transactions: Datetime, Direction (incoming or outgoing), Receiving address, Amount in BTC and transaction ID. Details are shown in table 6.1

6.2. Orders

The second part of the ground truth data are the placed orders. These are orders placed by users of the mixer and contain several data points, including the following: Deposit address, datetime, deposit amount, browser language, application cookie, UID, user agent, request uri, order id, service fee, address fee, payout addresses and ip information. Further details are in table 6.2.

Header	Description	Unit
Confirmed	Status of transaction as returned by the blockchain	True/False
Date	Datetime of the block in which transaction is present	Datetime
Type	Denotes if the transaction is incoming, outgoing or internal	'Received with', 'Sent to', 'Payment to yourself'
Address	The receiving blockchain address of the transaction	Blockchain Address (25-36 alphanum)
Amount (BTC)	Value of the transaction in Bitcoin	Decimal value
ID	Transaction ID	64 hexadecimal characters

Table 6.1: Wallet Data Fields

Header	Description	Unit
source	Source of the data	Alphanumeric String
Deposit address	Address on which Mixer receives payment	Blockchain address
datetime	Time on which the order was placed	Datetime
status	Status of the order	'Awaiting', 'Pending', 'Unconfirmed', Sending', 'Canceled', 'Complete', 'None'
Deposit amount	Amount of coins for this order	Decimal value
coin	Type of coin	'btc', 'btt', 'bch', 'ltc'
language	Language of the browser of the user	Two character country identifier
application	Session ID	Alphanumeric String
Partner id	ID of partner website linking to Bestmixer	Alphanumeric String
UID	User ID	Alphanumeric String
User agent	Type of browser of the user	Standard User agent format
Request uri	URI of the order	Alphanumeric String
Order id	Order ID	Alphanumeric String
Fee pa	Fee per payout address	Decimal Value
Fee sr	Service fee	Decimal Value
Out address 0 – 15	Payout addresses selected by user	Blockchain address
Unnamed: 32	Additional Payout address	Blockchain address
Ip range	IP range of user	Ip range
Ip hash	Hash of user IP	Hash

Table 6.2: Order Data Fields

7

Data Preparation

The process to prepare the data set for analysis consists of four steps: 1. Retrieve raw transaction data of wallet ground truth data, 2. Link input and output transactions to orders, 3. Retrieve labels and categories for each input and payout address from chainalysis, 4. Apply filters to improve correctness of the dataset for this analysis. In the following sections these steps are explained further.

7.1. Retrieving Transaction Data

For the first step, since the wallet contains only limited transaction information, the full transaction details were retrieved from the public blockchain. This is done using the transaction ID provided in the wallets. For clarity, the transactions were split into the categories “Received with” and “Sent to”. Therefore, the definition of the input and the output of a transaction is different depending on if the transaction is entering or leaving the mixer. For incoming transactions, the output of the transaction is on the deposit address and the input of the transaction is an address of the user. Inversely, for the outgoing transactions, the input of the transaction is a mixer address, and the output address is the payout address.

This results in two data sets including full transaction information on all transactions directly going into and coming from the mixer, within specified timeframes.

7.2. Connecting Input and Output transactions

In order to connect the inflow and outflow with each other, the orders dataset is used. Each order has a unique deposit address and desired payout address(es). With this, every transaction can be linked to a specific order. However, payout addresses are not necessarily unique, since the user can choose them. Therefore, the payout addresses can be re-used, which complicates linking order ids. For example, if three different orders use the same payout address, it cannot be determined from the address alone to which payout a transaction belongs. To eliminate this uncertainty, some filtering is applied. This is explained in the next section.

7.3. Data Filtering

Three filters are applied to the resulting dataset. Starting with the first filter, one of the main elements of the analysis is the correlation between deposit and payout transactions. If data on either of these transactions is missing, the comparison becomes impossible. Therefore, any order that does not have both an input and an output transaction linked is removed. The second and third filter relate to the uncertainty of payout transaction linking, since some payout transactions use the same payout address, linking them to multiple orders. The second filter makes use of the time delay of the mixer. Because a payout never occurs before the deposit occurs, the time delay cannot be less than zero. In addition, there is also a maximum time delay defined by the mixer, which is 72 hours. With time delay in hours as d_time, only transactions where $0 < d\_time < 72$ are kept. The third filter deals with any remaining cases. It is possible that after these filters, it is still uncertain to which specific order a payout transaction is linked. The transactions in this category are removed as well.

7.4. Value Calculation

The value of Bitcoin fluctuates heavily, in the period the data was gathered, the total price change could reach above 100% increase. While bitcoin is seen by some as an invesment opportunity, this research assumes that users value the Bitcoin at the exchange rate at time of use. As the values in the data set are given in Bitcoin, this value is converted to US dollars at the time of the transaction. The value in dollars was computed using a BTC/Dollar exchange rate retrieved from coincap.io, with the price updating each minute.

8

Characterization

8.1. General characteristics.

The data set consists of n=14,048 orders, spread out over five distinct time series, shown in Table 8.1. Linked to these orders are n=14,095 deposit transactions and n=18.847 payout transactions. In Figure 8.1 the number of orders per date is shown. Aside from the clearly visible time series, we can see that the average number of orders per day has grown after T1. In Figure 8.2 the number of deposit and payout transactions are plotted. Here it shows that the number of deposits is closely related to the number of orders, while there is a larger share of payout transactions. Since an order can have multiple payout addresses, but only one deposit address, it follows logically that an order will generally have a single deposit transaction and possibly multiple payout transactions. Worth to mention is the peak at the end of march where the number of payout transactions per order is much larger compared with the rest of the timeline. When viewing the order count in a daily perspective, as shown in 8.3, it can be seen that the number of orders follows a wave pattern. This could be similar to a day/night cycle.

Series	Start date	End date
T1	2018-07-17	2018-08-16
T2	2018-11-11	2019-01-07
T3	2019-02-07	2019-03-06
T4	2019-03-20	2019-04-15
T5	2019-05-06	2019-05-22

Table 8.1: Time series of orders

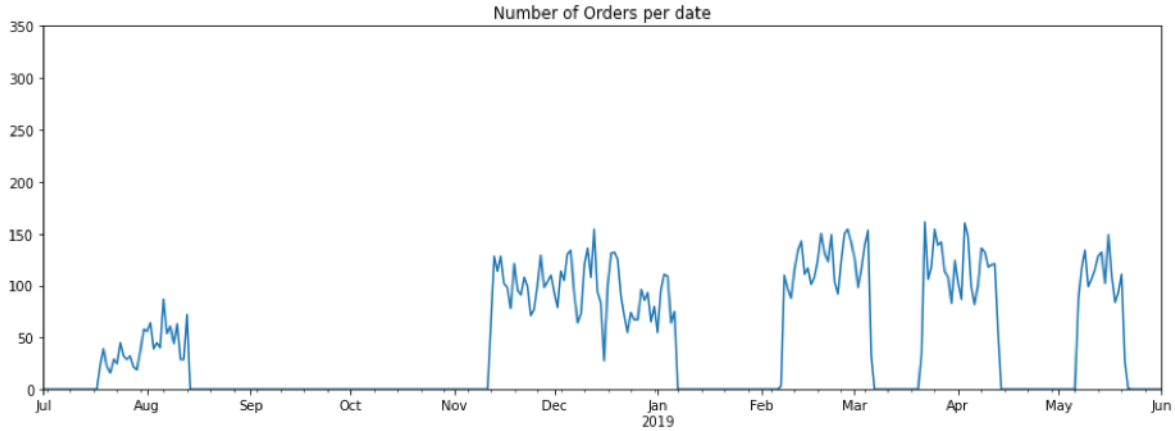

Figure 8.1: Order count per date

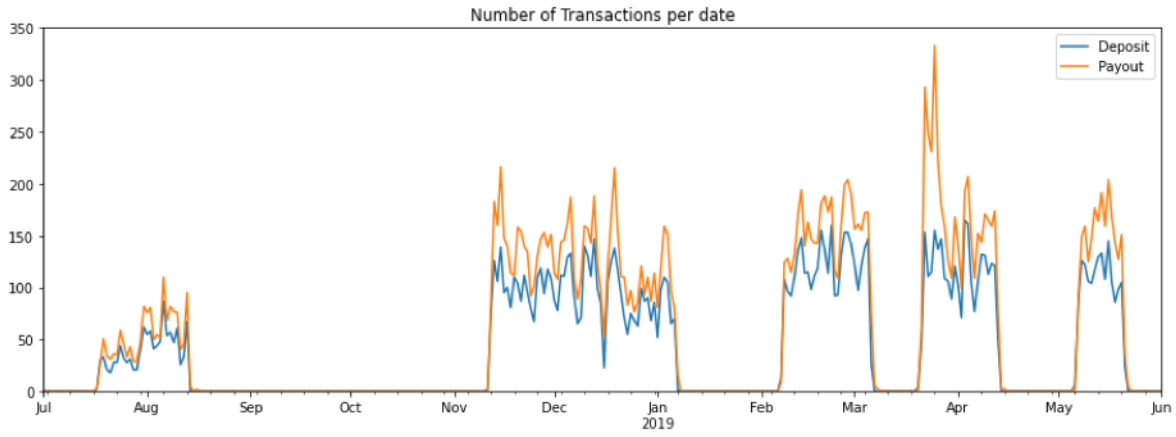

Figure 8.2: Transaction count per date

Figure 8.3: Order count per time of day

8.2. Value Characteristics

To describe the value of the orders some general statistics were computed. The mean value of an order is 0.64 BTC or $ 2,895 and the total value of all orders contained in the dataset amounts to 9,067 BTC or $ 40,671,262. When viewing the distribution of value across orders, shown in Figure 8.4, in it can be seen that the majority of orders are below the mean, with the third quartile at a value of $ 883 while the largest order has a value of $ 782,346. Comparing deposit and payout value shows that the deposit value is larger than the payout value, which relates to the service fee of the mixer. The mixer operates with a service fee between 0,5%-3% of the deposit amount and a flat fee for each additional payout address. Some of the orders show a larger difference or even a negative difference. This can be attributed to the fact that linking the transactions to the orders is not perfect, resulting in orders where part of the transactions is missing.

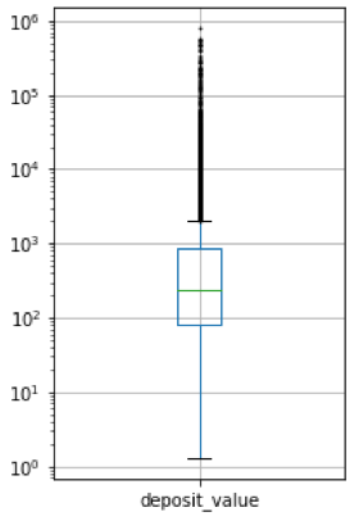

Figure 8.4: Distribution of Deposit Values

Looking at the value over time in figure 8.5, it can be seen that the value fluctuates more than the number of orders, with very noticeable peaks. What is interesting is that the outlier in number of payout transactions that was discovered earlier seems to be an outlier in value as well. On the other hand, the value per time, in figure 8.6 seems to be much more consistent, especially when compared to the number of orders per time of day.

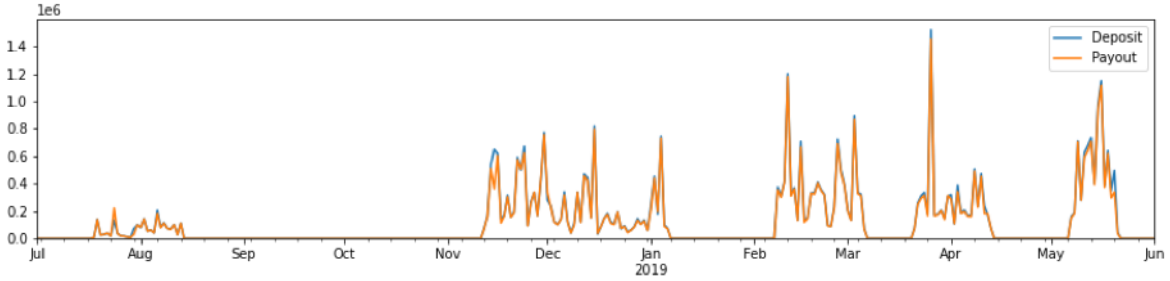

Figure 8.5: Value of transactions per date

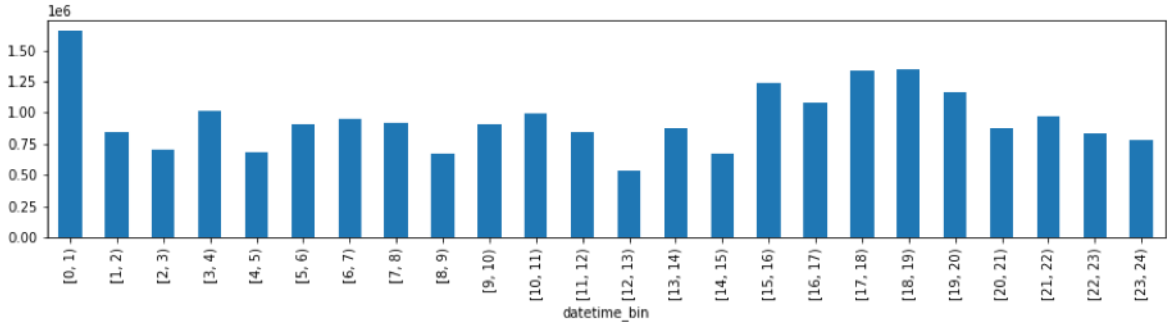

Figure 8.6: Value of orders per time of day

8.3. Category Characteristics

The analysis for categories is divided in two types: deposit category and payout category. From the first observation it is clear that the majority of non-mixer addresses in the deposit or payout transactions are not labelled as part of a service cluster by chainalysis. The percentage of addresses with a label is 28% and 15% for deposit and payout transactions respectively. In other words, 28% of the input

transactions and 15% of the output transactions have a direct and identifiable link to a service. The distribution is similar for the value of the transactions, with 21% and 8% for deposit and payout value respectively. From the security perspective, this indicates that for these transactions, the level of trail obfuscation is lower. No steps were taken to hide the use of a mixer from the service. In Figure 8.7 the total value distributed over the categories can be seen. From the total value it can be observed that the majority of the value related to unlabelled addresses, followed by unnamed services, exchanges and p2p exchanges.

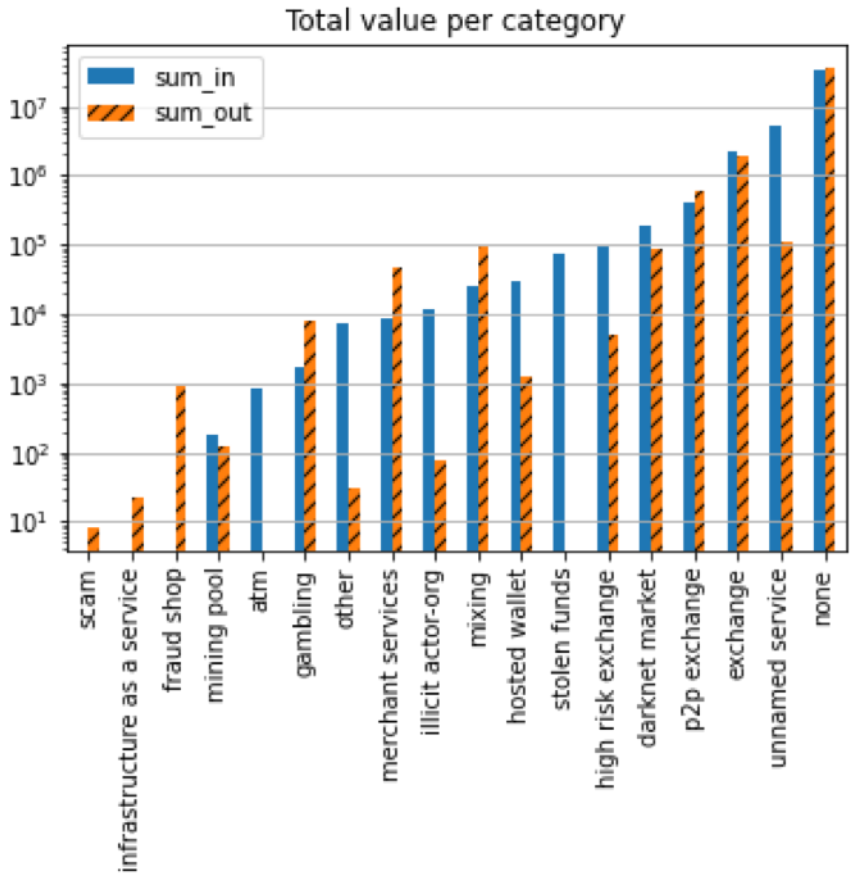

Figure 8.7: Total value per category

For several categories there is a significant difference between their value as an input and their value as an output. Most of the categories labelled are present in both input and output transactions, with the exception of ATM and stolen funds, which are exclusive to input transactions, and fraud shop, scam and infrastructure as a service, which are exclusive to output transactions. Furthermore, the categories unnamed service, illicit actor, hosted wallet, high risk exchange and other, while present in both input and output, have significantly more value present at the input of the mixer as opposed to the output. In addition, the categories gambling, mixing and merchant services have a larger share of value on the output side of the mixer. It is worth noting that most of the majority of categories which are present more on the input side have a security level which is either undefined, high or medium. This supports the expectation that for these categories, the mixer is used to hide/protect their origins.

To interpret these observations, some speculation can be performed based on the characteristics of the individual services. It can be argued for why these categories would not be present on both sides of the mixer. The first category exclusive to inputs is the ATM. A user of an ATM might decide to directly deposit his cash into a mixer, since the deposit address is chosen by the user. The bitcoin paid for a cash withdrawal is deposited to an address owned by the ATM operator, and for an ATM operator there is no direct business incentive to use a mixer. The second category exclusive to inputs is stolen funds. Since a mixer disconnects the inputs from outputs, an output of stolen funds is only

possible if the funds are directly stolen from the mixer. And if the funds were stolen from the mixer, the output address would not be present in the order data, therefore making stolen funds exclusive to inputs in this data set. The categories exclusive to outputs can similarly be argued for, although this distinction is less straightforward. With the first category exclusive to outputs, fraud shop, the funds only flow towards the shop. A fraud shop generally only has one operator and multiple customers, as opposed to darknet markets which generally have more vendors, as explained in 5.2.2. In light of this, there is a bigger likelihood that one or more of the customers use a mixer, compared to the sole operator putting his funds through a mixer. For the scam category and the infrastructure-as-a-service category, the sample size is very small (n=1), so it is not regarded as a significant observation. The fact that mixing as a category is present in this graph is interesting in itself, since it indicates at the minimum a double-mix. A double-mix describes the activity of putting funds through a mixing process twice, possibly in an attempt to increase security. For the other two categories, gambling and merchant services, it can be argued that they are part of a cash-out strategy. Firstly, in the category definition of gambling, chainalysis identifies it as having a potential for money laundering. Secondly, the merchant services category is used to enable business to accept cryptocurrency, which would explain their larger share of output value compared to input value. The fact that merchant services are also present in the input flow is unexpected, since it would mean that either the merchant service would see a need to use a mixer.

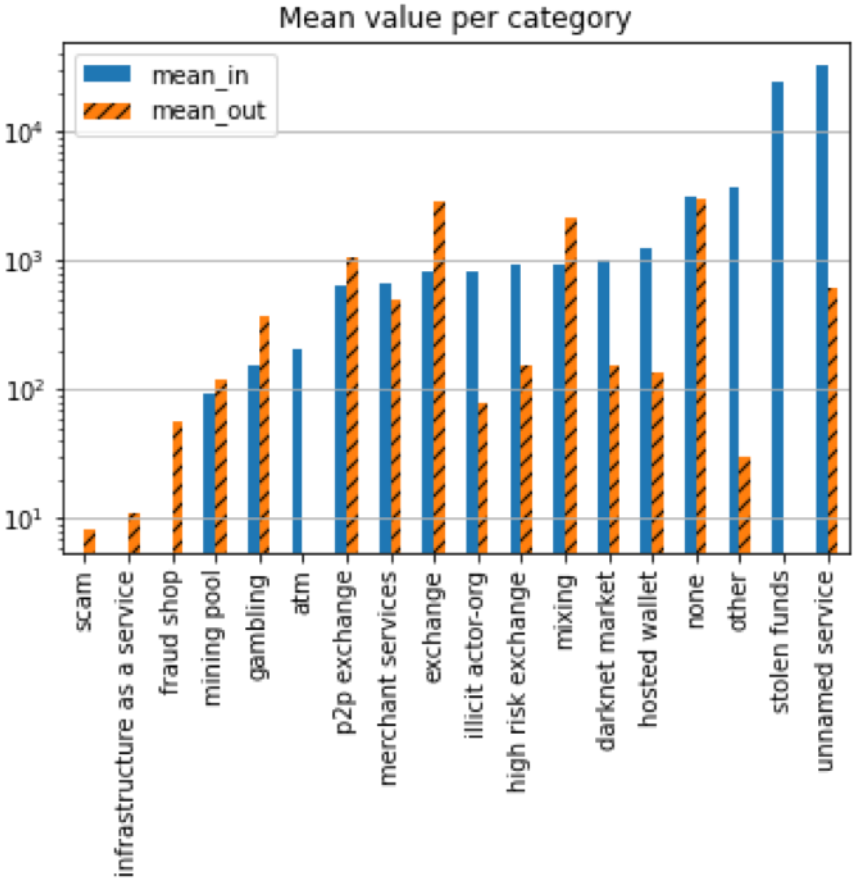

Figure 8.8: Mean value per category

In contrast to the categories with differences, there are also categories of which the difference between input and output value is very small. This is the case for the exchange, p2p exchange, none and mining pool.

Aside from looking at the total value of categories, the size of orders compared to their categories can show if categories tend to have larger or smaller orders. In Figure 8.8 the mean value of an order is shown per category. From this it can be observed that the largest transactions tend to be inputs, and

that categories of which the majority consists of input value, have even higher average input transaction value.

The distribution of value of each category is shown in Figure 8.9. The majority of the categories have their Q1-Q3 range close to 100-1000, similar to the overall distribution shown in Figure 8.4. Based on this, it can be derived that the differences in total and mean value are mainly created by the outliers in this dataset.

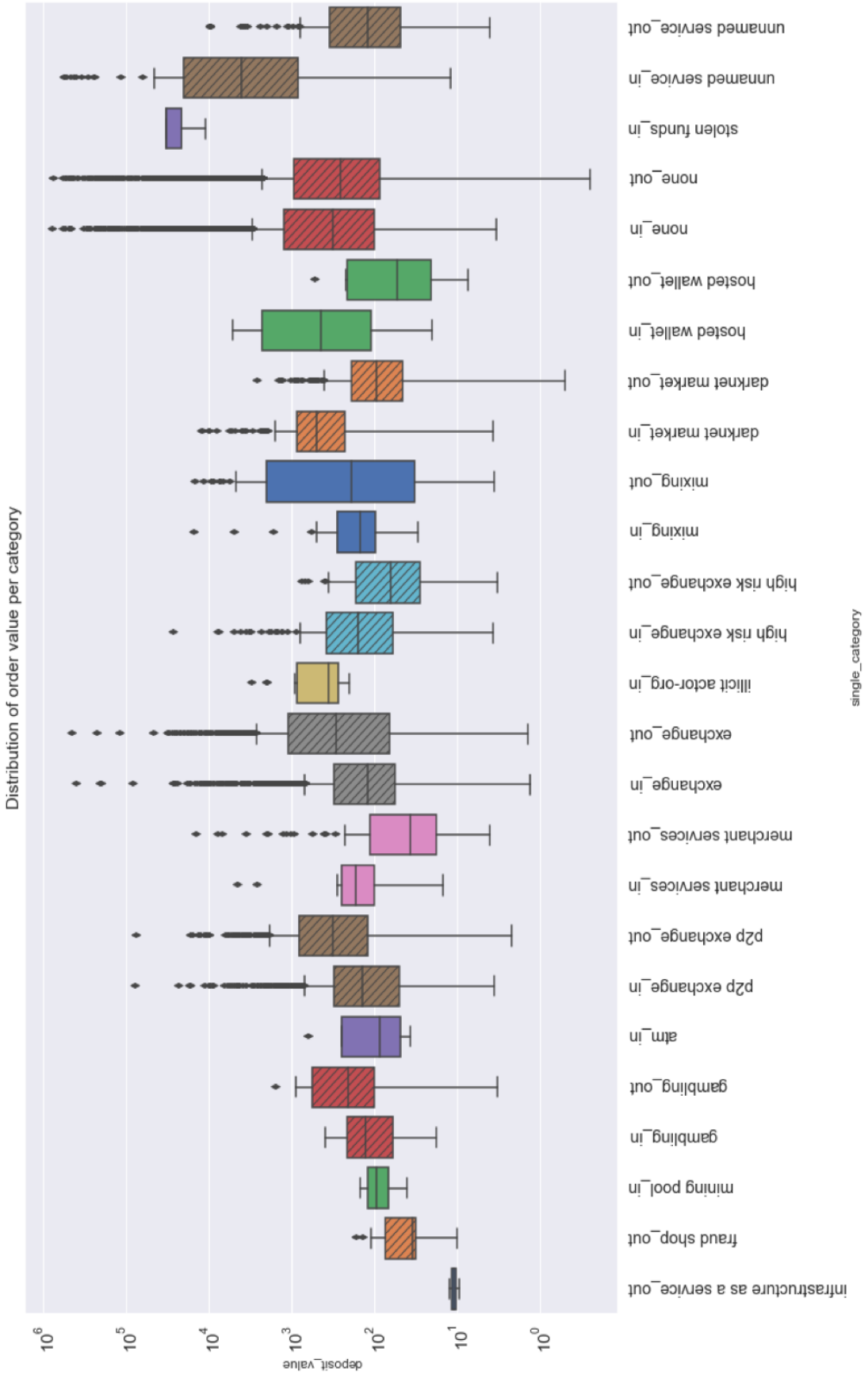

Figure 8.9: Distribution of order value per category

9

Input-Output relations

After having described the characteristics of the input and output transactions of BestMixer, this chapter is dedicated to investigating the relation between input and output transactions.

9.1. Categorising Unlabeled Addressses

From the characterisation it became clear that the majority of the connected addresses were not labelled. This large unlabelled category is further subdivided, to increase depth of the analysis, using two methods. The two methods, private wallet detection and Exposure analysis, are explained in the following section.

9.1.1. Private Wallet detection

A small share of the addresses is identified as a service, and the majority is not labelled as a service by chainalysis, as described in section 5.2. For the addresses without a label, there are two possible explanations. The first possibility is that the address does in fact belong to a service and chainalysis was not able to identify it as such. For some services there is value in not being identified as a service. This is the case with mixer services, for example. The second possibility is that the address does not belong to a service, but is part of a private wallet instead. The use of a private wallet is an understandable choice. The purpose of using a mixer is to complicate the tracking of transactions. Because of that, the use of a mixer is already an indicator that an entity desires their transactions not to be tracked. If the input or the output of the mixer is a service, then, with the use of tools such as chainalysis, it becomes a trivial matter to identify that the funds are put through a mixer. Using a private wallet as an intermediary complicates such identification. Some of the addresses without a label could be undetected services. Incorrectly assuming a service to be a private wallet has consequences for the analysis. Therefore, to further specify the addresses without a label, a heuristic is used. Chainalysis documentation states that the number of addresses contained in a cluster can indicate if it is a service or not. For clusters with more than 500 addresses, it is identified as a service. Due to this, the heuristic is based on the number of addresses in a cluster. The way the clusters are designated as private wallet heavily depends on the limit of addresses. As opposed to services, private wallet clusters generally contain fewer addresses. To ensure that no service is incorrectly labelled as a private wallet, the amount of addresses determining a private wallet is five addresses. The assumption here is that operating a service with five addresses is not possible. In addition, the majority of the unlabelled clusters have one or two addresses The heuristic is as follows: If the unlabelled cluster has five or fewer addresses, it is considered part of a private wallet, and otherwise it remains unknown if it is a service or a private wallet.

9.1.2. Exposure

When following transactions, if a service is reached, it is very difficult to continue following the trail without information from the service. This is due to the fact that multiple entities exchange funds with the service, making ownership unclear. With private wallets this is not the case. This is why the private wallets are investigated further. For this, the exposure information of chainalysis is used. Exposure is defined by chainalysis as the relationship between entities which is created by transactions. The

exposure of a specific entity or cluster can be measured in the amount of funds transacted with other known entities. There are four categories of exposure, based on two characteristics: Sent or received exposure, and direct or indirect exposure. Sent exposure pertains to outgoing transactions of the selected cluster, and received exposure to incoming transactions. Direct exposure refers to a relation between entities without intermediary wallets, whereas indirect exposure refers to a relation with any number of intermediary wallets. The relation of indirect exposure is measured by tracking funds from a cluster until a service is reached. There is a level of uncertainty in this tracking, as there is a possibility that undetected services are passed. In addition, after multiple transactions there is a possibility that the funds move to a private wallet with a different owner. Therefore, the indirect exposure is an estimation. Because of this, the indirect exposure is not used to categorize the addresses, and the distinction within the private wallet category is based on the direct exposure, and the category is split into private wallets with direct exposure and without. Furthermore, only received exposure is considered for input addresses, and only sent exposure is considered for payout addresses.

9.2. Results

As described in the data preparation section, the input and output transactions of the mixer were linked together via the order data. As there are large differences in scale between the presence of the categories, the results were divided into three sections: Total value flows, Exchanges and Medium/High severity services. In the total value flows, the flows without direct service labelling are discussed, as well as giving an overview of the total distribution of flows through the mixer. In the exchanges section, only flows containing an exchange on either side are presented. Exchanges are a convenient way to obtain or to sell Bitcoin in exchange for other currencies, most notably fiat currency, so they can be viewed in part as an ‘entrance’ or ‘exit’ to the Bitcoin environment. The last section describes all flows with at least one link to a medium or higher level severity service.

9.2.1. Total Value Flows

The two most secure categories are Private Wallets and Unknown as they are not directly linked to any kind of service in this analysis. Orders with solely these categories as in/outputs make up 28% of the total orders. This means that 72% of orders has at least one link with a service, either indirect via wallet exposure, or directly from the mixer. There is a large number of flows from Private Wallets (PW) to both PW (38%) and Private Wallets with Direct Exposure (PWDX) (32%). Since these are the largest categories, this is expected. The same can be said for flows from PWDX to PW (41%) and PWDX (28%). The flow relation between PW and PWDX shows an asymmetry in the use of a PW. This shows that among the users exhibiting more secure behaviour, there is still a difference of security behaviour before and after the mixer.

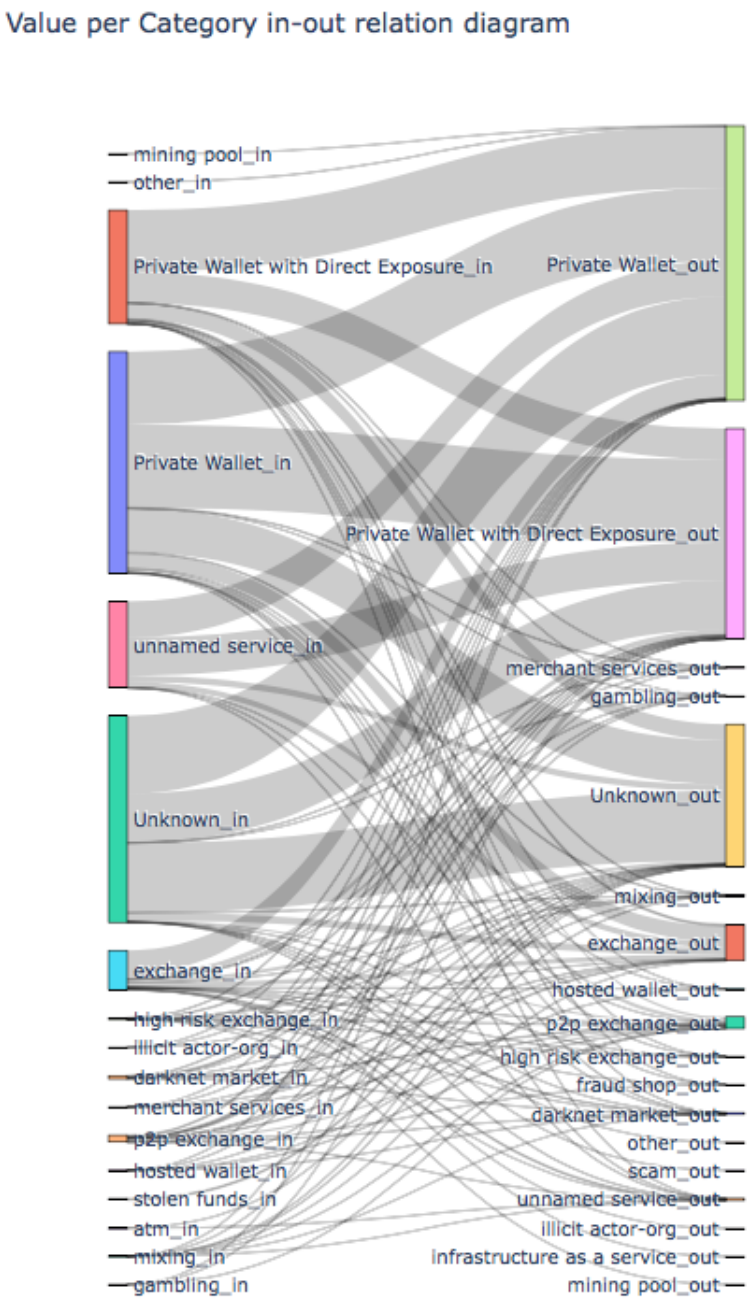

Figure 9.1: Value of flows with all inputs and outputs

9.2.2. Exchanges

To analyse the flows from and to exchanges, a filtered result is presented in 9.2. This selection contains 5 million dollar of flows (13 % of total flows). For normal exchanges, the majority of the funds flow from or towards Private Wallets, while the flows connected to Private Wallets with direct exposure make up a much smaller part of the exchange flows, when compared to the overall distribution. As the exchanges can be used as a direct gateway to and from other currencies, a direct link with an exchange shows that

immediately after entering or before leaving the Bitcoin environment, a mixer is used. As observed, the majority of those flows are linked to a private wallet, and do not have direct links to other services. The argument for no direct links with services can be made, as a user might decide to hold on to their funds in a wallet. This way they retain control over their funds, and are not dependent on the time delay of the mixer to choose a moment of depositing. However, if the funds were moved to a service after holding in a single wallet, the wallet would have direct exposure with this service. For the flows linked to a private wallet, this is not the case. This shows that additional steps have been taken which obfuscate the trail. Another observation is the presence of flows with an exchange as both input and output. Based on the assumption that exchanges have KYC requirements in place, there would be little incentive to put funds through a mixer to another exchange, as identification would be possible at either exchange, at the cost of the mixer fees. It is possible that the exchange does not have proper KYC procedures, however, based on the labeling, such an exchange would fall under the high risk exchange category. Another explanation for this could be that the flow represents a transaction between two users. Moving the funds through the mixer would disrupt the link between the two entities. The p2p exchanges category does not show such a large preference for private wallets. One explanation for this can also be based on the assumption that the use of the private wallet is more secure, as the p2p exchanges are also considered more secure. There is less incentive to use additional security measures if there are already other security measures in place.

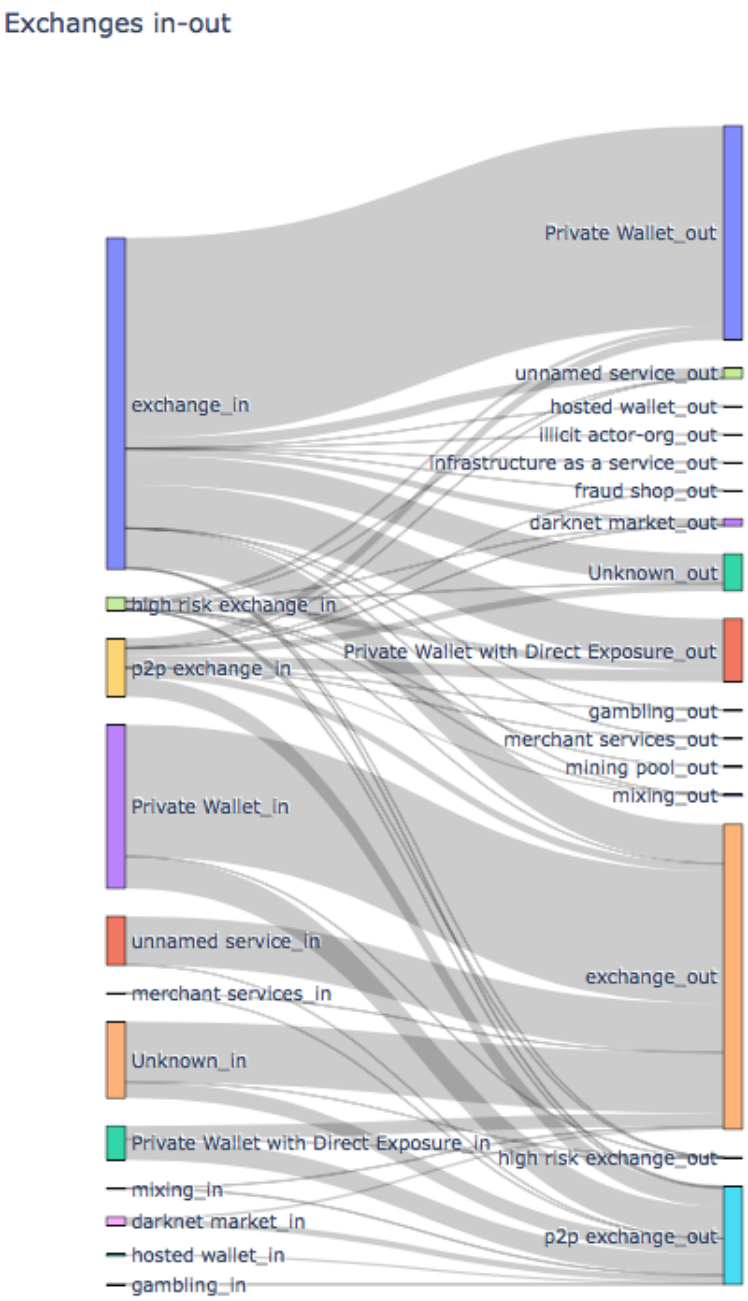

Figure 9.2: Value of flows with an exchange as either input or output

9.2.3. Medium or High Severity

The flows in this section make up an even smaller share of the total value passed through the mixer, 600 thousand dollars (1.5% of total flows) shown in figure 9.3. Slightly more than half of these flows is linked to solely high severity services, with approximately 370 thousand dollars of flows (0.9% of flows). Nonetheless, it is a noteworthy section of the data, as all flows in this section have at least one direct link to a medium or high severity service. Before investigating specific flows, this section indicates that

there are users which have a high level of trust in the mixer. If for some reason the anonymisation process does not work properly, be it intentional or accidental, a direct link can be made between the two services the users made use of. Without additional security measures, their obfuscation becomes minimal. One observation is the relationship with darknet markets. This is a category which has a severity category of high/medium, and a large portion of the value from a darknet market flows directly from or towards an exchange after mixing. Noteworthy is the difference between darknet market in and darknet market out. Of all funds flowing towards a darknet market, 51% originate from a regular exchange and 4% from a p2p exchange. In contrast, of the funds flowing from a darknet market, 22% flow towards p2p exchanges and only 9% towards regular exchanges. Looking at the darknet market specifically, it can be argued that funds flowing towards a darknet market could correspond to customers looking to purchase goods on the market, while funds originating from a darknet market could correspond to vendors looking to cash out their proceeds. It is difficult to prove if this is actually the case, as the means of validation would require identification of the address to the extent that it can be concluded if this address was used as either a customer or a vendor. Still, it often occurs in purchasing transactions that the customer provides funds while the vendor receives them, as compensation for the product or service provided. The flow of funds from the mixer to the darknet market indicates the presence of reverse money laundering.

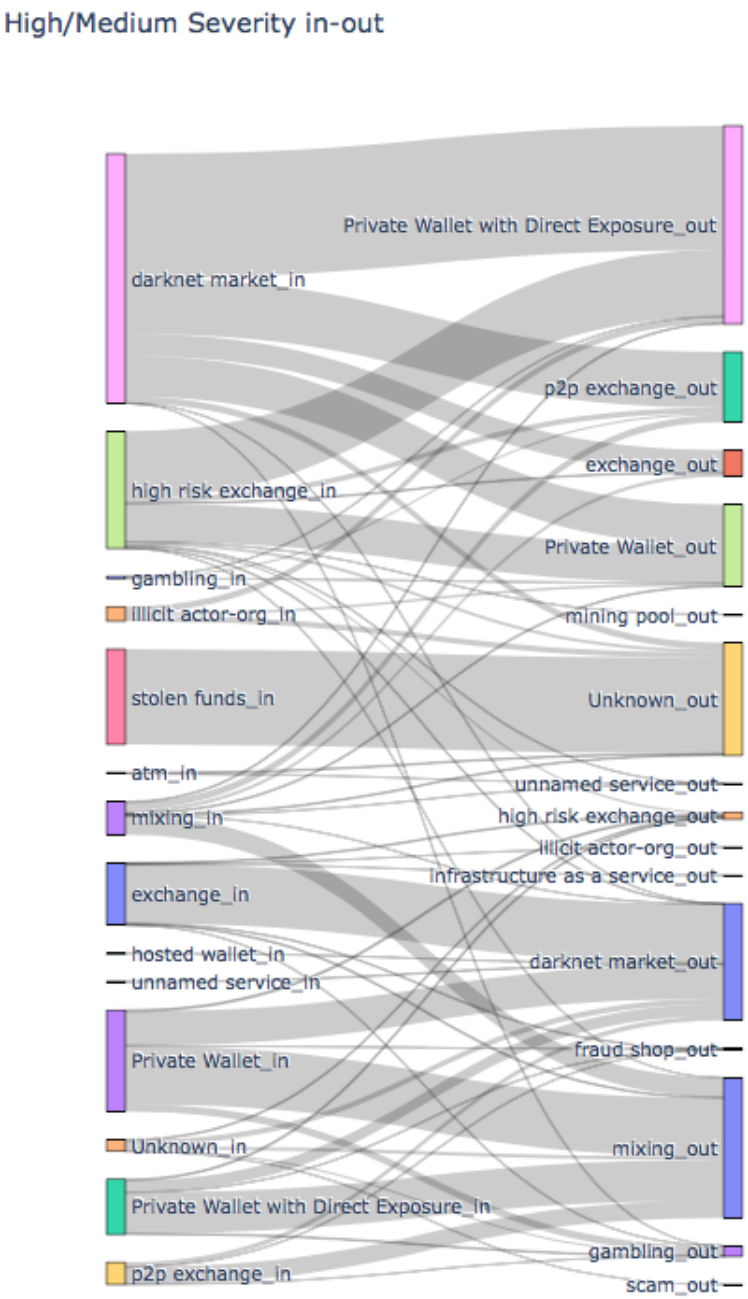

Figure 9.3: Value of flows with a medium/high severity category as input or output

9.2.4. Wallet Exposure

Diving deeper into the exposure of Private Wallets, we take a look at what this exposure actually consists of. In figure 9.4 and figure 9.5 the value distribution of the direct exposure is shown. The value presented here is the share of total exposure to services of all private wallets with any direct exposure. Because of this, interpreting the exposure should be done with caution, as there is no direct link between the exposure of the wallet and the funds that went through the mixer, except for the wallet itself.

It is wholly possible that the funds that did actually go through the mixer never interact with any of the services in the exposure graph. However, it does show that the wallets that are used to hold these funds, do have direct links with service, even if this link might be based on other funds which never passed through the mixer. Any category with a share smaller than 1 % is put in the miscellaneous category. The majority of the exposure is to exchanges, which is expected, as exchanges make up a large part of addresses for the general. The input side shows a larger share of value from exchanges than the output side.

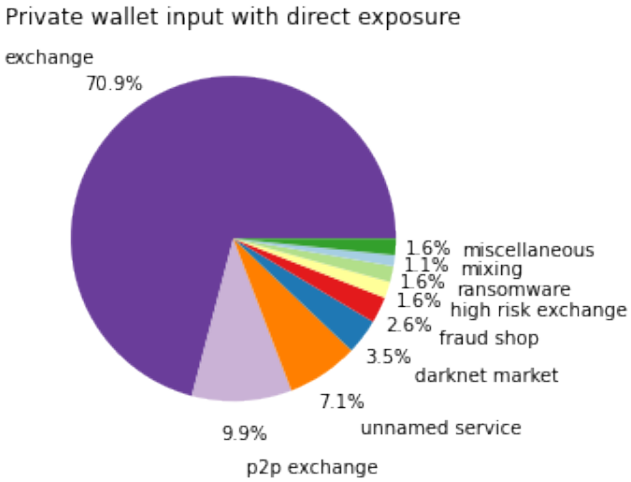

Figure 9.4: Distribution of service exposure of private wallets on the input

For the remaining categories, the input side shows a larger exposure to higher severity categories, with darknet markets and fraud shops making up 6% of the input exposure. The ransomware category makes up another 1.6% of the exposure. None of the direct mixer transactions were labeled as ransomware addresses. The output clusters do not have direct exposure to higher severity categories.

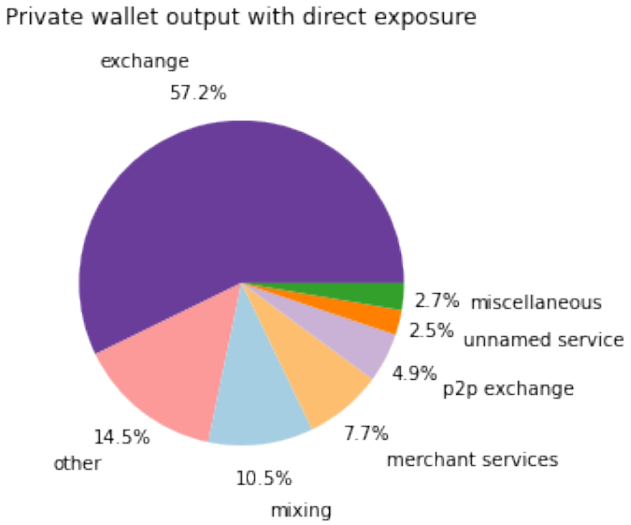

Figure 9.5: Distribution of service exposure of private wallets on the output

10 Discussion

This chapter is dedicated to the discussion of the results of the research, and it is divided into four sections. First, the limitations of the research are discussed. Second, a perspective on potential future work is given. Third, the societal relevance is discussed. For the last section, a recommendation is made towards law enforcement agencies.

10.1. Limitations

One of the limitations of this research is the dependency on chainalysis for the labelling. The identification of the services based on the addresses from the data set is thus from an external source. First off, chainalysis is a paid service, which is a barrier for anyone attempting to reproduce the analysis. Second, the methods used by chainalysis are not transparent. Chainalysis does mention what kind of methods they use to identify services, but the detailed working of their process is not available. Because of this, it is difficult to determine if the labelling of chainalysis contains a bias. Still, the use of chainalysis as a database for service labels does considerably reduce the resources required for such a research. The lack of motivations for behaviour is another limitation of the research. Based on the results of the analysis, some interpretations are made regarding the security behaviour of users. While the data does show certain choices that were made in the mixing process, it is uncertain what the motivation behind this choice was. For example, a user with little knowledge regarding traceability could exhibit similar security behaviour as a user who has more knowledge regarding traceability, but has high trust in the mixer as well. Another limitation to this research is the incompleteness of the data over time. Five distinct time series exist within the data, with large gaps in between. The number of orders and value of orders varies throughout the lifespan of the mixer, which has an impact on the results.

10.2. Future work

Some recommendations for future work can be made. The first recommendation is based on the fact that this research is a case study regarding one case: Bestmixer. To place the findings more firmly in a larger context, it is recommended to repeat the research on other mixers to see if the findings hold in general. The second recommendation is a change of scope of the research. Based on the same data set, using known tracing heuristics, a longer reach analysis of specific categories (e.g. private wallet) can be made. Following the funds that did not have a direct link to a service can be used to further identify actions in the money laundering process. This would come with other biases, as it is harder to trace the longer the trail goes, but it would allow for a more complete picture of the usage of mixers in the money laundering process.

10.3. Societal Relevance

From a regulatory perspective, one thing to reflect upon is the legality of mixing services. The results show us that there is a considerable share of the mixing funds that are directly linked to illegal services, and the expectation is that this share becomes larger the further the transaction paths are followed. Adding this to the notion that from an economic/business perspective there is very little incentive to

use mixing services, helps build the case that mixing services in general should be regarded as illegal. One might argue that if mixing services are located across borders, mixers might be difficult to stop. However, in the results there is a considerable proportion of funds flowing directly between exchanges and mixers. If mixing services were rendered illegal, users would have to be more cautious about moving funds between exchanges and mixers.

10.4. Recommendations Law Enforcement

Without the access to the data set obtained from the mixer, this research would not have been possible. In section 10.2 it was mentioned that performing the research on different cases would be beneficial. Therefore, the first recommendation is to continue retrieving data on mixers. This would not only help in identifying criminal activity, but also increase the possibility for further research. Furthermore, with a certain share of mixer traffic being directly identified as money laundering, the analysis of this traffic for prosecution should be considerably faster. If law enforcement is able to act fast after obtaining data on a mixer, it can show the danger of using a mixer, and could lower trust in such services.

11 Conclusion

In this thesis, the use of a large cryptomixer named Bestmixer was investigated regarding money laundering activities.

Sub Question 1: What are indicators of money laundering
Existing literature was reflected upon to determine money laundering indicators in the context of mixing services. In addition, services were categorised into three severity levels: high, medium and low. Two main elements were derived: Trail obfuscation and illicit origin. The money laundering indicators are based on these two elements. Three of the indicators are based on trail obfuscation, consisting of: The use of the mixer, the use of medium severity services, and the use of private wallets. One indicator, based on illicit origin, is the use of high severity services.

Sub Question 2: How to characterise bitcoin mixer transactions?
Mixer transactions were characterised based on general characteristics, including time and date frequencies, value characteristics and service categories. These characteristics have served to gain insight into the data set, and the service categories characteristics have shown that there are several categories which occur more often at either the input or the output side of the mixer, as opposed to an even division between input and output. The categories which occur either wholly or majorly at the input side of the mixer are: ATM, stolen funds, illicit actor, hosted wallet, high risk exchange, unnamed service and other. As for the output side, the following categories have shown an imbalance toward the output side: Fraud shop, gambling, mixing and merchant services.

Sub Question 3: Relation of input and output transactions
Using the links between input and output transactions resulted in several observations. Of all orders, a share of 72% have at least one link to a service, either by direct link or via wallet exposure. The exposure analysis has shown that 6% of the input exposure is linked to high severity categories. The share of value of orders directly linking to an exchange is 13%. Transactions directly linked to high severity service categories make up approximately 1% of the value of the analysed flows through the mixer. In addition, instances of reverse money laundering have been identified.

Research Question: To what extent can mixer transactions be labelled as money laundering?
Several money laundering indicators were derived for the analysis, and with the indicator of mixer use, every transaction in the data set can, to some degree, be labelled as money laundering. Attempting to increase the confidence of the labelling as money laundering activity becomes more difficult, but is not impossible. The research has shown that various degrees of obfuscation can be identified, and for a not insignificant part of the transactions it is possible to strongly identify money laundering with obfuscation combined with a direct illicit origin. However, for a large share of the transactions, the labelling as money laundering is based solely on the use of the mixer itself as obfuscation.

Bibliography

AMLC. (2022, March 6). *Witwassen* [AMLC]. https://www.amlc.nl/witwassen/

Bissias, G., Ozisik, A., Levine, B., & Liberatore, M. (2014). Sybil-resistant mixing for bitcoin [Journal Abbreviation: Proc ACM Conf Computer Commun Secur]. *Proc ACM Conf Computer Commun Secur*, 149–158. https://doi.org/10.1145/2665943.2665955

Bonneau, J., Narayanan, A., Miller, A., Clark, J., Kroll, J., & Felten, E. (2014). *Mixcoin: Anonymity for bitcoin with accountable mixes* (Bitcoin Foundation; CA Technologies; Google; Silent Circle, Trans.; Vol. 8437) [Journal Abbreviation: Lect. Notes Comput. Sci. Pages: 504 Publication Title: Lect. Notes Comput. Sci.]. Springer Verlag. https://doi.org/10.1007/978-3-662-45472-5_31

Bryans, D. (2014). Bitcoin and money laundering: Mining for an effective solution. *Indiana Law Journal*, *89*(1), 441–472. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84893947596&partnerID=40&md5=9c3259ce8582b07bbb47c8fa716b4719

Burden, K., & Palmer, C. (2003). Internet crime: Cyber crime — a new breed of criminal? *Computer Law & Security Review*, *19*(3), 222–227. https://doi.org/10.1016/S0267-3649(03)00306-6

Cassella, S. D. (2003). Reverse money laundering [Copyright - Copyright Henry Stewart Conferences and Publications Ltd. Summer 2003; Last updated - 2021-09-09]. *Journal of Money Laundering Control*, *7*(1), 92–94. https://www.proquest.com/scholarly-journals/reverse-money-laundering/docview/235831715/se-2?accountid=27026

Compin, F. (2008). The role of accounting in money laundering and money dirtying. *Critical Perspectives on Accounting*, *19*(5), 591–602. https://doi.org/10.1016/j.cpa.2007.01.001

Crawford, J., & Guan, Y. (2020). Knowing your bitcoin customer: Money laundering in the bitcoin economy. *2020 13th International Conference on Systematic Approaches to Digital Forensic Engineering (SADFE)*, 38–45. https://doi.org/10.1109/SADFE51007.2020.00013

Custers, B. H., Pool, R., & Cornelisse, R. (2019). Banking malware and the laundering of its profits. *European Journal of Criminology*, 18.

Ermilov, D., Panov, M., & Yanovich, Y. (2017). Automatic bitcoin address clustering. *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 461–466. https://doi.org/10.1109/ICMLA.2017.0-118

European Police Office. (2015). *Why is cash still king?: A strategic report on the use of cash by criminal groups as a facilitator for money laundering.* Publications Office. Retrieved January 29, 2021, from https://data.europa.eu/doi/10.2813/698364

Harrigan, M., & Fretter, C. (2016). The unreasonable effectiveness of address clustering. *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, 368–373. https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0071

Juels, A., Kosba, A., & Shi, E. (2016). The ring of gyges: Investigating the future of criminal smart contracts [Journal Abbreviation: Proc ACM Conf Computer Commun Secur]. *Proc ACM Conf Computer Commun Secur*, *24-28-October-2016*, 283–295. https://doi.org/10.1145/2976749.2978362

Levi, M. (2015). Money for crime and money from crime: Financing crime and laundering crime proceeds. *European Journal on Criminal Policy and Research*, *21*(2), 275–297. https://doi.org/10.1007/s10610-015-9269-7

Liu, X., Yu, X., Zhu, H., Yang, G., Wang, Y., & Yu, X. (2020). A game□theoretic approach of mixing different qualities of coins. *International Journal of Intelligent Systems*, *35*(12), 1899–1911. https://doi.org/10.1002/int.22277

Maksutov, A. A., Alexeev, M. S., Fedorova, N. O., & Andreev, D. A. (2019). Detection of blockchain transactions used in blockchain mixer of coin join type [ISSN: 2376-6565]. *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 274–277. https://doi.org/10.1109/EIConRus.2019.8656687

Maurer, F. K., Neudecker, T., & Florian, M. (2017). Anonymous CoinJoin transactions with arbitrary values [ISSN: 2324-9013]. *2017 IEEE Trustcom/BigDataSE/ICESS*, 522–529. https://doi.org/10.1109/Trustcom/BigDataSE/ICESS.2017.280

McGinn, D., Birch, D., Akroyd, D., Molina-Solana, M., Guo, Y., & Knottenbelt, W. (2016). Visualizing dynamic bitcoin transaction patterns [Publisher: Mary Ann Liebert Inc.]. *Big Data*, *4*(2), 109–119. https://doi.org/10.1089/big.2015.0056

Moser, M., Bohme, R., & Breuker, D. (2013). An inquiry into money laundering tools in the bitcoin ecosystem. *2013 APWG eCrime Researchers Summit*, 1–14. https://doi.org/10.1109/eCRS.2013.6805780

Motoyama, M., McCoy, D., Levchenko, K., Savage, S., & Voelker, G. M. (2011). An analysis of underground forums. *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference - IMC '11*, 71. https://doi.org/10.1145/2068816.2068824

Nakamoto, S. (2008). A peer-to-peer electronic cash system, 24.

Prado-Romero, M. A., Doerr, C., & Gago-Alonso, A. (2018). Discovering bitcoin mixing using anomaly detection, 8.

Ruffing, T., Moreno-Sanchez, P., & Kate, A. (2014). *CoinShuffle: Practical decentralized coin mixing for bitcoin* (Vol. 8713 LNCS) [Issue: PART 2 Journal Abbreviation: Lect. Notes Comput. Sci. Pages: 364 Publication Title: Lect. Notes Comput. Sci.]. Springer Verlag. https://doi.org/10.1007/978-3-319-11212-1_20

Seo, J., Park, M., Oh, H., & Lee, K. (2018). Money laundering in the bitcoin network: Perspective of mixing services [ISSN: 2162-1233]. *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, 1403–1405. https://doi.org/10.1109/ICTC.2018.8539548

Stokes, R. (2012). Virtual money laundering: The case of bitcoin and the linden dollar. *Information and Communications Technology Law*, *21*(3), 221–236. https://doi.org/10.1080/13600834.2012.744225

Valenta, L., & Rowan, B. (2015). *Blindcoin: Blinded, accountable mixes for bitcoin* (Bitcoin Foundation, Trans.; Vol. 8976) [Journal Abbreviation: Lect. Notes Comput. Sci. Pages: 126 Publication Title: Lect. Notes Comput. Sci.]. Springer Verlag. https://doi.org/10.1007/978-3-662-48051-9_9

van de Laarschot, J. (2020). Risky business: Analysing the security behaviour of cybercriminals active on a darknet market. Retrieved April 6, 2022, from https://repository.tudelft.nl/islandora/object/uuid%3A60a6d46c-e7bf-4621-af55-47467759ecec

van de Laarschot, J., & van Wegberg, R. (2021). Risky business? investigating the security practices of vendors on an online anonymous market using ground-truth data, 18.

van Wegberg, R., Oerlemans, J.-J., & van Deventer, O. (2018). Bitcoin money laundering: Mixed results?, 18.

Zabyelina, Y. G. (2015). Reverse money laundering in russia: Clean cash for dirty ends (P. Bill Tupman Dr. Yuliya Zabyelina, Ed.). *Journal of Money Laundering Control*, *18*(2), 202–219. https://doi.org/10.1108/JMLC-10-2014-0039

Zhang, Y., Wang, J., & Luo, J. (2020). Heuristic-based address clustering in bitcoin [Conference Name: IEEE Access]. *IEEE Access*, *8*, 210582–210591. https://doi.org/10.1109/ACCESS.2020.3039570

Ziegeldorf, J., Grossmann, F., Henze, M., Inden, N., & Wehrle, K. (2015). CoinParty: Secure multi-party mixing of bitcoins [Journal Abbreviation: CODASPY - Proc. ACM Conf. Data Appl. Secur. Priv.]. *CODASPY - Proc. ACM Conf. Data Appl. Secur. Priv.*, 75–86. https://doi.org/10.1145/2699026.2699100