

Characterizing Bitcoin use for Illicit Activities

Karaktäriserar användning av Bitcoin för illegala aktiviteter

Hampus Rosenquist

Supervisor : David Hasselquist
Examiner : Niklas Carlsson

Upphovsrätt

Detta dokument hålls tillgängligt på Internet - eller dess framtida ersättare - under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida <http://www.ep.liu.se/>.

Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a period of 25 years starting from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: <http://www.ep.liu.se/>.

Abstract

Bitcoin's decentralized nature enables reasonably anonymous exchange of money outside of the authorities' control. This has led to Bitcoin being popular for various illegal activities, including scams, ransomware attacks, money laundering, black markets, etc. In this thesis, we characterize this landscape, providing insights into similarities and differences in the use of Bitcoin for such activities. Our analysis and the derived insights contribute to the understanding of Bitcoin transactions associated with illegal activities through three main aspects. First, it offers a comprehensive characterization of money flows to and from Bitcoin addresses linked to different abuse categories, revealing variations in flow patterns and success rates. Second, a temporal analysis captures long-term trends and weekly patterns across categories. Finally, an analysis of outflow from reported addresses uncovers differences in graph properties and flow patterns among illicit addresses and between abuse categories. These findings provide valuable insights into the distribution, temporal dynamics, and interconnections within various categories of Bitcoin transactions related to illicit activities.

Acknowledgments

I am very grateful for the continuous assistance and great interest I have received from Professor Niklas Carlsson, my examiner. Thanks for course-correcting the ship and giving your insight on a weekly basis. It has been pleasant working with you on this project.

I am very thankful for the help I have received from David Hasselquist, my supervisor, and Martin Arlitt.

I would like to thank Axel Flodmark and Markus Jakum for supplying a copy of the downloaded reports from the Bitcoin Abuse Database. Due to the malfunctioning of the database's API since then, this thesis would not have been possible without your copy.

I would like to thank Vera Antonov and Karl Söderbäck for their work on the script they created for their *Bitcoin address analysis* in TDDD17. I based my structure for storing the Bitcoin addresses on your script.

Hampus Rosenquist
Linköping, May 2023

Contents

Abstract	iii
Acknowledgments	iv
Contents	v
List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Motivation and aim	1
1.2 Approach	1
1.3 Contributions	2
1.4 Delimitations	3
1.5 Outline	3
1.6 Paper Included in Thesis	3
2 Background	4
2.1 Bitcoin	4
2.2 Bitcoin abuse	5
2.3 Bitcoin Abuse Databases	6
2.4 Probability distributions	7
2.5 Probability distribution fitting	7
2.6 Related work	8
3 Method	10
3.1 Tools	10
3.2 Data collection	10
4 High-level Characterization	13
4.1 How successful is each address?	13
4.2 Model of the tail distribution	15
4.3 Category-based analysis	16
4.4 Transactions-based analysis	18
4.5 Report frequencies	19
5 Temporal Analysis	20
5.1 Longitudinal timeline	20
5.2 Time of the week	21
5.3 Initial report date analysis	22
6 Following the money	23
6.1 Following the money methodology	24

6.2	One-step concentration or dispersion	25
6.3	Multi-step analysis	26
7	Discussion	29
7.1	Results	29
7.2	Method	30
7.3	The work in a wider context	31
8	Conclusion	32
	Bibliography	34
A	Appendix	38
A.1	Additional statics	38
A.2	Listings	39

List of Figures

3.1	Landscape of Bitcoin abuse.	11
3.2	Summary of primary dataset.	11
4.1	Distribution statistics. Rank plots of (a) the received BTC per address and (b) the cumulative fraction of received BTC. The last two sub-plots show (c) the CDF and (d) CCDF of received BTC per address.	13
4.2	Curve fitting comparison of the CCDF when computing x_{min} for each distribution class.	15
4.3	CDF of received bitcoins.	16
4.4	CCDF of received bitcoins.	16
4.5	Categories overview.	16
4.6	CDF of received transaction sizes per category.	17
4.7	Received bitcoins vs. the number of received transactions. Per-category scatterplots (red) overlayed on overall heatmaps (gradient color).	18
4.8	Mean vs. median number of received bitcoins per transaction. Per-category scatterplots (red) overlayed on overall heatmaps (gradient color).	18
4.9	Successfulness of addresses with different numbers of reports. Per-category scatterplots (red) overlayed on overall heatmaps (gradient color).	19
5.1	Timeline of received bitcoins and number of reports between Jan 2017 and July 2022.	20
5.2	Received bitcoins by the time of the week. Weekday ticks show 12 pm UTC.	21
5.3	CDFs of the relative timing of the transactions compared to the first reporting date of an address.	22
6.1	Scatterplot of received and sent bitcoins.	23
6.2	Visualization of the one-step analysis (the node in-degrees are in the orange circles).	24
6.3	Node in-degrees distributions of receiving addresses being sent money from addresses in each category.	25
6.4	Multi-step analysis. Basic chain of transactions (black) and "concentration" edges (blue arrows).	26
6.5	CDF and rank plot of the node degrees of addresses in the chain, for respective category, when counting only blue "concentration" edges.	26
6.6	CDF of the node <i>in</i> -degrees of addresses in the chain, for respective category, when counting only blue "concentration" edges.	27

List of Tables

4.1	Overview of the top-10 highest receiving reported addresses.	14
4.2	Summary of distributions and their model fits with individual x_{min} and their goodness-of-fit (Kolmogorov-Smirnov distance) to the empirical data.	15
4.3	Power-law fitting of per-category CCDFs.	17
6.1	Address expansion ratio comparison of categories when going one step deep. . . .	24
6.2	Comparison of graph metrics and transaction metrics calculated on each category's flow graph.	27
6.3	Transactions across categories.	28
A.1	Comparison of recent reporting rates and the volume of new addresses being reported in the past five months, and the transactions they receive.	38



1 Introduction

Bitcoin, a decentralized digital currency, is attempting to revolutionize finance by facilitating reasonably anonymous exchanges outside the oversight of traditional authorities. This unique feature has propelled Bitcoin's widespread popularity. However, its inherent pseudonymous design and lack of regulation has also made it a very popular tool for illicit activities, including scams, ransomware attacks, money laundering, and black market transactions. Consequently, Bitcoin presents significant challenges to law enforcement agencies worldwide and strains traditional legal frameworks.

1.1 Motivation and aim

Despite prior works having considered a wide range of criminal activities with relations to Bitcoin [31, 41, 40, 56, 18], most prior work either try to estimate the global cybercrime Bitcoin revenue [45, 22, 27] or focus only on a single category of illegal activities, including money laundering (using tumblers) [32], ransomware [51, 42, 52, 25, 57, 21, 15, 28], sextortion [39, 37], cryptojacking [53], darknet markets [14, 9, 26], and human trafficking [43]. In contrast to these works, we present a comprehensive characterization of the money-flow to and from a large set of addresses that are associated with different categories of illegal activities, and provide insights into similarities and differences in the use of Bitcoin usage for these categories of addresses.

1.2 Approach

This thesis relies on two primary data sources, namely the Bitcoin Abuse Database and Bitcoin's blockchain, for the analysis presented. The Bitcoin Abuse Database [5] provides information on attacks and the corresponding Bitcoin addresses used by attackers, gathered from reports submitted by victims and other individuals or organizations. The reports specify the type of attack and often include additional details like email examples. Additionally, we employ tools to extract information directly from the Bitcoin blockchain, focusing on the identified Bitcoin addresses used by attackers. Through careful combination of these data resources, we provide a comprehensive comparison of the quantity of funds directed to addresses associated with different types of attacks. Importantly, this methodology enables the

observation of transactions involving a larger number of victims beyond those who reported an attack, acknowledging that many actual victims may not report their experiences, while some reports may come from individuals who did not fall victim themselves.

1.3 Contributions

First, we perform a high-level characterization (Chapter 4) of the transactions received by the Bitcoin addresses reported to the Bitcoin Abuse Database [5] from May 16, 2017 to April 25, 2022. Our characterization reveals a high skew in the distribution of funds attracted by different Bitcoin addresses involved in illicit activities: a small subset of addresses received a significant portion of the funds, with the top-10 addresses responsible for 55% of the total bitcoins received, indicating a heavy-tailed distribution. While the overall tail-behavior is best modeled using a power-law distribution with exponential cutoff, we found the tails of the individual abuse categories (together making up the overall distribution) to be well-modeled using power-law functions with slope parameters between 1.3 and 1.4, depending on the category.

Our observations also highlight significant variations in the success of different abuse categories, with "Blackmail scams" and "Sextortion" receiving numerous reports but attracting smaller funds, while categories like "Ransomware" and "Darknet markets" receive fewer reports but attract substantial funds, indicating differences in effectiveness and financial impact. The category attracting the most transactions and funds, however, is an "Other" category, which includes many top addresses that attract the most and the biggest individual transactions. This analysis also reveals that, across and within each category, the most reported addresses (typically associated with high-volume campaigns) are often not the most successful at attracting funds.

Second, we perform a temporal analysis (Chapter 5) that captures both long-term trends, differences in the weekly patterns associated with the different categories, and temporal correlations with when reports of an illicit address are first reported. While the number of reports in the Bitcoin Abuse Database has remained relatively steady since 2019, the daily number of bitcoins received by reported addresses has increased by a factor of 100 over the same period, indicating a substantial rise in funds transferred to these addresses. Weekly variations were observed, with higher volumes and more funds transferred during weekdays compared to weekends, with notable patterns in "Ransomware", "Darknet markets", and the "Other" category. Although the reports typically were obtained around the time that the addresses saw peak activity and there were significant variations between abuse categories, most transactions occur before the first report is filed, suggesting that victims may not report abusive addresses. This raises questions about the effectiveness of using reports to ban addresses.

Third, we analyze the outflow of bitcoins from the reported addresses associated with each abuse category (Chapter 6). Also, this analysis reveals several interesting observations. For example, when considering the outgoing money from reported addresses, there is a concentration of funds towards specific addresses, but the majority of receiving addresses have a node degree of one, indicating dispersion of funds after the first step. In our multi-step tracking of money flows, Bitcoin tumblers stand out with higher node degrees and concentration, suggesting fewer actors are involved in this category, possibly due to the increased effort associated with running such addresses. There are significant differences in graph structure and transaction patterns between categories, with Bitcoin tumblers having more connecting edges and loops, and the "Other" category receiving significantly more transactions going back to reported addresses. Finally, transactions between categories show increased inflow/outflow to/from Bitcoin tumblers, indicating interest in their money laundering services.

Summary of contributions

We present a comprehensive characterization of money flows to and from a large set of Bitcoin addresses associated with different categories of illegal activities. Our analysis and the derived insights contribute significantly to the understanding of Bitcoin transactions associated with illegal activities through three main aspects. First, our high-level characterization reveals variations in flow patterns and success rates both within and across addresses associated with the different categories. Second, a temporal analysis captures long-term trends, weekly patterns, and the relative timing of when illicit addresses of each category are first reported. Finally, an analysis of the outflow from reported addresses uncovers differences in graph properties and flow patterns among illicit addresses and between abuse categories. Overall, the thesis provides comprehensive insights into the characteristics of money flow in Bitcoin transactions associated with various illegal activities, shedding light on the distribution, temporal patterns, and interconnections between different categories.

1.4 Delimitations

The thesis contains the following delimitations:

- Only addresses reported to the Bitcoin Abuse Database between May 16, 2017 and April 25, 2022, are analyzed.
- The addresses' information was gathered from Bitcoin's blockchain on February 8, 2023. Any transactions made by the addresses beyond this date are not analyzed.
- The thesis submits to the categorization chosen by the Bitcoin Abuse Database and only analyzes addresses reported for ransomware, darknet markets, Bitcoin tumbling, black-mail scam, sextortion, and "other".

1.5 Outline

We begin with a background (Chapter 2), including related work (Section 2.6). After presenting our data collection methodology and dataset (Chapter 3), the next three chapters presents our high-level characterization (Chapter 4), temporal analysis (Chapter 5), and outflow analysis (Chapter 6). Finally, we present a discussion (Chapter 7) and conclusions (Chapter 8).

1.6 Paper Included in Thesis

This thesis is based on our recently accepted research paper [46]:

- Hampus Rosenquist, David Hasselquist, Martin Arlitt, Niklas Carlsson. On the Dark Side of the Coin: Characterizing Bitcoin use for Illicit Activities. In Proceedings of the Passive and Active Measurement Conference (PAM), Mar. 2024.

While the work and thesis were completed in 2023, to ensure that we adhered to anonymous submission rules, we held the publication of this thesis until the paper was accepted (in 2024).



2 Background

This chapter describes the necessary background and concepts for the thesis, as well as related works.

2.1 Bitcoin

Bitcoin is a peer-to-peer electronic cash system [34]. Unlike conventional electronic cash transactions, it does not rely on a trusted third party. Instead, peers can exchange electronic cash directly without going through any financial institution, like banks.

In any electronic cash system, the fundamental flaw of double-spending must be solved. Double-spending is when the same electronic cash is spent more than once, which is naturally easy to do in information technology: copy and paste. This is usually solved by letting a trusted third-party keep track of electronic cash transactions. For a peer-to-peer system, however, the double-spending problem is not as easily solved. In 2008, Satoshi Nakamoto proposed “a solution to the double-spending problem using a peer-to-peer distributed timestamp server to generate computational proof of the chronological order of transactions. The system is secure as long as honest nodes collectively control more CPU power than any cooperating group of attacker nodes.” [34]. The proposal was very successful, making Bitcoin the first peer-to-peer electronic cash system to solve the double-spending problem.

The use of Bitcoin is nowadays widespread and accessible to many. But like any tool, it can be used for both good and bad. Due to its decentralized nature, being mostly outside of authorities’ control, and the lack of enforcement of Know-Your-Customer guidelines, Bitcoin is naturally useful for illicit activities.

A melting pot of transactions

Most of the analysis in this thesis does not require us to keep track of *who* has transferred funds to *whom*, but is rather performed by simply counting the funds transferred. However, when following the money paid from one address to another (as done in the following-the-money analysis in chapter 6), greater attention to detail is needed. We describe the main challenge with this type of analysis in this section.

Basics: All Bitcoin transactions are made up of one or more inputs and one or more outputs. The inputs are previous transactions that are transferred to the outputs.

Trivial cases: Since each input and output is labeled with how many bitcoins an address is contributing/receiving as the result of a transaction, for cases where there is a single input the transactions can easily be determined regardless if there are one or more outputs, since these can be split into several separate (virtual) transactions. By symmetry, the same approach works when there is a single output but one or more inputs that are part of the transaction.

Challenging case: However, when there are multiple inputs *and* multiple outputs, the situation is like a melting pot and it is not obvious which bitcoins ended up where. In some cases, however, it is still possible to determine who transferred the funds to whom (with some accuracy) using heuristics based on the input/output sizes.

The decisions (and their limitations) we made to address this, for the following-the-money analysis, are presented in conjunction with its result (Chapter 6).

2.2 Bitcoin abuse

There exist many types of Bitcoin abuse. Since the data set used for this thesis divides it into six categories: "Ransomware", "Darknet markets", "Bitcoin tumbler", "Blackmail scam", "Sextortion" and "Other", we need to know what these mean.

Ransomware

Ransomware is a malicious type of software used to blackmail the target, usually by taking the target's files hostage by encrypting them. The perpetrator then promises to decrypt the files or refrain from publishing sensitive information, if the victim pays the ransom (in bitcoins).

Darknet markets

Darknet markets are illegal online markets that are operating on the so-called dark web and are only accessible through darknets such as Tor or I2P. Bitcoin is often used to exchange money on these markets.

Bitcoin tumbler

A Bitcoin tumbler is a service that mixes bitcoins by lumping many input sources all together over a long and random time frame and then distributing them to the destination addresses chosen by each customer beforehand. This makes it very difficult for a bystander to follow the flow of the bitcoins. It is therefore commonly used to launder "tainted" bitcoins.

Blackmail scam

A blackmail scam is when the perpetrator tries to trick the target into thinking he has a hold over him, in order to blackmail him. It is often in the form of a spam email containing vague false claims, such as "I have hacked your webcam" or "I know your password", blackmailing the victim to pay (in bitcoins).

Sextortion

Sextortion, short for sexual extortion, is a form of sexual blackmail. Usually by threatening to share nude or explicit images if the victim does not pay a ransom (in bitcoins).

2.3 Bitcoin Abuse Databases

To the best of our knowledge, there currently exist two public databases of Bitcoin abuse: the Bitcoin Abuse Database and Chainabuse. The websites are joining forces as of spring, 2023, but both websites are still mostly operative as of May 2023.

Bitcoin Abuse Database

The Bitcoin Abuse Database (bitcoinabuse.com) [5] is a public database of Bitcoin addresses used for various types of abuse. Since 2017, people have been able to report Bitcoin addresses for abuse. The website informs people who have been the target of Bitcoin abuse (such as having received a blackmail email) with: "Do not pay ransoms. Extortion emails are 100% fake". On the website, you can search for a Bitcoin address to see if there are any reports made, and read what they all say. This can be useful if you want to check before interacting with a stranger, or suspect malicious activity. The reports are all made by other users, and each report contains the fields:

id, address, abuse_type_id, abuse_type_other, abuser, description, from_country, from_country_code, and created_at,

where "id" is a unique number assigned to each report by the database. "Address" is the Bitcoin address that is reported for abuse. "Abuse type id" is a number representing the abuse type (category): 1:Ransomware, 2:Darknet markets, 3:Bitcoin tumbler, 4:Blackmail scam, 5:Sextortion and 99:Other. "Abuse type other" is an optional free text column where the reporter may describe the abuse type whenever choosing "Other". "Abuser" is a free text field where the reporter may describe the abuser's identity. "Description" is a free text field where the reporter usually describes the abuse in more detail. Many simply paste an email they have received from a perpetrator. "From country" represents the reporter's (victim's) home country. "From country code" is the reporter's home country code. "Created at" is a date and time field representing the time the *report* was made.

The database is accessible through an API where you can report addresses, lookup abuse types, lookup distinct reports, check an address, and download all reports [3].

Sometime during the spring of 2023, the website stopped allowing new reports to be made. Instead it redirected you to a website called Chainabuse with the following message: "We are happy to announce that BitcoinAbuse and Chainabuse have joined forces. Chainabuse.com is like Bitcoinabuse.com but with more superpowers." [16].

Chainabuse

Chainabuse (chainabuse.com) [11] is a more recently established website that has a database of not only Bitcoin abuse but of blockchain abuse in general. Users can report abuse for blockchains such as Ethereum, Litecoin, Cardano, and many more. The website features the same basic use cases as the Bitcoin Abuse Database: informing, reporting, and address lookup. They describe themselves as the "leading reporting platform for malicious crypto activity worldwide" and are backed by TRM Labs among other "leading organizations and foundations with an interest in making crypto safer for the next billion users" [10].

Compared to the Bitcoin Abuse Database, Chainabuse makes an extra effort to exclude spam from the database of reports and sells access to their address lookup API, as well as other partnerships [10].

Chainabuse's API has fewer endpoints and a lower rate limit for free users. The endpoints include: address lookup (get reports for one address), post a report, and get a report by id [13]. It is not possible to download the entire data set and they were unfortunately not ready to hand it over for our research purposes either.

As mentioned, Chainabuse has collaborated with the older service, the Bitcoin Abuse Database. In practice, this means that Chainabuse has ingested the reports previously made to the Bitcoin Abuse Database. The old reports were filtered for spam before being ingested into Chainabuse's database, according to an email conversation we had with Chainabuse.

In summary, Chainabuse's database of spam-filtered Bitcoin abuse reports would have been of great interest to the thesis, but the data set was unfortunately not accessible.

2.4 Probability distributions

A probability distribution is a mathematical function that tells us how likely each possible value of a random variable is to occur. In this section, four different distributions, that will later be fitted to the empirical data, are presented.

Power law

A power law is a functional relationship between two quantities where a relative change in one quantity is proportional to a change in another quantity raised to a power. It can be expressed as $f(x) = Cx^{-\alpha}$, where C is a constant and α is called the shaping parameter.

Power law with exponential cutoff

A power law with exponential cutoff (also known as a power law with exponential tail) is simply a combination of a power law and an exponential distribution, defined as $f(x) = Cx^{-\alpha}e^{-x/\beta}$, where C is constant and α and β are the shaping parameters. The result is a distribution with a power law relationship for small values of x , and an exponential cutoff for large values of x , where $f(x)$ decreases exponentially. In other words, it has a heavier decrease in the probability of large values of x than a normal power law. A power law with exponential cutoff is one of the more commonly observed probability distributions found in nature and social systems, such as distributions of wealth, city sizes, and earthquake magnitudes.

Lognormal

A lognormal distribution is a continuous probability distribution function (PDF) of a random variable, where its logarithm is normally distributed. I.e., if $y = \ln(x)$ has a normal distribution, x has a lognormal distribution. A normal distribution is a continuous PDF where the values are distributed in a symmetrical fashion mostly situated around the mean. The lognormal distribution can be expressed as $f(x) = \frac{1}{x} \cdot e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$, where μ is the location parameter and σ is the scale parameter of the distribution.

Stretched Exponential

A stretched exponential distribution (also known as a complementary cumulative Weibull distribution) is a fractional power law in an exponential function and can be expressed as $\lambda\beta x^{\beta-1}e^{-\lambda x^\beta}$, where λ is the scale parameter and β is the shaping (stretching) parameter of the distribution.

2.5 Probability distribution fitting

Fitting a probability distribution to empirical data, in terms of repeated measurement of a variable phenomenon, is called probability distribution fitting. The fitting enables predictions of the probability of the magnitude of this variable phenomenon. The fitting is typically limited to a certain interval of the data.

To fit a probability distribution to the data, the distribution's model parameters need to be chosen wisely to find the best fit. First, the interval must be chosen. In this thesis, we are interested in the tail of a curve. Therefore we need to decide a starting point x_{\min} , while the end is simply the end of the curve. An optimal x_{\min} can be found by trying different values and using the Kolmogorov-Smirnov test [29] to find which x_{\min} gives the best goodness-of-fit. Secondly, the model parameters can be chosen wisely using maximum likelihood estimation [33].

Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test, often abbreviated as the KS test, is a statistical hypothesis test used to determine if a sample of data follows a specific probability distribution, such as a normal or exponential distribution. It compares the cumulative distribution function (CDF) of the sample data to the theoretical CDF of the chosen distribution. The test yields a p-value, indicating the likelihood that the sample data comes from the specified distribution; a low p-value suggests a significant deviation from the distribution. It is useful for assessing goodness-of-fit in statistical analysis.

Maximum likelihood estimation

Maximum Likelihood Estimation (MLE) is a statistical method used to find the values of one or more parameters in a statistical model that maximize the likelihood of observing the given data. It assumes that the data are generated from a specific probability distribution and aims to find the parameter values that make the observed data most probable under that distribution.

2.6 Related work

Anonymity: Many works study the anonymity aspects of Bitcoin and identify ways to deanonymize users. Reid and Harrigan [44] present an early study of the anonymity aspects of Bitcoin. Herrera-Joancomart [20] provides an exhaustive review of Bitcoin anonymity. Biryukov et al. [4] show that combining Bitcoin with the anonymizing service Tor creates a new attack vector, jeopardizing their privacy. Androulaki et al. [2] investigate user privacy in Bitcoin by simulating usage of Bitcoin in accordance with Bitcoin's recommended privacy measures, finding that almost 40% of the simulated participants could be profiled using behavior-based clustering techniques with large accuracy. Meiklejohn et al. [31] discuss the challenges Bitcoin's public flow of transactions causes for larger-scale criminal and fraudulent activity. Harrigan et al. [19] explain how "unreasonably" effective address clustering is — i.e., heuristics that group addresses together.

Robustness: Bitcoin's founder Nakamoto [34] and Karame et al. [24] cover Bitcoin and its solution to the double-spending problem. Garay et al. [17] analyze the core of the Bitcoin protocol and prove two of its fundamental properties.

Criminal activities: As we are not the first to characterize the landscape of Bitcoin abuse, many previous works have studied criminal activities related to Bitcoin [41, 40, 56, 18, 45, 22, 27]. For example, Pastrana et al. [41] perform a large measurement of 4.5M crypto-mining malware samples, revealing campaigns with multi-million dollar earnings. Pastrana et al. [40] measure the practice of "eWhoring" (selling photos and videos with sexual content of another person). While most transactions involved PayPal and Amazon gift cards, Bitcoin was found to be a popular tool for offloading eWhoring profits.

Money laundering and tumbling: Möser et al. [32] present the first study on Bitcoin money laundering (tumblers) and conclude that applying a Know-Your-Customer principle to Bitcoin is likely not possible. Others have created protocols that facilitate the service of mixing (tumbling) transactions. Bonneau et al. [8] propose a protocol called MixCoin, later

improved to BlindCoin by Valenta and Rowan [55]. Concurrently, Ruffing et al. [47] proposes a decentralized mixing system called CoinShuffle.

Ransomware: Many works study the use of Bitcoin with ransomware [51, 42, 52, 25, 57, 21, 15, 28]. For example, Kharraz et al. [25] present a long-term study of observed ransomware attacks between 2006 and 2014. More recently, Wang et al. [57] present a large-scale empirical analysis of ransomware relating specifically to Bitcoin, based on data from 2012–2021. Huang et al. [21] study the landscape of ransomware and trace the money flow from when the victim acquires Bitcoins to when the perpetrator converts it back to fiat. For a two-year period, they trace 19,750 victims’ (likely) ransom payments of more than \$16M. Conti et al. [15] conduct a large study of the economic impact of many different ransomwares from a perspective of Bitcoin transactions, including ransoms like WannaCry, Jigsaw and many more. Liao et al. [28] focus on one particular family of ransomware called CryptoLocker, i.e., ransomware that simply encrypts files until the ransom is paid.

Sextortion: Paquet-Clouston et al. [39] study sextortion spam that requires a payment in Bitcoin using a dataset of 4M entries, concluding that one entity is likely behind the majority of them and has gained around \$1.3M over an 11-month period. Oggier et al. [37] also analyze sextortion, but focus on those where the victim is blackmailed (scammed) *with* compromising sexual information, rather than being blackmailed *into* committing sexual actions.

Darknet markets: Christin [14] perform a measurement analysis of the darknet market Silk Road over an 8-month period in 2012. Broséus et al. [9] study the structure and organization of darknet markets from a Canadian perspective. Lee et al. [26] study how criminals abuse cryptocurrencies on the dark web using over 27M dark webpages and 10M Bitcoin addresses, learning that more than 80% of the addresses on the dark web were used for malicious activity.



3 Method

In this chapter, the data collection method will be presented.

3.1 Tools

Scripting language

The Python programming language (version 3.11.3) was used for all scripts that were created for collecting, organizing, calculating, and plotting the data. For on-the-fly data manipulation, the Unix shell bash was used.

Libraries

Several Python libraries were used to facilitate the development. The data manipulation library Pandas [38] was used to handle, manipulate and make calculations efficiently on the parts of the dataset that was saved to comma separated (CSV) files. After the data had been prepared, the plotting library Matplotlib [30] was used. The computational library Numpy [36] was also used to manipulate arrays of the dataset. The network analysis tool NetworkX [35] was used to create graphs to represent the relationships between addresses in the dataset. The analysis library powerlaw [1] and the scientific calculation library SciPy [49] was used to find and compare probability distribution fits to empirical data.

3.2 Data collection

We rely on two primary data sources in this thesis: the Bitcoin Abuse Database and Bitcoin's blockchain. Figure 3.1 presents an overview of our data collection framework in the context of these two information sources.

High-level overview: First, we use the "abuse reports" collected by the Bitcoin Abuse Database to obtain knowledge about attacks and the Bitcoin addresses that the attackers used in these attacks. These reports are typically submitted by victims and other persons/organizations and contain information about what type of attack was performed (e.g., blackmail scam, ransomware, sextortion, etc.) and typically some additional information about the attack (e.g., an email example) and the Bitcoin address(es) used in the attack.

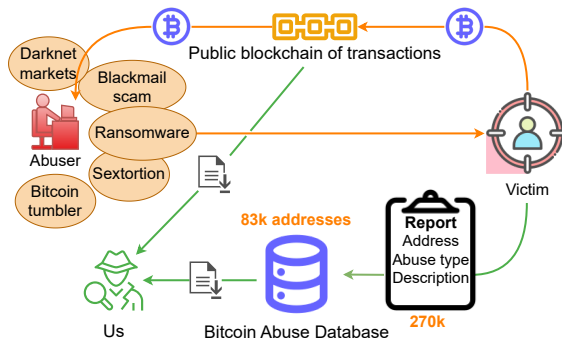


Table 3.2: Summary of primary dataset.

Time period of reports	2017-05-16– –2022-04-25
Reports	267,708
Unique addresses	82,527
Transactions	5,092,489
Received bitcoins	31,346,586
Received in USD	815,011,236,000

Figure 3.1: Landscape of Bitcoin abuse.

Second, we use a series of tools to extract various information about the identified Bitcoin addresses that the attackers used directly from the Bitcoin blockchain itself. Using this information, we compare and contrast how successful attackers were in attracting funds from potential victims to the addresses associated with different types of attacks.

This methodology allows us to observe transactions made by much more victims than only those who report an attack. This is important since we would expect that many victims never report that they have been attacked and reports instead may be filled by people who did not fall victim themselves.

Bitcoin Abuse Database

The Bitcoin Abuse Database [5] contains reports dating back all the way to 2017-05-16; new reports are still being added. We initially obtained all records between 2017-05-16 and 2022-04-25 from this database. After that time, an API issue prevented us from getting all of the data newer than 2022-04-25. The second and third line in Table 3.2 summarizes the number of reports (267,708) and unique Bitcoin addresses (82,527) included in the dataset.

Dataset information: Like mentioned in the background, the Bitcoin Abuse dataset includes the following fields: id, address, abuse type id, abuse type other, abuser, description, from country, from country code, and created at. Each individual field is explained in the background (Section 2.3).

Blockchain information

We have tried several different APIs to retrieve information about each observed address from the blockchain. For the analysis presented here, we used the Blockchain.com API [7]. The data for each address was saved to two files each. One includes the raw JSON form retrieved from the API and one containing a list of the address’ transactions in CSV format with the headers: hash, timestamp, received/sent, address, and value. The purpose of creating the list was to summarize the data points of interest in an easy-to-process format for later analysis.

Pre-processing and summary files

To simplify our data analysis, we created two summary files: one with summary information about each reported address and one with information about each transaction associated with these addresses. Both these files contained summary information based on (1) all reports from the Bitcoin Abuse Database containing the address, (2) the raw JSON files from the blockchain API that were associated with the address, and (3) the list of transactions (and their properties).

Per-address summary file: Each row of this CSV file contains the fields: address, received BTC, sent BTC, balance in BTC, # of received transactions, # sent transactions, # total transactions, average received BTC/transaction, average sent BTC/transaction, median received BTC/transaction, median sent BTC/transaction, date of last transaction, most common abuse type id, abuse type ids, abuse type free-text, # reports, date of the first report, country that the address was most commonly reported in, abuser identity, abuse description. Here, we note that the field “most common abuse type id” captures only the most common abuse type (i.e., the category selected by the reporter). To capture the full set of abuse types that an address has seen reports associated with, we included the field “abuse type ids”, which contains a list of all different abuse types that the address has been labeled with. The same applies to the fields for in which country the abuse was reported in.

Per-transaction summary file: The file containing all transactions was made from each of the address’s individual list of transactions, with some additional fields added based on the per-address summary file, including a field for the abuse type (category) the transaction’s address belongs to and a field for the date that the transaction’s address was first reported on. These fields were later useful when characterizing the transactions. The complete list of headers for the transaction file were: the hash, timestamp, value, address1 (the reported address), received/sent, address2 (other sender/receiver), most common abuse type id, first reported.

Dataset summary

Table 3.2 presents an overview of the dataset for our analysis (based on reports made from May 16, 2017, to April 25, 2022). We also use an extra dataset (based on reports made in the last five months) to confirm that we observe similar reporting rates (114 vs. 147 reports/day) and transactions per address (62.9 vs. 61.7 transactions/address) as seen in the past few years. The statistics for this dataset are provided in the appendix.

Focusing on our main dataset, based on 268K reports, we identified 83K unique Bitcoin addresses that together have received 31M bitcoins across 5M transactions. Using the average price of a bitcoin (BTC) thus far this year (estimated at roughly \$26K USD / BTC) [6], this amounts to a staggering \$815 billion USD in total transaction value.¹

¹Throughout the remainder of this thesis, we use \$26K / BTC when giving a dollar estimate. The price was volatile over the measurement period, with peaks in 2021 and 2022 reaching above \$61K and \$64K / BTC, respectively.

4 High-level Characterization

This chapter examines Bitcoin usage as seen when studying the reported addresses. We first perform an aggregate analysis of how successful each address is (Section 4.1), look closer at how to best model the amount of BTC obtained by the set of addresses (Section 4.2), and present a per-category-based analysis (Section 4.3), before turning our attention to the transactions (Section 4.4) and reports (Section 4.5) themselves. The results and accompanying discussions will be presented. The more general and broader discussion is left for Chapter 7.

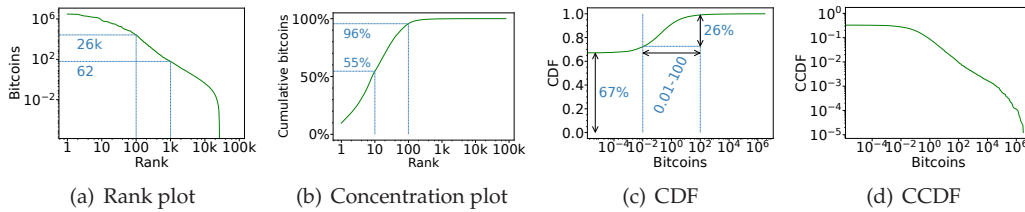


Figure 4.1: Distribution statistics. Rank plots of (a) the received BTC per address and (b) the cumulative fraction of received BTC. The last two sub-plots show (c) the CDF and (d) CCDF of received BTC per address.

4.1 How successful is each address?

In our Bitcoin Abuse dataset, 83K unique Bitcoin addresses were reported. As one might expect, not all addresses were equally successful in (illicitly) attracting funds.

High skew: We observed a significant skew in the number of funds that each address attracted, with a relatively small subset of addresses attracting most of the funds. This skew is characterized and quantified in Figures 4.1(a) and 4.1(b). Figure 4.1(a) shows the total bitcoins received per address (reported to bitcoinabuse.com) as a function of the rank of each address. Figure 4.1(b) shows the cumulative fraction of the total observed bitcoins that the top-X addresses have attracted as a function of the cumulative rank X. The top-10 addresses each received more than 700K bitcoins; together these ten addresses were responsible for 55% of the total bitcoins received across all 83K addresses (i.e., 17M out of 31M bitcoins). Similarly,

Table 4.1: Overview of the top-10 highest receiving reported addresses.

Received [BTC]	Median [BTC]	Category	Description
3,048,040	40.0	Other	Trading investment scam.
2,845,086	18.0	Other	Foreign exchange trading scam, "investment in terror".
2,009,608	25.0	Other	"Investment in terror", begs for treatment money.
1,815,619	800	Other	"Investment in terror".
1,535,341	45.0	Other	"Investment in terror".
1,459,182	160	Ransomware	"Investment in terror".
1,378,975	800	Other	"Investment in terror".
1,259,824	0.50	Other	"Investment in terror".
1,030,376	505	Other	"Inhumane" bank account theft via remote desktop.
724,340	1,150	Other	"Investment in terror", begs for treatment money.

the top-100 all have each received more than 26K bitcoins, combining for more than 29M bitcoins or 96% of the total observed bitcoins in the dataset. The top-1K have each received more than 62 bitcoins (together being responsible for 99.8% of the total bitcoins observed). While 62 bitcoins may seem small relative to the most successful addresses, we note that this still suggests that there are more than 1K abusive addresses that have attracted at least \$1.6M USD (based on \$26K USD / BTC) and that these 1K addresses together have attracted an estimated \$814B USD.

Big hitters: Table 4.1 provides an overview of the number of bitcoins that each of the top-10 accounts have received and the type of reports that have been filed against these accounts. Perhaps most noteworthy, the address that received the most bitcoins during our study received more than 3M bitcoins (worth \$79B USD). This staggering amount is of the same order of magnitude as the Gross Domestic Product (GDP) of entire US states such as Maine and North Dakota [54] or European countries such as Luxembourg [23]. The reports associated with this attack list it as being associated with trading investment scams and link it to CapitalBullTrade. Also, the second-ranked address is reported to be associated with a foreign exchange trading scam (ROFX). The sixth-ranked address has primarily been associated with many ransomware attacks, and the ninth-ranked address is associated with "inhumane" bank account theft through remote desktop software. The remaining addresses on the top-10 list have been reported as organized Bitcoin scam groups that also make worldwide financial "investment in terror", especially in the US, Russia, and Eastern and Central Europe. The reporter of several of these reports claims to have worked for the organized criminals using these addresses for various illicit Bitcoin abuse (e.g., financial scams, begging scams, etc.) and for financial support of "terror".

The most common cases: We next identify the most common cases and how much money these accounts attract. For this, refer to the cumulative distribution function (CDF) of the total received bitcoins per address shown in Figure 4.1(c). 67% of the reported addresses did not receive any bitcoins at all (i.e., the CDF starts at 0.67), suggesting that many of the reported addresses were not successful in attracting funds. Furthermore, among the addresses that received some funds, most received between 0.01 and 100 bitcoins, with the frequency in this interval being s-shaped on log-scale, suggesting a log-normal-like distribution for this region. In total, these addresses make up 26% of all the reported addresses.

Heavy-tailed distribution: As seen in Figures 4.1(a) and 4.1(b), a smaller subset of addresses are responsible for the majority of the received bitcoins, suggesting that the distribution may be heavy-tailed. This is confirmed in Figure 4.1(d), where we plot the CCDF of the number of bitcoins received per address (with both axes on log-scale). While the distribution clearly is heavy-tailed (i.e., heavier than an exponential), the curvature towards the end suggest that the tail is not power law (as often seen in the wild). We next model this tail behavior.

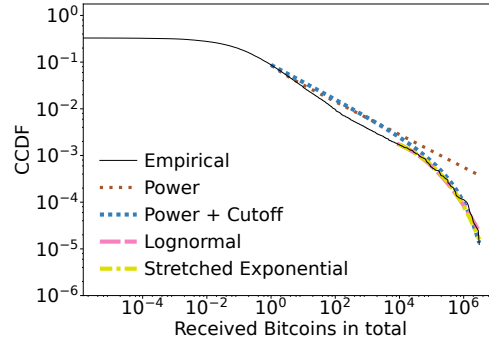


Figure 4.2: Curve fitting comparison of the CCDF when computing x_{\min} for each distribution class.

Table 4.2: Summary of distributions and their model fits with individual x_{\min} and their goodness-of-fit (Kolmogorov-Smirnov distance) to the empirical data.

Distribution	$f(x)$	x_{\min}	Shape parameter(s)	KS
Power law	$f(x) = Cx^{-\alpha}$	1	$\alpha = 1.35$	0.057
Power + Cutoff	$f(x) = Cx^{-\alpha}e^{-x/\beta}$	1	$\alpha = 1.36, \beta = 4.71 \cdot 10^{-7}$	0.099
Lognormal	$\frac{1}{x} \cdot e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$	8,372	$\mu = 9.72, \sigma = 2.17$	0.035
Stretched Exponential	$\lambda \beta x^{\beta-1} e^{-\lambda x^\beta}$	9,444	$\beta = 0.30$	0.036

4.2 Model of the tail distribution

To better understand the shape of the tail, we applied model fitting using the following probability distributions: (1) power law, (2) power law with an exponential cutoff, (3) lognormal, and (4) stretched exponential. For each class, we determined both the x_{\min} from which the distribution gave the best goodness-of-fit (using the Kolmogorov–Smirnov test [29]) and the best model parameters (using maximum likelihood estimation [33]). Table 4.2 summarizes the selected parameters and Figure 4.2 shows the curves fitted from their respective x_{\min} values. While lognormal and stretched exponential provide the best fits for the range they are fitted, we note that they only capture the very end of the tail (as they use x_{\min} values of 8,372 and 9,444, respectively). In contrast, the power-law-based distributions capture a much bigger portion of the tail properly (both models using $x_{\min} = 1$). Of these two distributions, we note that the power-law distribution with the exponential cutoff better captures the shape of the distribution visually, while the pure power-law function has a smaller Kolmogorov–Smirnov distance (as it better captures the convex-shaped portion of the body of the distribution).

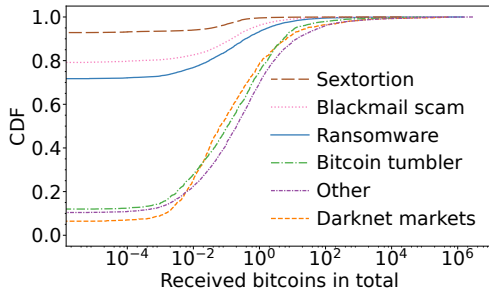


Figure 4.3: CDF of received bitcoins.

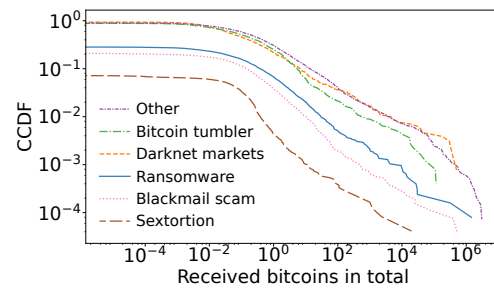


Figure 4.4: CCDF of received bitcoins.

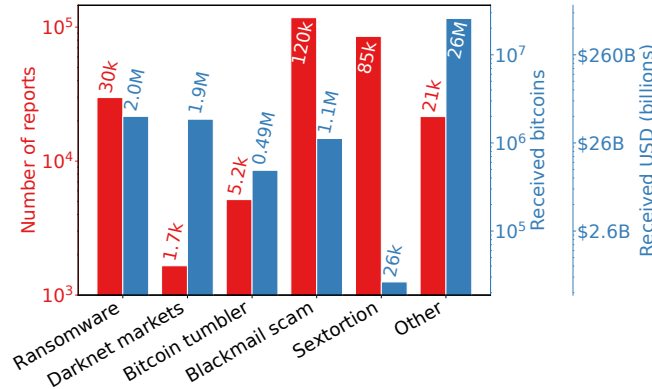


Figure 4.5: Categories overview.

4.3 Category-based analysis

High-level comparison: Consider first the number of reports received by Bitcoin Abuse regarding each abuse category and the number of bitcoins that each category of addresses received. Figure 4.5 summarizes these statistics for each of the abuse categories used by Bitcoin Abuse. To put the number of received bitcoins (shown in blue) in perspective we also show the corresponding amounts in USD.

The two most reported abuse categories (i.e., “Blackmail scam” and “Sextortion”) are among the three categories that attracted the least funds to the addresses associated with their reported attacks. While this may suggest that these attacks are not very successful, we note that the amount of money they attracted still are non-negligible. For example, while the addresses reported in the 120K reports about Blackmail scams “only” received a modest 1.1M bitcoins, this still corresponds to \$28.6B USD. Similarly, the modest 26K bitcoins obtained via addresses associated with Sextortion campaigns are still worth roughly \$676M USD.

Having said that, these amounts are very small compared to the amounts paid to the addresses of the reported Ransomware attacks (2.0M bitcoins worth \$52B USD) or Darknet markets (1.9M bitcoins or \$49B USD), not to mention the “Other” category (26M bitcoins and \$676B USD). As noted, this category includes the top addresses observed in our dataset (e.g., Table 4.1), including trading/investment scams, remote bank account theft, and “investment[s] in terror”.

Also, the reported Bitcoin tumbler sees significant funds passing through them. Here we note that Bitcoin tumbler, also known as a mixing service, combines and shuffles bitcoins from different sources to obscure their original origin, making them an attractive service to be used by the organizations behind many of the illicit addresses. (The tendency to use Bitcoin tumbler for such purposes is clearly demonstrated and further investigated in Chapter 6,

where we trace the bitcoin flows within and across addresses associated with different classes of illicit activities.)

Fraction of addresses not attracting any funds: The relative success of the addresses associated with each abuse category becomes even clearer when looking at the distribution statistics. Figure 4.3 shows a CDF for each category. The leftmost point on each CDF indicates the fraction of accounts in that category that did not receive any bitcoins. Sextortion (93% of addresses), Blackmail scams (79%) and Ransomware (72%) had significantly more accounts that did not receive any funds (i.e., zero bitcoins) compared to the addresses associated with Bitcoin tumbler (12%), the “Other” category (10%), and Darknet markets (6%).

Distribution comparisons: While most addresses that received funds ranged between 0.01 and 100 bitcoins (seen by the s-shaped step in the CDFs for this region), we observe a noticeable shift in the distributions. For example, referring to the CDFs in Figure 4.3, we observe a clear separation between where the different distributions approach one. This separation is more visible in the CCDFs shown in Figure 4.4. Here, the labels of each class are ordered based on the number of accounts that received at least one bitcoin. We observe three distinct groups: (1) Sextortion addresses obtained the least funds, (2) Blackmail scams and Ransomware addresses in general received distinctly more but typically not as much as (3) the addresses associated with Darknet markets, Bitcoin tumbler, and the addresses in the “Other” category.

Table 4.3: Power-law fitting of per-category CCDFs.

Category	Slope estimate		Confidence interval
	x_{\min}	α (σ)	95%
Sextortion	1	1.423 (0.041)	$\alpha \pm 0.000518$
Blackmail scam	1	1.419 (0.013)	$\alpha \pm 0.000161$
Ransomware	1	1.388 (0.013)	$\alpha \pm 0.000234$
Darknet markets	1	1.309 (0.019)	$\alpha \pm 0.00101$
Bitcoin tumbler	1	1.391 (0.016)	$\alpha \pm 0.000612$
Other	1	1.329 (0.005)	$\alpha \pm 0.0000851$

Furthermore, when broken down on a per-category basis, the CCDFs become significantly more power-law-like (compared with the aggregate curve in Figure 4.2), with clear straight-line behavior when plotted on log scale. This is further confirmed by the power-law fitting of each curve. Table 4.3 summarize these fittings, with corresponding confidence intervals on the slope parameter. The slopes are relatively clustered around the range $1.31 \leq \alpha \leq 1.42$, each with a relatively tight confidence interval, and together encompassing the slope of the aggregate curve ($\alpha=1.35$). Rather than the small slope variations, the most visible difference between the CCDFs is instead their relative shift to each other.

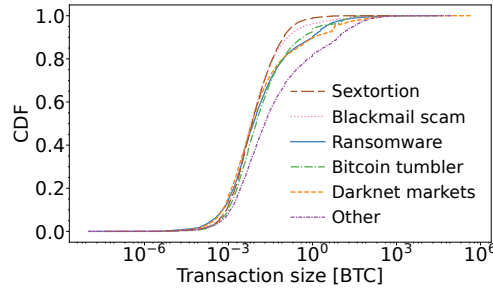


Figure 4.6: CDF of received transaction sizes per category.

4.4 Transactions-based analysis

There are two primary contributors to the differences seen in the distributions of the number of bitcoins received per address when comparing the different abuse categories: the transaction sizes and number of transactions. First, as shown in Figure 4.6, the size distributions of individual transactions differ substantially between the categories. Here, we note a clear shift in the size distributions, captured by noticeable differences when looking at the upper percentiles, for example.

Second, in addition to bigger transactions, the most successful addresses in these categories also received more transactions. To capture the strong correlation between how successful individual addresses were at attracting funds and the number of victims, Figure 4.7 shows per-category scatterplots of the received bitcoins (per address) and the number of received transactions (per address) for each category. To simplify comparisons between the categories, the scatterplots (shown using red points) are overlaid on a heatmap of the overall per-address distribution (across all categories). Here, the color in the heatmap shows the probability density function (PDF) of addresses observed with that combination. While the distributions for the first four categories (i.e., Ransomware, Darknet markets, Bitcoin tumbler, and Blackmail scams) look relatively similar, with a clear cluster receiving up-to 100 bitcoins spread over up-to 1K incoming transactions, Sextortion and the “Other” category stand out. Again, the Sextortion addresses receive a much smaller number of bitcoins and typically see noticeably fewer transactions than the addresses associated with the other categories. Meanwhile, the “Other” category is responsible for most of the highest receiving addresses (e.g., receiving 1K–2M bitcoins) as well as some of the addresses that received the most transactions.

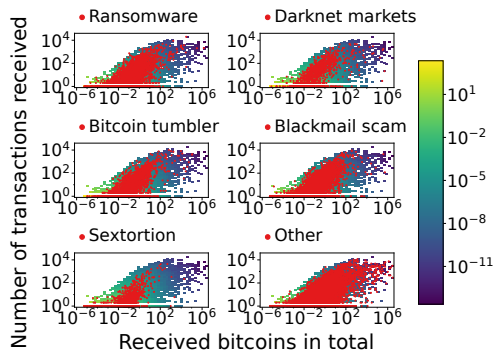


Figure 4.7: Received bitcoins vs. the number of received transactions. Per-category scatterplots (red) overlaid on overall heatmaps (gradient color).

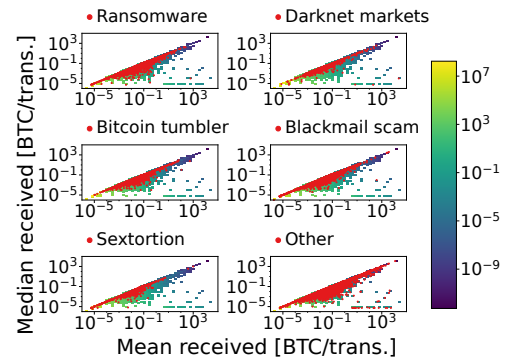


Figure 4.8: Mean vs. median number of received bitcoins per transaction. Per-category scatterplots (red) overlaid on overall heatmaps (gradient color).

High skew in transaction sizes within each category: While we observe a high correlation between the number of transactions and the number of received bitcoins, especially when looking at individual categories, there are several noticeable exceptions. One reason for this is the high skew in the size distribution of transactions (CDFs in Figure 4.6).

High skew in transaction sizes for individual addresses: We have also seen that the sizes can differ substantially for the transactions of individual addresses, best visualized in Figure 4.8, where we plot the median vs. mean number of received bitcoins per transaction and address. Here, the addresses with relatively symmetric size distributions fall close to the diagonal and those with high skew fall below the diagonal. Perhaps most noticeable is a “line” of addresses at the bottom right of “Other”. Those addresses have a very high mean but low median, due to a few very large incoming transactions driving up the mean

significantly together with many smaller transactions dragging down the median. We expect that addresses with higher skew may be used for a more diverse set of abuses, targeting both “big” and “small” fish.

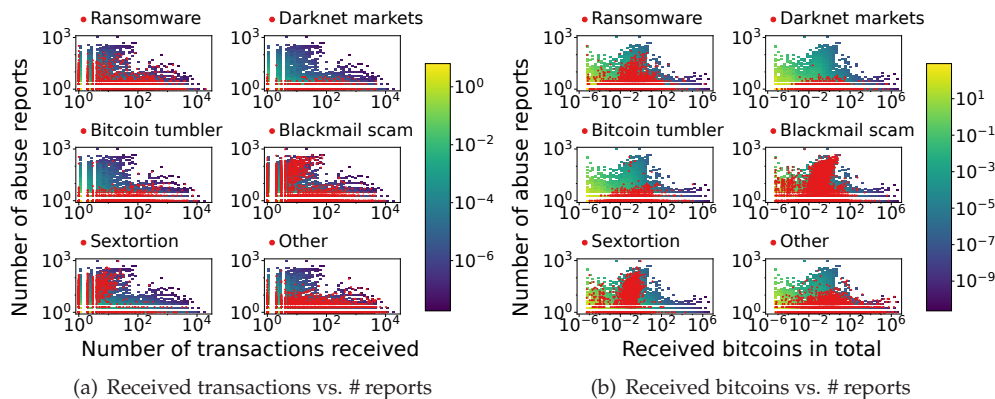


Figure 4.9: Successfulness of addresses with different numbers of reports. Per-category scatterplots (red) overlaid on overall heatmaps (gradient color).

4.5 Report frequencies

It is expected that (low-effort) attacks targeting many users will see many reports. It is therefore not surprising to see the much higher report frequencies of Blackmail scams and Sextortion abuse in Figure 4.5. What is perhaps more interesting is that all categories, including these two categories, include a noticeable mix of low-effort attacks (e.g., spam campaigns) and high-effort, directed attacks. In addition to explaining the high skews observed in how successful different addresses within a category are at attracting funds (both in terms of bitcoins and incoming transactions), we note that these differences also can be observed in the relatively lower correlation between the number of transactions and reports (Figure 4.9(a)), as well as between the number of received bitcoins and the number of observed reports (Figure 4.9(b)).

Addresses best at attracting funds are not highly reported: Referring to Figure 4.9(b), we note that the most successful addresses at attracting funds are only reported a few times (perhaps representing targeted efforts) and that the number of reports per address are relatively independent of the quantity of received funds when considering the three most successful categories: Darknet markets, Bitcoin tumbler, and “Other”. In contrast, the most reported addresses (most belonging to the other three categories: Ransomware, Blackmail, Sextortion) typically received less than 100 bitcoins.

5 Temporal Analysis

This chapter examines the reported addresses temporally. We perform a longitudinal analysis of received bitcoins and report count (Section 5.1), examine weekly patterns associated with the different categories (Section 5.2), and look at temporal correlations with when reports of an illicit address is first reported (Section 5.3). The results and accompanying discussions will be presented. The more general and broader discussion is left for Chapter 7.

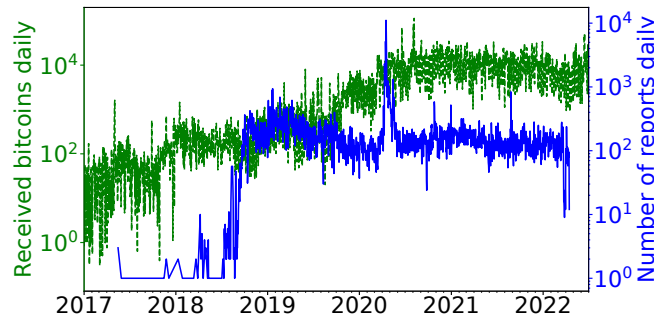


Figure 5.1: Timeline of received bitcoins and number of reports between Jan 2017 and July 2022.

5.1 Longitudinal timeline

High-level timeline: Figure 5.1 shows the daily number of reports between January 2017 and July 2022 (blue) together with the total daily number of bitcoins received by the reported set of addresses (green). We note that the Bitcoin Abuse Database was created in 2017 and gained popularity in late 2018 when it saw a steep rise in the number of reports. The daily report count has remained relatively steady (at an order of 100's per day) since the beginning of 2019, with exceptions for some temporal peaks and dips. The timeline of the number of received bitcoins per day is more concerning, as there has been a substantial (roughly 100x) increase from $O(100)$ to $O(10,000)$ of bitcoins transferred to these addresses per day over the three-year period that reporting has been relatively stable (i.e., 2019–2022).

Noteworthy spikes: There are several noteworthy spikes in the reporting. The biggest spike by far was observed on April 16th, 2020. On this day, 11K reports were filled, which is roughly 100 times more than the daily average (of 100) for the surrounding days. Our investigation revealed that many news articles around that time warned about a particular style of scam emails reported by both the US [48] and Australian [50] governments. In these emails, the attacker (falsely) claims that they have recorded the victim visiting an adult website, while also showing the victim one of their passwords (likely obtained from a leak) as part of the email.

Looking at the reports for this day, it is clear that a lot of the reports are talking about the same type of attack, matching the descriptions in the articles mentioned. We observed a mix of descriptions written by the victims themselves as well as copies of the emails they received. In many cases, the scammers asked for \$1,000 or \$2,000 to be paid in bitcoins and displayed one valid password belonging to the targeted victim as “proof” that they know the password.

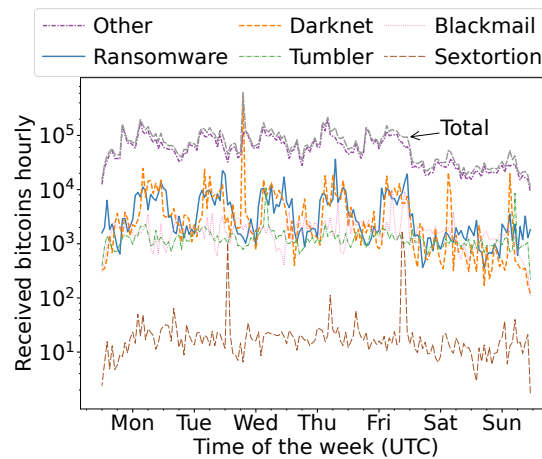


Figure 5.2: Received bitcoins by the time of the week. Weekday ticks show 12 pm UTC.

5.2 Time of the week

Figure 5.2 shows the time of the week that bitcoins were received for each category. This figure reveals both daily diurnal patterns (with much bigger volumes during daytime/evenings (UTC)) and more funds being transferred during weekdays than weekends. These observations are clearly seen by looking at the total volume transferred per hour (black line) in Figure 5.2 as well as the Ransomware, Darknet markets, and “Other” categories. In contrast, Bitcoin tumbler sees the least pronounced patterns, possibly suggesting some level of automation. Here it should be noted that Bitcoin tumbler typically aim at pooling and redistributing the funds at random intervals, with the aim to enhance the anonymity of Bitcoin and achieve effective money laundering.

Finally, the biggest relative spikes (due to specific events and campaigns) can be seen for Sextortion, Blackmail scams, and Darknet markets. Here we note that the Sextortion curve (due to smaller volumes) is more sensitive to larger transactions.

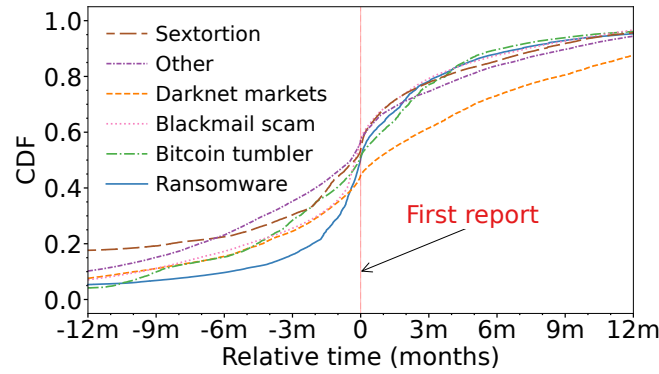


Figure 5.3: CDFs of the relative timing of the transactions compared to the first reporting date of an address.

5.3 Initial report date analysis

We next consider the timing of the payments to an address relative to the first time that the address was reported to Bitcoin Abuse. This is illustrated on a per-category basis in Figure 5.3. This shows the CDFs of the relative time each transaction was made in relation to the *first* time the address was reported.

This figure provides several interesting insights. First, addresses are reported around the time that their incoming transaction count is high, indicating that the first report often is made around the time of the abuse’s highest activity. This may be a reflection of a significant portion of the addresses only being used for specific attacks. However, here we observe significant differences between the categories, with Ransomware having the biggest concentration around the time of the address first being reported and Darknet markets being the least concentrated.

Second, for all categories except Darknet markets, most transactions take place before the first report is even filled. This may in part be a reflection of most victims not reporting addresses engaging in illicit behaviors.

While some might report the addresses somewhere other than Bitcoin Abuse, the large share of transactions before the first time an address is reported (to Bitcoin Abuse) also suggests that unless reporting behaviors change, there may be limited effectiveness to using such reports to “ban” addresses. In general, there is a need for a recognized central anti-fraud mechanism for cryptocurrencies. As we have seen here, without a centralized mechanism, malicious activity will exploit the gaps.

6 Following the money

In this section, we share insights learned from following the outflow of bitcoins from the reported addresses. We first discuss why, then how (Section 6.1), followed by a one-step and multi-step analysis (Section 6.2 and 6.3). The results and accompanying discussions will be presented. The more general and broader discussion is left for Chapter 7.

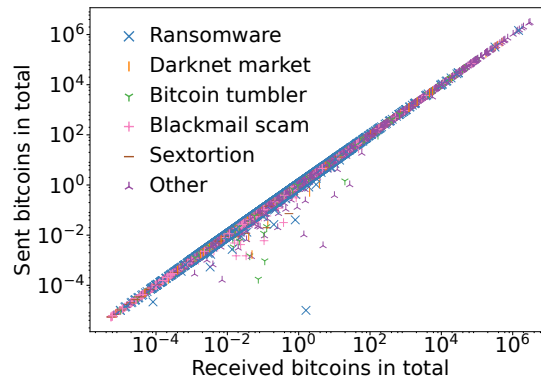


Figure 6.1: Scatterplot of received and sent bitcoins.

Bitcoins temporarily passing through reported addresses: This analysis is of particular interest since nearly all reported addresses have sent as many bitcoins as they received, leaving a balance of zero. This is illustrated in Figure 6.1, where we show the total number of bitcoins sent vs. received (per address). This shows that the bitcoins only temporarily pass through the reported addresses, suggesting that these addresses typically are not the wallets that the perpetrators use to store their ill-obtained monetary gains. The following section is dedicated to providing insights from following the money associated with different categories of reported addresses.

Scope of analysis: The main goal of this analysis is to study and compare the potential concentration or dispersion of money for different categories of abuse accounts, rather than trying to pinpoint people/organizations extracting or using the money (we leave such tasks for government agencies, police forces, and the like).

6.1 Following the money methodology

The analysis thus far has not required us to keep track of who has transferred funds to whom and was performed by simply counting the funds transferred. However, when following the money paid from one address to another (as done next), greater attention to detail is needed. We described the main challenge with this type of analysis in the background (Section 2.1). We next describe the decisions (and their limitations) we made to address this.

Challenging case and our solution: We remind ourselves that when there are multiple inputs *and* multiple outputs, the situation is like a melting pot and it is not obvious which bitcoins ended up where. In some cases, it is still possible to determine who transferred the funds to whom (with some accuracy) using heuristics based on the input/output sizes. However, to avoid introducing potential inaccuracies, for the analysis presented here we opted to not use any transactions matching this challenging case of our follow-the-money analysis (coming next) and instead only consider the transactions for which we are sure exactly who sent what bitcoins to whom. Fortunately, there are very many transactions that have limited inputs or outputs, allowing us to identify how funds flow between a series of Bitcoin addresses. (We again note that this limitation only is applied from here on and that it does not impact any of the analysis presented earlier in the thesis.)

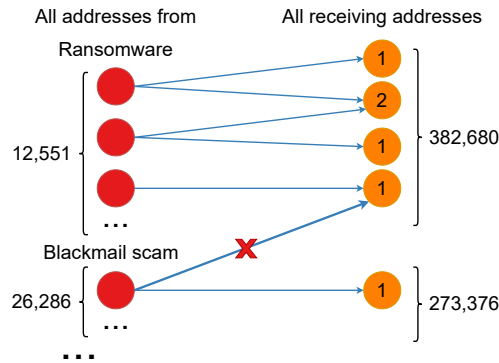


Figure 6.2: Visualization of the one-step analysis (the node in-degrees are in the orange circles).

Table 6.1: Address expansion ratio comparison of categories when going one step deep.

Category	Abuse addresses (#)	Addresses one level deep (#)	Expansion ratio
Sextortion	24,218	32,982	1.36
Blackmail scam	26,286	273,376	10.40
Ransomware	12,551	382,680	30.49
Darknet markets	1,289	168,542	130.75
Bitcoin tumbler	2,507	334,160	133.29
Other	13,736	1,419,902	103.37

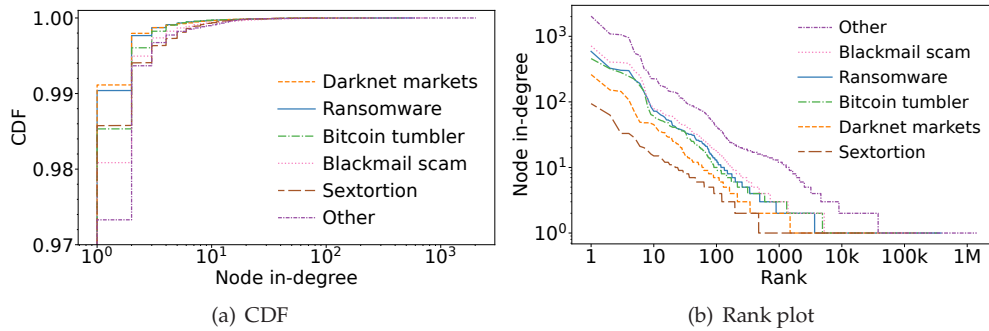


Figure 6.3: Node in-degrees distributions of receiving addresses being sent money from addresses in each category.

6.2 One-step concentration or dispersion

Let's first follow the outgoing money from the reported addresses only *one* step. In particular, we consider the concentration of addresses that the reported addresses transfer funds directly to. Figure 6.2 illustrates how we did this analysis. First, for each category, we added a link from the reported addressees (red circles on the left) belonging to that category to any address that it directly transferred funds to (conclusively). Second, for each receiving first-hop address (orange circles) we count and report how many reporting addresses each such address has received at least some funds from. This corresponds to the in-degree of each (orange) address to the right in the graph. For example, if two reported addresses both send money to address x , then address x has a node in-degree of two. Please note that this metric does not consider how many transactions an address x receives, only how many unique senders (in this case from the set of reported addresses) that it receives some funds from.

Figures 6.3(a) and 6.3(b) show the CDFs and rank plots, respectively, of the per-category node degrees. First, we note a Zipf-like distribution (i.e., relatively straight-line behavior in the rank plots shown on log-log scale), capturing a high skew among the nodes that do receive money flows from multiple abuse addresses. For example, each category has at least one address that obtained money from 100 or more abuse addresses, with one of the addresses of the "Other" category obtaining money from 2K abuse addresses. This behavior matches well with the reports about some of the addresses in the "Other" category being used by organized crime to pool funds from several attack vectors to make worldwide financial "investment in terror".

However, perhaps the main observation is the very long tail of addresses with node in-degree one. For example, looking at the CDF, 97-99% of the addresses associated with each category have a node in-degree of one (the minimum), meaning that nearly all receiving addresses are not visibly related when only tracing the money one step. This suggests that the money mostly is spread out across even more addresses after the first step, when going only one step deep from the reported addresses. This is confirmed when looking at the total addresses seen one level deep and the relative expansion ratio of each category, shown in Table 6.1. It is interesting to see the big differences in the expansion ratios, with Bitcoin tumbler, Darknet markets, and "Other" having much bigger ratios than the other categories.

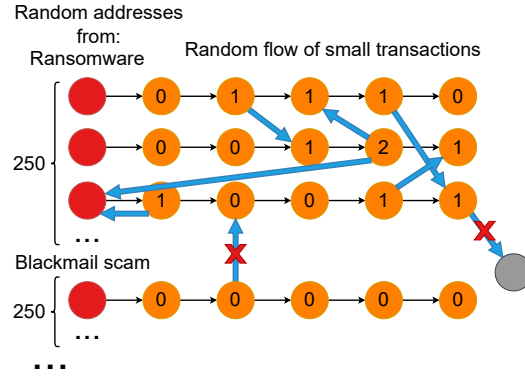


Figure 6.4: Multi-step analysis. Basic chain of transactions (black) and “concentration” edges (blue arrows).

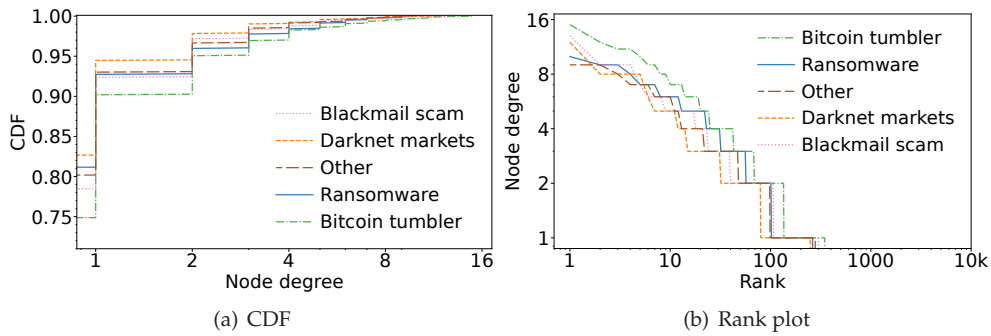


Figure 6.5: CDF and rank plot of the node degrees of addresses in the chain, for respective category, when counting only blue “concentration” edges.

6.3 Multi-step analysis

Having seen little concentration when following the transactions only one step, we next performed a multi-step analysis to track the money flow several steps deep. For this analysis, we again wanted to compare the categories fairly head-to-head and be able to answer questions such as whether some categories are more likely to shuffle around the money a couple of steps only to collect the money a few steps later.

For fair head-to-head comparison, we used a “tracking the pennies” approach in which we tracked an equal amount of “penny flows” (flow of small transactions) as bitcoins were moved five steps deep. This approach is not based on any, to us known, already established methods. Figure 6.4 provides a visual overview of our sampling and tracking methodology. More specifically, for each category, we used a random (depth first) search to find 250 randomly selected reported addresses from which we were able to trace back at least one random chain of money five steps deep from that address. For each step in this search, a new random transaction was chosen from the last address in the chain. We typically gave preference to a small transaction (under 0.1 bitcoins), and if no such transaction existed, we used any random transaction as a fallback. Finally, if a chain of transactions reached a dead end (i.e., where there are no more outgoing transactions), the latest address was removed from the chain and a new random transaction (new path) was chosen.

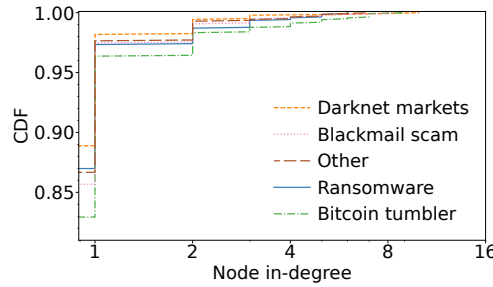
The choice of picking 250 was made to ensure that we have a substantial number of random paths for each category (so that numerical properties can be compared with some sta-

Table 6.2: Comparison of graph metrics and transaction metrics calculated on each category's flow graph.

Category	Graph metrics		Transaction metrics		
	Connecting edges	Loops	To reported	To reported category	To reported 250
Ransomware	254	4	2,526	399	169
Darknet	209	17	2,240	269	263
Tumbler	354	25	2,524	585	97
Blackmail	267	16	2,239	307	154
Other	247	10	5,164	3,155	1,299

tistical confidence). However, we note that this choice forces us to drop Sextortion from the analysis since we did not find 250 full five-step paths for this category. Therefore, the following analysis focuses only on the other five categories.

Furthermore, for the node degree analysis, we did not count all types of address relationships that we identify (but report on these separately). For example, as illustrated with an \times in Figure 6.4, we did not count the basic chain (black arrows), cross-category links (e.g., vertical arrow in the figure), or links to addresses further away than five steps deep.

Figure 6.6: CDF of the node *in*-degrees of addresses in the chain, for respective category, when counting only blue "concentration" edges.

Fewer addresses with the minimum in-degree: While all categories still had nodes with the minimum in-degree (one for one-step and zero for multi-step), compared to the one-step analysis, this fraction reduced noticeably (from 97-99% to 83-89%). See Figure 6.3(a) and 6.6. Here, Bitcoin tumbler saw the biggest reduction (from 98.5% down to 83%).

Bitcoin tumblers: In general, Bitcoin tumbler stands out in our multi-step analysis. For example, in addition to the above finding, it stands out with higher node degrees (e.g., see CDFs of "concentration" edges in Figure 6.5(a)) and substantially higher concentration (e.g., see the rank plot in Figure 6.5(b)) than the other categories. These findings suggest that fewer actors are involved with this category, typically used for anonymity/money laundering. These differences can perhaps be explained by the increased effort associated with running such addresses. In particular, we note that tumbling bitcoins typically requires more effort and expertise than simply sending blackmail scam emails (which has the lowest node degree and concentration of the categories).

Money-flow comparisons: When looking closer at the structure of the graphs formed by the 1,250 edges (250 chains \times 5 steps) we observed significant differences between the categories. The first two columns of Table 6.2 summarize some of these properties. Here, the "concentration edges" represent the blue arrows in Figure 6.4, which are transactions among the set of addresses in the graph, excluding the "penny flow" itself (the white arrows in Figure 6.4). "loops" measures the number of distinct cycles that exist in the graph structure for that category. Looking at these two metrics, Bitcoin tumbler again stands out with signifi-

Table 6.3: Transactions across categories.

From/to	Ransomware	Darknet	Tumbler	Blackmail	Other
Ransomware	-	982	2,658	1,566	1,228
Darknet	767	-	2,126	664	915
Tumbler	2173	1,638	-	2,282	2,588
Blackmail	1,512	914	4,087	-	2,230
Other	1,086	1,825	3,015	1,903	-

cantly higher “connecting edge” (354) and “loop” (25) counts than the other categories. This again matches the intuition that the addresses in this category are more likely to send money among a relatively smaller set of addresses. In contrast, Darknet markets have the fewest “connecting edges” (209) and Ransomware has by far fewest “loops” (4).

Transaction-based analysis on the graph: Finally, we have found that some of the edges goes back to the original reported addresses, and that these in some cases carry a non-negligible number of transactions. The remaining columns of Table 6.2 summarizes the metrics we used here, where “to reported” counts the number of transactions going back to any of the 267K reported addresses, “to reported category” counts transactions to any reported address in the category, and “to reported 250” only counts transactions back to the 250 randomly picked reported addresses of that category.

Here, we again observe some major differences between the categories. First, the “Other” category has a much higher number of transactions going back to reported addresses, especially to its own category (3,155 transactions compared to the 585 for the second-ranked category using this metric) as well as back to the 250 random (reported) addresses of its own category (1,299 compared to 263 of the second-ranked category). These findings suggest that a significant number of transactions are directed towards some of the reported addresses in the “Other” category.

Transactions across categories: To better understand how money flowed between addresses associated with the different categories, we next counted the transactions made between the subgraphs of each category (note that these were not included above, since we did not include that type of cross-category “connecting edges” in the original graph analysis (marked with \times in Figure 6.4). Table 6.3 summarizes the total number of transactions over such cross-category edges. Here, Bitcoin tumbler again stands out with both a higher inflow and outflow of transactions to/from the category compared to to/from the other categories. Given the nature of Bitcoin tumbler (money laundering), it also makes sense that other categories are interested in their service, which may explain why all categories have more outgoing transactions to Bitcoin tumbler to any of the other categories; e.g., Ransomware (2,658 vs. 1,566 for 2nd ranked), Darknet markets (2,126 vs. 915 for 2nd ranked), Blackmail scams (4,087 vs. 2,230 for 2nd ranked), and “Other” (3,015 vs. 1,903 for 2nd ranked).



7 Discussion

In this chapter, a broader discussion will be presented. For readability, the discussion of individual results is presented in conjunction with the results in Chapter 4, 5, and 6 respectively.

7.1 Results

While most of the results have already been discussed, some overall remarks will be made. First, we have witnessed staggering sums of money (\$815B) passing through the reported addresses. The magnitude of this number is not easily grasped. It could for example pay off Sweden's entire national debt, eight times over. That being said, turnover does not equal take-home profit. Considering the nature of the various abuse categories, the total received bitcoins might give an accurate hint of the magnitude of the profits for categories like Blackmail scam, Sextortion, and Ransomware, since the money flow is typically simple: the victim pays the abuser. Meanwhile, Darknet markets and Bitcoin tumblers have a more complex flow of money, where only a fraction of the turnover is likely to be take-home profits, since a Bitcoin tumbling service or a darknet market likely incurs a fee for their service, while most of the money passing through belongs to their clients.

Second, we note that the overall picture is clear: many people fall victim to Bitcoin abuse continuously, with no end in sight. It would be naive to assume that this situation could be solved with merely a bit more effort from the police/government. More likely, an overhaul of the current strategies is necessary, and local efforts might suffer since Bitcoin knows no borders. Tainted bitcoins will likely find their way to whatever jurisdiction looks the other way when it is time to convert them into fiat money.

7.2 Method

Large dataset

The dataset consisting of 267K reports and 83K unique addresses is considered sufficiently large to base the characterization on. That being said, it is unfortunate that we were unable to collect data from the most recent year due to the malfunctioning of the Bitcoin Abuse Database's API, and were not allowed to use Chainabuse's data. It would have been valuable to compare the results with that of Chainabuse's cleaned dataset, especially considering Chainabuse's more fine-grained abuse categorization.

Self-reported abuse types may be inaccurate

There are mainly two potential problems with the self-reported categorization of abuse. First, the reporter may easily be mistaken in their choice. All five abuse-type options that were available cannot be assumed to be common knowledge. There was no readily accessible explanation of all five abuse types on bitcoinabuse.com. For example, a victim that receives a blackmail scam email that claims to have recorded sexual activity on their webcam, may incorrectly report this as sextortion. Even though sextortion typically refers to real extortion based on nude or explicit images, and is not a blackmail *scam*. We simply cannot expect the average user to be aware of this distinction, as an example. This uncertainty may also lead many to simply choose "other" unnecessarily.

Second, a lot of abuse types were not available as an option. If we look at the top-10 highest receiving addresses in Figure 4.1 for example, we can see that financial scams and potentially organized crime may be worthy of categories themselves.

The new abuse database Chainabuse has created a more involved categorization with an accompanied quiz people can take to help them choose. Their categorization might be useful for any potential future work. Categories and sub-categories they use [12]:

- Blackmail: Sextortion scam, ransomware, and other blackmail scam.
- Fraud: NFT airdrop scam, donation impersonation scam, romance scam, pig butchering scam, fake project investment scam, rug pull investment scam, fake returns investment scam, other phishing scam, other impersonation scam, other fraud scam.
- Hack: Sim swap, contract exploit, other hack scam.

Collection of blockchain data

There are various ways to retrieve information from the blockchain and during the study several of them were explored, while the Blockchain.com API was ultimately used for the analysis presented. This API is deemed a credible source and the study's reliance on its results should not pose an issue. That being said, making use of two different methods/APIs to validate the accuracy of the retrieved information could potentially strengthen the credibility of the results. This was however not done, since time was deemed better spent on analyzing the dataset as well as possible.

Following the money

While our calculations of node degrees among addresses at one or multiple steps away were useful to gain insight into the magnitude of concentration versus dispersion of funds, more could be done to further characterize the outflow of funds, if time would have allowed it. For example, we have touched on the topic of address clustering heuristics in related work (Section 2.6). Efforts could be made to apply a carefully chosen heuristic to either strengthen our results or gain further insight into the outflow of money.

7.3 The work in a wider context

Our research highlights the significant challenges posed by Bitcoin’s reasonably anonymous transactions and lack of regulation, particularly in relation to its use in illicit activities. In the thesis, we provide a comprehensive analysis of Bitcoin transactions associated with different categories of illegal activities, conducted so as to ensure that ethical considerations were upheld.

Addresses an important societal problem: As the thesis demonstrates, Bitcoin use for illicit activities is clearly widespread and turns over very large sums of money. Many of the abuse types prey on the weak, and it is clear that these activities have increasingly negative societal effects. While the general public often focuses on Bitcoin’s energy consumption, much less attention has been put on the numerous victims that fall prey to Bitcoin’s role in various illicit activities. We note that while Bitcoin’s energy consumption *might* have long-term effects on humans, the effect of Bitcoin abuse on humans is both *apparent* and *current*.

Addressing the abuse is in Bitcoin’s interest: Addressing the abuse would also be of great value for those who see the value of the new technology Bitcoin brings to the table — a functioning peer-to-peer electronic cash system — since the abuse sheds a negative light on a neutral tool. The exchange of money outside of the authorities’ control is useful for people under tyrannical governments and the like. Widespread Bitcoin abuse might lead to bans that hinder the good uses of Bitcoin.

Respect privacy and confidentiality of individuals: The data used for analysis is sourced from the Bitcoin Abuse Database, which collects information from reports submitted by victims and other individuals or organizations, and the public Bitcoin chain itself. While the database provides insights into attacks and associated Bitcoin addresses, care is taken to ensure the anonymity of victims and attackers. We recognize that not all victims report their experiences and that reports may come from individuals who did not fall victim themselves.

Adhere to legal and ethical guidelines: The analysis focuses on publicly available blockchain data and information obtained from the Bitcoin Abuse Database. No attempts are made to compromise the security or integrity of the Bitcoin network or any other systems.

Promote responsible and ethical use of the findings: Law enforcement agencies, regulatory bodies, and cryptocurrency service providers can leverage our insights to enhance their strategies, policies, and compliance measures. Overall, the study aims to contribute to the development of effective strategies to mitigate the risks and vulnerabilities associated with Bitcoin’s potential for misuse, promoting a safer and more secure financial landscape.



8 Conclusion

This thesis presented a comprehensive analysis of money flows to and from Bitcoin addresses linked to different abuse categories. Our analysis uncovered valuable insights contributing to the understanding of Bitcoin transactions associated with illegal activities, that can shape future efforts in combating illicit activities in the cryptocurrency space and that have implications for legitimate users and stakeholders.

First, our high-level characterization of money flows revealed substantial variations in flow patterns, report rates, and success rates within and across addresses associated with different abuse categories (e.g., high skew, heavy tails, and big differences between categories). This understanding aids law enforcement and regulators in identifying patterns and trends in illegal Bitcoin activities, and improve their strategies to detect, prevent, and mitigate illicit activities.

Second, our temporal analysis captured long-term trends, weekly patterns, and the relative timing of when illicit addresses of each category are reported. The observed increase in the daily number of Bitcoins received by reported addresses over time (e.g., roughly 100 times over a three-year period) indicates a significant rise in funds transferred to these addresses. This finding calls for continuous vigilance and adaptive approaches to keep up with the evolving landscape of illegal transactions. Moreover, the variations in activity levels and transaction volumes during weekdays and weekends (especially for Ransomware, Darknet markets, and the "Other" category) highlight the importance of targeted enforcement efforts during periods of heightened activity.

Third, our analysis of the outflow of bitcoins from reported addresses sheds light on significant differences in graph properties and flow patterns among illicit addresses and between abuse categories. For example, the concentration of funds toward specific addresses, the dispersion of funds after the initial step, and the presence of loops highlight the complexity of money laundering schemes and the need for enhanced measures to track and disrupt these networks. The significant differences in graph structure and transaction patterns between categories, particularly the prominence of Bitcoin tumblers (used for money laundering), underscore the importance of addressing the role of specific services in facilitating illicit financial flows.

Finally, our results also highlight that authorities need a coordinated effort to monitor all Bitcoin activities, and cryptocurrency activities in general. While there is a lot of illicit activity on Bitcoin, only a fraction of it gets reported, and at least to Bitcoin Abuse many of the reports

come in relatively late. Having governments globally agree on an official centralized abuse monitoring effort for all cryptocurrencies is important to capture the money flows spanning different types of illicit activities and national borders. While we stay away from pointing to specific actors, as demonstrated by our follow-the-money analysis, sophisticated analysis techniques could also be used to help identify big threat actors that dabble in several types of illicit activities.




Bibliography

- [1] Jeff Alstott, Ed Bullmore, and Dietmar Plenz. “powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions”. In: *PLOS ONE* 9 (Jan. 2014), pp. 1–11. DOI: 10.1371/journal.pone.0085777.
- [2] Elli Androulaki, Ghassan O Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. “Evaluating user privacy in bitcoin”. In: *Proc. Financial Cryptography and Data Security (FC)*. 2013.
- [3] *API Documentation | BitcoinAbuse.com* — *bitcoinabuse.com*. <https://www.bitcoinabuse.com/api-docs>. [Accessed 15-May-2023].
- [4] Alex Biryukov and Ivan Pustogarov. “Bitcoin over Tor isn’t a good idea”. In: *Proc. IEEE Symposium on Security and Privacy (S&P)*. 2015.
- [5] *Bitcoin Abuse Database* — *bitcoinabuse.com*. <https://www.bitcoinabuse.com>. [Accessed 08-May-2023]. 2023.
- [6] *Bitcoin Historical Data - Investing.com* — *investing.com*. <https://www.investing.com/crypto/bitcoin/historical-data>. [Accessed 17-May-2023].
- [7] Blockchain. *Blockchain Developer APIs*. <https://www.blockchain.com/explorer/api>. [Accessed 25-May-2023]. 2023.
- [8] Joseph Bonneau, Arvind Narayanan, Andrew Miller, Jeremy Clark, Joshua A Kroll, and Edward W Felten. “Mixcoin: Anonymity for bitcoin with accountable mixes”. In: *Proc. Financial Cryptography and Data Security (FC)*. 2014.
- [9] J. Broséus, D. Rhumorbarbe, C. Mireault, V. Ouellette, F. Crispino, and D. Décary-Héту. “Studying illicit drug trafficking on Darknet markets: Structure and organisation from a Canadian perspective”. In: *Forensic Science International* 264 (2016). Special Issue on the 7th European Academy of Forensic Science Conference, pp. 7–14.
- [10] *Chainabuse - About* — *chainabuse.com*. <https://www.chainabuse.com/about>. [Accessed 15-May-2023].
- [11] *Chainabuse - Home* — *chainabuse.com*. <https://www.chainabuse.com/>. [Accessed 08-May-2023]. 2023.
- [12] *Chainabuse - New report* — *chainabuse.com*. <https://www.chainabuse.com/report>. [Accessed 18-May-2023].

-
- [13] *Chainabuse Public API (v1.2)* — *docs.chainabuse.com*. <https://docs.chainabuse.com/docs>. [Accessed 15-May-2023].
 - [14] Nicolas Christin. “Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace”. In: *Proceedings of the International Conference on World Wide Web (WWW)*. 2013.
 - [15] Mauro Conti, Ankit Gangwal, and Sushmita Ruj. “On the economic significance of ransomware campaigns: A Bitcoin transactions perspective”. In: *Computers & Security* (2018).
 - [16] *File Bitcoin Abuse Report | BitcoinAbuse.com* — *bitcoinabuse.com*. <https://www.bitcoinabuse.com/reports/create>. [Accessed 15-May-2023].
 - [17] Juan Garay, Aggelos Kiayias, and Nikos Leonardos. “The bitcoin backbone protocol: Analysis and applications”. In: *Proc. EUROCRYPT*. 2015.
 - [18] Gibran Gomez, Pedro Moreno-Sanchez, and Juan Caballero. “Watch Your Back: Identifying Cybercrime Financial Relationships in Bitcoin through Back-and-Forth Exploration”. In: *Proc. ACM Computer and Communications Security (CCS)*. 2022.
 - [19] Martin Harrigan and Christoph Fretter. “The unreasonable effectiveness of address clustering”. In: *Proc. UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld*. 2016.
 - [20] Jordi Herrera-Joancomarti. “Research and challenges on bitcoin anonymity”. In: *Proc. Workshop on Data Privacy Management (DPM)*. 2015.
 - [21] Danny Yuxing Huang, Maxwell Matthaios Aliapoulos, Vector Guo Li, Luca Invernizzi, Elie Bursztein, Kylie McRoberts, Jonathan Levin, Kirill Levchenko, Alex C Snoeren, and Damon McCoy. “Tracking ransomware end-to-end”. In: *Proc. IEEE Symposium on Security and Privacy (S&P)*. 2018.
 - [22] Danny Yuxing Huang, Hitesh Dharmdasani, Sarah Meiklejohn, Vacha Dave, Chris Grier, Damon McCoy, Stefan Savage, Nicholas Weaver, Alex C Snoeren, and Kirill Levchenko. “Botcoin: Monetizing stolen cycles”. In: *Proc. Network and Distributed System Security Symposium (NDSS)*. 2014.
 - [23] International Monetary Fund. *IMF.org*. <https://www.imf.org/en/Publications/WEO/weo-database/2023/October/weo-report?c=512,914,612,171,614,311,213,911,314,193,122,912,313,419,513,316,913,124,339,638,514,218,963,616,223,516,918,748,618,624,522,622,156,626,628,228,924,233,632,636,634,238,662,960,423,935,128,611,321,243,248,469,253,642,643,939,734,644,819,172,132,646,648,915,134,652,174,328,258,656,654,336,263,268,532,944,176,534,536,429,433,178,436,136,343,158,439,916,664,826,542,967,443,917,544,941,446,666,668,672,946,137,546,674,676,548,556,678,181,867,682,684,273,868,921,948,943,686,688,518,728,836,558,138,196,278,692,694,962,142,449,564,565,283,853,288,293,566,964,182,359,453,968,922,714,862,135,716,456,722,942,718,724,576,936,961,813,726,199,733,184,524,361,362,364,732,366,144,146,463,528,923,738,578,537,742,866,369,744,186,925,869,746,926,466,112,111,298,927,846,299,582,487,474,754,698,&s=NGDPD,&sy=2021&ey=2028&ssm=0&scsm=1&scc=0&ssd=1&ssc=0&sic=0&sort=country&ds=.&br=1>. [Accessed 15-Oct-2023]. 2023.
 - [24] Ghassan O. Karame, Elli Androulaki, and Srdjan Capkun. “Double-Spending Fast Payments in Bitcoin”. In: *Proceedings of the ACM Computer and Communications Security (CCS)*. 2012.

- [25] Amin Kharraz, William Robertson, Davide Balzarotti, Leyla Bilge, and Engin Kirda. "Cutting the gordian knot: A look under the hood of ransomware attacks". In: *Proceedings of the Detection of Intrusions and Malware, and Vulnerability Assessment: International Conference (DIMVA)*. Springer. 2015, pp. 3–24.
- [26] Seunghyeon Lee, Changhoon Yoon, Heedo Kang, Yeonkeun Kim, Yongdae Kim, Dongsu Han, Soeul Son, and Seungwon Shin. "Cybercriminal minds: an investigative study of cryptocurrency abuses in the dark web". In: *Proc. Network and Distributed System Security Symposium (NDSS)*. 2019.
- [27] Xigao Li, Anurag Yepuri, and Nick Nikiforakis. "Double and Nothing: Understanding and Detecting Cryptocurrency Giveaway Scams". In: *Proc. Network and Distributed System Security Symposium (NDSS)*. 2023.
- [28] Kevin Liao, Ziming Zhao, Adam Doupé, and Gail-Joon Ahn. "Behind closed doors: measurement and analysis of CryptoLocker ransoms in Bitcoin". In: *Proc. Electronic Crime Research (eCrime)*. 2016.
- [29] Frank J Massey Jr. "The Kolmogorov-Smirnov test for goodness of fit". In: *Journal of the American statistical Association* (1951).
- [30] *Matplotlib x2014; Visualization with Python — matplotlib.org*. <https://matplotlib.org/>. [Accessed 29-May-2023].
- [31] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. "A Fistful of Bitcoins: Characterizing Payments among Men with No Names". In: *Commun. ACM* 59.4 (2016), pp. 86–93.
- [32] Malte Möser, Rainer Böhme, and Dominic Breuker. "An inquiry into money laundering tools in the Bitcoin ecosystem". In: *2013 APWG eCrime researchers summit*. IEEE. 2013, pp. 1–14.
- [33] InJae Myung. "Tutorial on maximum likelihood estimation". In: *Journal of Mathematical Psychology* (2003).
- [34] Satoshi Nakamoto. *Bitcoin: A Peer-to-Peer Electronic Cash System*. [Accessed 08-May-2023]. 2008. URL: <https://bitcoin.org/bitcoin.pdf>.
- [35] *NetworkX 2014; NetworkX documentation — networkx.org*. <https://networkx.org/>. [Accessed 29-May-2023].
- [36] *NumPy — numpy.org*. <https://numpy.org/>. [Accessed 29-May-2023].
- [37] Frédérique Oggier, Anwitaman Datta, and Silivanxay Phetsouvanh. "An ego network analysis of sextortionists". In: *Social Network Analysis and Mining* (2020).
- [38] *pandas - Python Data Analysis Library — pandas.pydata.org*. <https://pandas.pydata.org/>. [Accessed 29-May-2023].
- [39] Masarah Paquet-Clouston, Matteo Romiti, Bernhard Haslhofer, and Thomas Charvat. "Spams Meet Cryptocurrencies: Sextortion in the Bitcoin Ecosystem". In: *Proceedings of the ACM Conference on Advances in Financial Technologies (AFT)*. 2019, pp. 76–88.
- [40] Sergio Pastrana, Alice Hutchings, Daniel Thomas, and Juan Tapiador. "Measuring EWhoring". In: *Proc. ACM Internet Measurement Conference (IMC)*. 2019.
- [41] Sergio Pastrana and Guillermo Suarez-Tangil. "A First Look at the Crypto-Mining Malware Ecosystem: A Decade of Unrestricted Wealth". In: *Proc. ACM Internet Measurement Conference (IMC)*. 2019.
- [42] Stijn Pletinckx, Cyril Trap, and Christian Doerr. "Malware coordination using the blockchain: An analysis of the cerber ransomware". In: *Proc. IEEE Communications and Network Security (CNS)*. 2018.

- [43] Rebecca S Portnoff, Danny Yuxing Huang, Periwinkle Doerfler, Sadia Afroz, and Damon McCoy. "Backpage and bitcoin: Uncovering human traffickers". In: *Proc. Knowledge Discovery and Data Mining (KDD)*. 2017.
- [44] Fergal Reid and Martin Harrigan. *An analysis of anonymity in the bitcoin system*. Springer, 2013.
- [45] Dorit Ron and Adi Shamir. "How did dread pirate roberts acquire and protect his bitcoin wealth?". In: *Proc. Financial Cryptography and Data Security (FC)*. 2014.
- [46] Hampus Rosenquist, David Hasselquist, Martin Arlitt, and Niklas Carlsson. "On the Dark Side of the Coin: Characterizing Bitcoin use for Illicit Activities". In: *Proceedings of the Passive and Active Measurement Conference (PAM)*. Mar. 2024.
- [47] Tim Ruffing, Pedro Moreno-Sanchez, and Aniket Kate. "Coinshuffle: Practical decentralized coin mixing for bitcoin". In: *Proc. ESORICS*. 2014.
- [48] *Scam emails demand Bitcoin, threaten blackmail* — [consumer.ftc.gov](https://consumer.ftc.gov/consumer-alerts/2020/04/scam-emails-demand-bitcoin-threaten-blackmail). <https://consumer.ftc.gov/consumer-alerts/2020/04/scam-emails-demand-bitcoin-threaten-blackmail>. [Accessed 17-May-2023]. 2020.
- [49] *SciPy documentation x2014; SciPy v1.10.1 Manual* — [docs.scipy.org](https://docs.scipy.org/doc/scipy/). <https://docs.scipy.org/doc/scipy/>. [Accessed 31-May-2023].
- [50] *Sextortion email campaign impacting Australians* — [cyber.gov.au](https://www.cyber.gov.au/about-us/alerts/sextortion-email-campaign-impacting-australians). <https://www.cyber.gov.au/about-us/alerts/sextortion-email-campaign-impacting-australians>. [Accessed 17-May-2023]. 2020.
- [51] Michele Spagnuolo, Federico Maggi, and Stefano Zanero. "Bitiodine: Extracting intelligence from the bitcoin network". In: *Proc. Financial Cryptography and Data Security (FC)*. 2014.
- [52] Tsuyoshi Taniguchi, Harm Griffioen, and Christian Doerr. "Analysis and takeover of the bitcoin-coordinated pony malware". In: *Proc. ACM Asia Computer and Communications Security (ASIACCS)*. 2021.
- [53] Ege Tekiner, Abbas Acar, A Selcuk Uluagac, Engin Kirda, and Ali Aydin Selcuk. "SoK: Cryptojacking malware". In: *Proc. IEEE European Symposium on Security and Privacy (EuroS&P)*. 2021.
- [54] U.S. Bureau of Economic Analysis (BEA). *GDP by State*. <https://www.bea.gov/data/gdp/gdp-state>. [Accessed 15-Oct-2023]. 2023.
- [55] Luke Valenta and Brendan Rowan. "Blindcoin: Blinded, accountable mixes for bitcoin". In: *Proc. Financial Cryptography and Data Security (FC)*. 2015.
- [56] Anh V Vu, Jack Hughes, Ildiko Pete, Ben Collier, Yi Ting Chua, Ilia Shumailov, and Alice Hutchings. "Turning up the dial: the evolution of a cybercrime market through set-up, stable, and covid-19 eras". In: *Proc. ACM Internet Measurement Conference (IMC)*. 2020.
- [57] Kai Wang, Jun Pang, Dingjie Chen, Yu Zhao, Dapeng Huang, Chen Chen, and Weili Han. "A Large-Scale Empirical Analysis of Ransomware Activities in Bitcoin". In: *ACM Trans. Web* 16.2 (2021).



A Appendix

Table A.1: Comparison of recent reporting rates and the volume of new addresses being reported in the past five months, and the transactions they receive.

	Primary dataset	Latest reports
Reports time frame	2017-05-16 to 2022-04-25	2022-12-20 to 2023-05-19
Reports	267,708	17,116
Unique addresses	82,527	2,249
Transactions	5,092,489	141,485
Received bitcoins	31,346,586	269,070
Received in USD	815,011,236,000	6,995,820,000

A.1 Additional statics

Table A.1 provides a high-level overview of the recent reporting rates, the volume of new addresses having been reported in the past five months, and the transactions they receive. We note that reporting rates (114 reports/day on average) are similar to those observed in Figure 5.1 and that the average transactions/address is almost the same (61.7 for primary vs. 62.9 for latest dataset). The main difference is a reduction in the average number of bitcoins received/address (380 for the primary dataset vs. 120 for the latest dataset) and the rate at new unique addresses are observed. These later differences are easily explained by (1) these addresses being earlier in their lifecycle (e.g., Figure 5.3) and/or (2) a bias towards many of the most successful addresses (in terms of attracting funds; e.g., the top-hitters in the “Other” category) already having been reported. Yet, the large number of funds that these *newly reported* addresses obtain shows that there continually are many more (new) illicit addresses being reported that are attracting significant funds, including at the present moment.

A.2 Listings

Listing A.1: Example of address info from blockchain.com.

```

1 {
2   "hash160": "{redacted}",
3   "address": "{redacted}",
4   "n_tx": 1,
5   "n_unredeemed": 1,
6   "total_received": 3551,
7   "total_sent": 0,
8   "final_balance": 3551,
9   "txs": [
10    {
11      "hash": "{redacted}",
12      "ver": 1,
13      "vin_sz": 1,
14      "vout_sz": 1,
15      "size": 192,
16      "weight": 767,
17      "fee": 2433,
18      "relayed_by": "0.0.0.0",
19      "lock_time": 0,
20      "tx_index": {redacted},
21      "double_spend": false,
22      "time": {redacted},
23      "block_index": {redacted},
24      "block_height": {redacted},
25      "inputs": [
26        {
27          "sequence": {redacted},
28          "witness": "",
29          "script": "{redacted}",
30          "index": 0,
31          "prev_out": {
32            "addr": "{redacted}",
33            "n": 0,
34            "script": "{redacted}",
35            "spending_outpoints": [
36              {
37                "n": 0,
38                "tx_index": {redacted}
39              }
40            ],
41            "spent": true,
42            "tx_index": {redacted},
43            "type": 0,
44            "value": 5984
45          }
46        ]
47      ],
48      "out": [
49        {
50          "type": 0,
51          "spent": false,
52          "value": 3551,
53          "spending_outpoints": [],
54          "n": 0,
55          "tx_index": {redacted},
56          "script": "{redacted}",
57          "addr": "{redacted}"
58        }
59      ],
60      "result": 3551,
61      "balance": 3551
62    }
63  ]
64 }

```