TRACKING MALICIOUS TRANSACTIONS IN CRYPTOCURRENCIES

Bhavish Dhanda
Supervisor: Dr Hassan Asghar
Co-Supervisor: Dr Benjamin Zhao

Department of Computing and Engineering
Macquarie University
Australia

ACKNOWLEDGEMENTS

STATEMENT OF CANDIDATE

Contents

List of Figures

Listings

1 Abstract

Cryptocurrencies form what is called the new decentralized finance world of digital finances. They have become increasingly popular in the recent few years and with this technology becoming more and more accessible, the number of users of this form of payments have grown exponentially.On of the other side of coin, this mass usage of this technology gives a small percentage of people of hide in plain sight and do illegal activities. These people who are involved in such malicious activites are also supported by the anonymity features of this form of payments.

2 Introduction

Cryptocurrencies have become the modern form of decentralized finance giving the general public a level of visibility and control over financial systems that we could never imagine while using traditional financial systems. Their adoption has increased exponentially in the last few years with a few countries even deciding to make it a legal tender for their day to day transactions. Such countries include but are not limited to Central African Republic, El Salvador [**?**] where Bitcoin has been deemed as a legal form of tender by the local authorities.

The striking difference between so-called decentralized finance[**?**] formed by cryptocurrencies and traditional banking systems the type of ledger they maintain, and how the currency is controlled. Decentralized as the word suggests means that the control is not the in hands of a single person or entity rather the entire network, in this case, all the people using and servicing the cryptocurrency influence all the decisions that are made which include, the price largely controlled by supply and demand and the number of people mining and other factors, validation of transactions is normally done by validation nodes inside the networks and similar functions all are done by the public nodes inside the network. The second major factor that contributes to this decentralization is the public ledger these currencies maintain in contrast to a centralized secured ledger maintained by traditional banking systems where only certain authorized users can access the ledger, in the case of cryptocurrencies, anyone can access the ledger and see all sorts of information stored in a typical currency ledger which includes account balances, transactions made etc.

Our goal in this paper is to find ways to track down on these malicious transactions and to try and get as close as possible to the real world entity behind these activities.

3 Background

3.1 What are cryptocurrencies

Cryptocurrencies are a modern form of digital currency which

3.2 How do cryptocurrencies work?

Before we actually start exploring our area of research we need to be aware of how cryptocurrencies techincally operate. To simplify things we will be using Bitcoin as our reference to explain the technical architecture. Blockchain could be any is essentially is a LinkedList [] of Blocks put one after another. Each Block inside the blockchain is a primarily a collection of transaction with someother metadata associated. Modern cryptocurrencies work on maintaining a public ledger which is the distinguishing factor of these currencies with the standard paper currencies that we have been using for ages.

3.3 What exactly are malicious transactions?

In this paper we will at times refer to a transaction or address being malicious. What we mean when are saying that an address or transaction is malicious, is that the particular address or the addresses involved in a particular transaction are trying to do something illegal. These illegal activites can include money laundering, malware, scams, ponzi schemes amongst other ways to cause harm to the general public. The basic intention over here to signify that the entity controlling the address or performing a transaction is doing something which would be deemed illegal in the context of local laws and is thereby prohibited by law.

[**?**]

4 Related Work

In this section we will be going through the work that has been previously done in this area of research. As we will go through this section we will realise that a lot of people are already working to break through this issue of governance in cryptocurrencies. As we saw in the last section that illegal and malicious transactions in cryptocurrencies is an ever increasing problem so its very important that some action is taken to add some sort of regulation.

Muller et al[**?**]

5 Methodology

5.1 Introduction

Since there is not much information about cryptocurrencies except the information in the public ledger we will need to look outside that scope and see what information we can find about addresses and transactions to make some justified conclusion so as to eventually draw a suspicion explaining how is the address malicious and the conclusions we are making to justify its link with the real world entity.

The aim over here is now to find a target address either directly or through a transaction and find all the information I can related to it. We have shortlisted a few sources which keep records of information reported by the public, so we will need to scrape through all that information and its related reports to see if we can find something which can identify the person itself. If not we will need to move to the neighbour address and perform a similar search to observe the behaviour of the neighbour address and draw some conclusions.

The next plan is to identify a few strategic address whose physical location we are aware of with complete certainty and try and find interactions with those addresses.

5.2 Step 1 : The target address

The first step of the process for us is the data collection. For data collected there are couple of sources are we are using which includes a number of public APIs. We are using the public APIs to start with to get 2 different types of data.

5.3 Step 2

Collection of metadata about the address or transaction. This is the tricky part for us since our research purely relies on data publically available which can be on the blockchain or information reported by private entitites on the interent. This information is very limited specially if we pick up a random address for our investigation and the reason for this is due to the pseudoanonymous nature of cryptocurrencies. In cryptocurrencies a person can essentially set up as many addresses or so called account numbers as he wishes without any personally identifiable information in contrast to traditional banking system where a proper KYC is mandatory before one can use those services. Due to this the only information readily available is the public key of the user using the services which is his wallet address. What we can do is use that wallet address to query different sources to get as much information as possible about it. For this we be utilizing a mix of API services and also build a web scraping service to extract information no available from API due to reasons highlighted below.

In case our starting point is a transaction we will be targeting one of the addresses in the transaction. The information about getting the addresses from

the transaction is available readily on the internet thorugh APIs.

Our application can work both on an address or a transaction so our initial call will depend on what we are targeting.

To extract the address from the transaction we can utilize the public API available from Blockchain.info[**?**].

Blockchain.info ..... **Information about this source V IMP**

Url : GET `https://blockchain.info/rawtx/{{tx_hash_or_id}}` Every transaction in the blockchain gets a hash value when generated and verified so this API takes in that as a path variable and returns the response below for a sample random transaction we picked up. The data model for the response is also shown in figure [**?**].

Sample Response:

Listing 1: A single transaction from the blockchain


```
{
    "hash": "
       d16f321d0ed28dadb8310948aeaba0906380c605fe04b592788f3eab1a9f1a92
       ",
    "ver": 2,
    "vin_sz": 1,
    "vout_sz": 1,
    "size": 288,
    "weight": 534,
    "fee": 258,
    "relayed_by": "0.0.0.0",
    "lock_time": 733593,
    "tx_index": 5140577156163537,
    "double_spend": false,
    "time": 1650944563,
    "block_index": 733594,
    "block_height": 733594,
    "inputs": [
        {
            "sequence": 1,
            "witness": "032076
               f49003207b8bf9a8451f080c6ebe7d4d298d40e1f4f82a8b4c838446922c72473
               ",
            "script": "",
            "index": 0,
            "prev_out": {
                "tx_index": 8757097163138133,
                "value": 500000,
                "n": 0,
                "type": 0,
                "spent": true,
```

```
                "script": "0020
                    c720308c4d9f87406d5abb087ce71384839b2d963c439ee149e6dcbe5cd18
                    ",
                "spending_outpoints": [
                    {
                        "tx_index": 5140577156163537,
                        "n": 0
                    }
                ],
                "addr": "
                    bc1qcusrprzdn7r5qm26hvy8eecnsjpektvk83peac2fumwtuhx33adqrjrpm
                    "
            }
        }
    ],
    "out": [
        {
            "type": 0,
            "spent": true,
            "value": 499742,
            "spending_outpoints": [
                {
                    "tx_index": 869868793508427,
                    "n": 1
                }
            ],
            "n": 0,
            "tx_index": 5140577156163537,
            "script": "0014
                d78cb1e16955806504b5b8351b89b59cc5929520",
            "addr": "
                bc1q67xtrctf2kqx2p94hq63hzd4nnze99fq978f0s
                "
        }
    ]
}
```


As we can see from the sample response we can directly access the input and output address as our target and move on to the next step which is getting data about the individual address itself.

Now the task for us is to pick up the address in question and identify as much information as possible about it before we move on.

Firstly we will be utilizing an API from Blockchain.info[**?**] to get basic information about the address. The API below from this source gives us **Url : GET** `https://blockchain.info/rawaddr/{{address_id}}`

This API takes in the wallet address which is the publicly available on the

blockchain as a path variable and gives all the information about it directly from the blockchain. The data structure of the response is shown in the figure[] from which we will observe the interactions of the address inside the blockchain.

Sample Response:

Listing 2: A single address from the blockchain


```
{
    "hash160": "5e9b23809261178723055968d134a947f47e799f
        ",
    "address": "19dENFt4wVwos6xtgwStA6n8bbA57WCS58",
    "n_tx": 10638,
    "n_unredeemed": 160,
    "total_received": 6672872758903,
    "total_sent": 6609785308209,
    "final_balance": 63087450694,
    "txs": [
        {
            "hash": "68
                a0b7e0ee87f580fbf17afefdf8e397efac32215d47cd02f52496f288247d8f
                ",
            "ver": 2,
            "vin_sz": 1,
            "vout_sz": 2,
            "size": 217,
            "weight": 760,
            "fee": 0,
            "relayed_by": "0.0.0.0",
            "lock_time": 0,
            "tx_index": 5048564692439748,
            "double_spend": false,
            "time": 1660349777,
            "block_index": 749197,
            "block_height": 749197,
            "inputs": [
                {
                    "sequence": 4294967295,
                    "witness":
                        "012000000000000000000000000000000000000000000000000000000000000000
                    "script": "038
                        d6e0b0451edf6622f466f756e64727920555341205 06f6f6c20236472
                        ",
                    "index": 0,
                    "prev_out": {
                        "tx_index": 0,
                        "value": 0,
```

```
                        "n": 4294967295,
                        "type": 0,
                        "spent": true,
                        "script": "",
                        "spending_outpoints": [
                            {
                                "tx_index":
                                    5048564692439748,
                                "n": 0
                            }
                        ]
                    }
                }
            ],
            "out": [
                {
                    "type": 0,
                    "spent": false,
                    "value": 638621401,
                    "spending_outpoints": [],
                    "n": 0,
                    "tx_index": 5048564692439748,
                    "script": "76
                        a9145e9b23809261178723055968d134a947f47e799f88ac
                        ",
                    "addr": "19
                        dENFt4wVwos6xtgwStA6n8bbA57WCS58"
                },
                {
                    "type": 0,
                    "spent": false,
                    "value": 0,
                    "spending_outpoints": [],
                    "n": 1,
                    "tx_index": 5048564692439748,
                    "script": "6
                        a24aa21a9ed8855481a2704f6e6220201c7311fda2128b81689a33a4a
                        "
                }
            ],
            "result": 638621401,
            "balance": 63087450694
        },
        {
            "hash": "
                dd2dde28f1b7a33163f959b5a20c4670554542dc040fe03c1dcb73c7456dddce
```

```
                ",
            "ver": 2,
            "vin_sz": 1,
            "vout_sz": 2,
            "size": 217,
            "weight": 760,
            "fee": 0,
            "relayed_by": "0.0.0.0",
            "lock_time": 0,
            "tx_index": 7278413324283513,
            "double_spend": false,
            "time": 1660349287,
            "block_index": 749196,
            "block_height": 749196,
            "inputs": [
                {
                    "sequence": 4294967295,
                    "witness":
                        "012000000000000000000000000000000000000000000000000000000000000000
                    "script": "038
                        c6e0b0468ebf6622f466f756e6472792055534120506f6f6c20236472
                        ",
                    "index": 0,
                    "prev_out": {
                        "tx_index": 0,
                        "value": 0,
                        "n": 4294967295,
                        "type": 0,
                        "spent": true,
                        "script": "",
                        "spending_outpoints": [
                            {
                                "tx_index":
                                    7278413324283513,
                                "n": 0
                            }
                        ]
                    }
                }
            ],
            "out": [
                {
                    "type": 0,
                    "spent": false,
                    "value": 645134567,
```

```
                    "spending_outpoints": [],
                    "n": 0,
                    "tx_index": 7278413324283513,
                    "script": "76
                        a9145e9b2380926117872305596 8d134a947f47e799f88ac
                        ",
                    "addr": "19
                        dENFt4wVwos6xtgwStA6n8bbA57WCS58"
                },
                {
                    "type": 0,
                    "spent": false,
                    "value": 0,
                    "spending_outpoints": [],
                    "n": 1,
                    "tx_index": 7278413324283513,
                    "script": "6
                        a24aa21a9ed668b12b36b8b1e107f9190d36bea153086e2e32e119123
                        "
                }
            ],
            "result": 645134567,
            "balance": 62448829293
        },
        {
            "hash": "5312
                aff6f09a1366c7b24036967ce687636dafa2759228941548a6a4aa688a48
                ",
            "ver": 2,
            "vin_sz": 1,
            "vout_sz": 2,
            "size": 217,
            "weight": 760,
            "fee": 0,
            "relayed_by": "0.0.0.0",
            "lock_time": 0,
            "tx_index": 2552297558414537,
            "double_spend": false,
            "time": 1660348141,
            "block_index": 749193,
            "block_height": 749193,
            "inputs": [
                {
                    "sequence": 4294967295,
                    "witness":
                        "012000000000000000000000000000000000000000000000000000000000000
```

```
151                     "script": "03896
                            e0b04e7e6f6622f466f756e64727920555341205 06f6f6c202364726f
                            ",
152                     "index": 0,
153                     "prev_out": {
154                         "tx_index": 0,
155                         "value": 0,
156                         "n": 4294967295,
157                         "type": 0,
158                         "spent": true,
159                         "script": "",
160                         "spending_outpoints": [
161                             {
162                                 "tx_index":
                                        2552297558414537,
163                                 "n": 0
164                             }
165                         ]
166                     }
167                 }
168             ],
169             "out": [
170                 {
171                     "type": 0,
172                     "spent": false,
173                     "value": 630066392,
174                     "spending_outpoints": [],
175                     "n": 0,
176                     "tx_index": 2552297558414537,
177                     "script": "76
                            a9145e9b23809261178723055968d134a947f47e799f88ac
                            ",
178                     "addr": "19
                            dENFt4wVwos6xtgwStA6n8bbA57WCS58"
179                 },
180                 {
181                     "type": 0,
182                     "spent": false,
183                     "value": 0,
184                     "spending_outpoints": [],
185                     "n": 1,
186                     "tx_index": 2552297558414537,
187                     "script": "6
                            a24aa21a9ed0c4d24d3af3a82f76a0979c986824a0a1990b2a1eacbe
                            "
```

```
                }
            ],
            "result": 630066392,
            "balance": 61803694726
        },
        {
            "hash": "More transactions depending on the
                address"
        }
    ]
}
```

The main information we need to extract from here is the interactions of the address throughout the blockchain. As we can see from the sample response we get a list of transactions in the **txs** object though which we can loop through to get to the immediate neighbours of the address. We refer immediate neighbours to addresses which directly interact with the address though a transaction where our target address can either be the sender or the receiver.

5.4 Step 3: Web Scraping

Now that we have most of the metadata about the address we need to understand the nature of the address, we will starting digging into the address to find links to the real world. For this we will be building a web scraping service to go through a couple of identified online sources which seem have data which can help us get as close to the real world entity as possible.

When it comes to web scraping we will be downloading the entire page as an HTML instead of using APIs and then parsing the page element by element. This approach has both advantages and dis-advantages but in our cases the advantages outweigh the negatives so we are using this approach. Firstly when it comes to APIs, these sources are not best known to provide these services so getting an API key is very hard. We applied for an API key but its been over months since we have got anything so we decided to use the web crawling service as it will also give us more control over what data we can process. And finally APIs always have this security throttling feature against Denial of Service attacks so we can bypass this while scraping the web as we are acting as if a lot of users are accessing the web page which is a very normal behaviour and most websites do not have any restriction to the number of users who can access their page.

The sources we are targeting right now are

1. Bitcoin Who is Who [**?**]
2. Hash XP [**?**]

In the sections below we will go though in detail the data provided by each source in detail but breifly these 2 sources provide nearly the same information except in some case one is known to have more information about the address

than the other so it is always good to have 2 sources and also helps in better data validation. In the end we cannot completely rely on this information as this is publicly sourced and there are high chances of false positives as well so we have to take a cautious approach when analyzing this data and making further conclusions.

In short these sources provide some very good insights about an address such as the basic information from the blockchain such as balance(s), transaction(s), but the main data we are interested in from these sources is the scam alert which we will be scraping using our service. These are reports submitted by people from all over the world wherein an address was known to perform a malicious activity. We will analyzing that data to make further conclusions about the address.

In the web scraping service what we are doing is starting to get data about the actual address itself and then moving on to the neighbours of the address identified earlier through it transactions and getting data about them to observe and understand the behaviour of the address though the neighbours and if needed might move the second neighbours as well which is the addresses to which the direct neighbours have interacted with.

1. BitcoinWhoIsWho : From this source when we download the page we get HTML of the sample page attached in fig [] . From here the main data we are after is the scam alerts section which in this case a bunch of reports sent through by random people online who have either witnessed or been targeted by a particular address.

2. HashXp :

After we have extracted the data from the 2 sources we are saving it in a seperate file to analyze and understand what sort of information we can get and what conclusions we can make about the target address from the data about the address and its neighbours.

5.5 Step 4 : Analyzing data obtained to get some results

6 Critical Judgement and Evaluation of Results

7 Conclusion

8 Potential Future Work