STUDY OF BITCOIN ADDRESS GRAPH: LINKING ADDRESSES AND ESTIMATING OVERALL BALANCE OF A USER

By

NIKHIL JAIN

Bachelor of Computer Science

University of Petroleum and Energy Studies

Dehradun, Uttarakhand

2015

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2017

STUDY OF BITCOIN ADDRESS GRAPH: LINKING ADDRESSES AND ESTIMATING OVERALL BALANCE OF A USER

Thesis Approved:

Thesis Advisor :

Dr. Eric Chan Tin

Dr. Christopher Crick

Dr. Yanmin Gong

ACKNOWLEDGEMENTS

The Master's Degree from Computer Science department, Oklahoma State University has given me immense experience and knowledge in my field of interest. I would take interest to thank my Thesis Advisor Dr. Eric Chan Tin for his continuous assistance, for believing in me and encouragement to learn new technologies.

I would express my gratitude to committee members, Dr. C. Crick and Dr. Yanmin Gong for their guidance and support. I would extend my thanks to bitcoin community for their continuous support and help throughout the duration of my thesis.

Finally, I would express my profound gratitude to my parents Mr. Vineet Jain, Mrs. Vandana Jain and my family members for supporting me throughout my life. I thank my my friends Meghana Kiran, Prateek Dwivedi, Ashwin Kanan, who helped me out with their constant moral support when I needed the most. Without their support I would not be able to complete my thesis.

Lastly I would like to thank all of the Computer Science Department. You all gave the opportunity to strengthen my technical and communication skills.

Name: NIKHIL JAIN

Date of Degree: JULY, 2017

Title of Study: STUDY OF BITCOIN ADDRESS GRAPH: LINKING ADDRESSES AND FINDING OVERALL BALANCE OF A USER

Major Field: COMPUTER SCIENCE

Abstract: Bitcoin has become a popular electronic currency in recent years. It is a virtual currency with no central authority and relies on a peer to peer network. The bitcoin transactions that are carried out over the network are recorded and stored in a public ledger which is accessible to every peer in the network. The sender and receiver for each transaction are identified only by cryptographic public Bitcoin ID or addresses that are generated through bitcoin wallets. While Bitcoin's presumed anonymity offers new avenues for commerce, several recent studies have suggested otherwise. In this paper we explore the anonymity of Bitcoin system. We start by collecting user data from a popular bitcoin forum, where we scrape the address along with the username from the posts made by 1,460 users. Next we extract the transaction details from the blockchain, and construct an address graph that shows the flow of bitcoins in the network. We annotate the graph by linking all public Bitcoin ID to a user. Linking addresses gives an estimate of the overall bitcoin balance of a user. We also look up the transaction details of each user, in order to find a link between them and some well-known addresses. We were able to find few users from our study who made donations to WikiLeaks and who were frequent visitors to gambling sites. We also found a user who performed a transaction with Silkroad owner Dread Robert Pirates, which suggests that the user might have purchased some illicit item from the website.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Bitcoin is a peer-to-peer electronic system that was first described in a paper by its creator Satoshi Nakamoto in 2008 [1]. It is a decentralized cryptographic currency that enables instant payments to anyone, anywhere in the world. Bitcoin uses hashing technique and digital signatures to ensure that the bitcoin belongs to a particular person or has been created or spent by him. Bitcoin maintains a chain of blocks to record all the transactions that are being carried out. The history of transactions is shared with the whole network in a peer-to-peer fashion and is agreed using proof-of-work system [2, 3].

1.1 BITCOIN SYSTEM

Bitcoin is an electronic currency without any issuer or central authority. They are mined continuously by the miners at such a rate that the eventual total number would be 21 million [4]. There is no third party controlling the production or supply of bitcoin nor there is a requirement for one while making a transaction.

Miners mine a block which contains thousands of transactions that take place and links all these blocks together using the hashing technique by adding the hash of the previous blocks that was accepted universally to the new block, in order to prevent the attackers from double spending [5] the bitcoin.

All the transactions within the block are maintained in a data structure called *Ledger*. In order to validate a transaction the party that is receiving a Bitcoin should wait for 6 confirmations i.e. wait for the transaction to appear in 6 blocks.

Every 10 minutes a block is created which contains a record of transactions that are linked together in a form of a Merkle tree [6]. Since bitcoin is a decentralized mechanism the blocks are created by various nodes across the globe, which are continuously listening for any transaction that is being carried out and build a block containing all of those transactions. Every node needs to solve the hash puzzle wherein they need to calculate a nonce value [7], which when hashed with previous block hash generates a new hash that should fit within the specified target space.

It's a hit and trial mechanism wherein whichever machine finds the nonce value first gets to publish the block and is transmitted to all the nodes and every node validates the transaction listed on the block and once it is verified the block is added to the long chain of blocks. To carry out this mechanism in an honest fashion the Bitcoin developers introduced a reward feature where the node that publishes the block gets 50 Bitcoin as a reward. This value halves every four years currently it is 12. This is also known as Bitcoin mining. Apart from this there is also a transaction fees collected by the peers as a reward.

1.2 BITCOIN WALLETS

Just like normal wallets are used for storing money, bitcoin wallets are used for storing bitcoins, which can be used for spending and receiving BTC (bitcoins). Bitcoin wallet, as described by John Villasenor in his blog "something that stores the digital credentials for your bitcoin holdings" [8].

The mechanism deployed by bitcoin wallets is public key cryptography, where in two keys are generated public-key and a private-key. There are several types of wallet online, offline, paper wallet etc.

Wallets generate an address on the user's behalf, which it can use to carry out transactions without revealing his/her public identity. It can generate several addresses, allowing the user to use a new address every time to make a transaction. The figure below depicts an online bitcoin wallet.

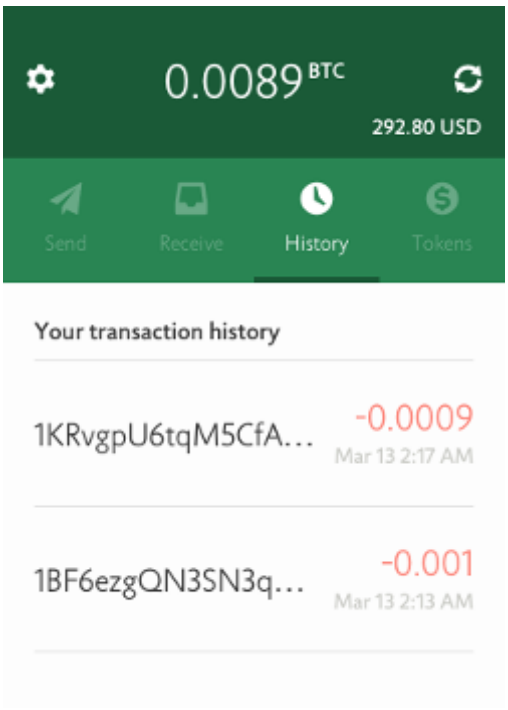


Figure 1.1 Bitcoin Wallet

1.3 ANONYMITY

Anonymity is a major part of the Bitcoin network. It uses peer to peer network where the identities of the nodes are unknown. But bitcoin doesn't offer complete anonymity, the

anonymity is compromised in order to gain the decentralization property and it is known as pseudo anonymity. The linking address mechanism is well known technique for deanonymizing a client who is using wallet software that doesn't offer any mechanism to prevent the identity of the client. Another mechanism is the Idioms of use where in you know the addresses of the major vendors and study the transactions carried out with that vendor by the users with the help of address graph.

Maintaining anonymity in today's era of networked data and online services is becoming a daunting task for the developers. The study on privacy attacks by Narayanan and Shmatikov [9] and Backstrom et al. [10] revealed the difficulty of maintaining user's identity in the presence of network data and user behavior. They de-anonymized the Netflix Prize dataset using information available from IMDB website.

Such linking could also be achieved in Bitcoins using the publically available information stored inside the blocks in a ledger. The addresses involved in a transaction could be tied to an identity, thereby compromising the anonymity of that person, as you can now keep a track on his/her activity. One such study was conducted by Dan Kaminsky, a security researcher who studied various aspects of anonymity in the bitcoin system. His work dealt with the 'linking problem' and was published in a security conference in 2011 [11].

Various other addresses that belong to a user can also be identified from the publically available information and tied to them, only if you can find a link to one of the addresses from the user. These links could be easily gathered from various social networking sites like youtube, twitter, facebook or from various blogging sites or forums like reddit, bitcointalk etc. The addresses for various famous vendors are made publically available on their website.

Linking multiple bitcoin addresses to a user, gives you an estimate of the overall balance in the person's account, which clearly violates the idea of anonymity. The motivation for this analysis is to demonstrate the inherent limits of anonymity when using Bitcoin. This will ensure that people have no false expectations from the system.

There are two main goals for this work:

1. Tie real "names" to transactions:

Real names could be a person's true name or username from an online public source or any other data source. The goal is to associate numerous unrelated cryptographic ID's with an actual user. There have been instances when users have accidentally leaked their real ID's to the world. Silk Road owner Dread Robert Pirates unintentionally revealed his public bitcoin key in online forum bitcointalk.org

2) Tie Bitcoin addresses to the real "names" involved in a transaction:

There are various Bitcoin addresses that are involved in a transaction; in Bitcoin every user has many addresses that it uses in order to maintain anonymity. While analyzing the blockchain and transactions these addresses could be tied to the user.

CHAPTER II

RELATED WORK

2.1 MECHANISM

Crypto currencies grew in the market strongly because of their mechanism for substantiating ownership, defense against double spend, ensuring anonymity/privacy of users [12] and their ability to generate themselves without involving any third party authorization. It is combined mixture of cryptography and peer-to-peer protocol. Double spend was a significant threat to the bitcoin model before the proof-of-work [2] system came into existence. In order to reduce the time gap between the transactions made and receiving confirmation about the same, developers decided to introduce the fast pay transaction [13] methodology. Though the waiting time would have reduced drastically, the chances of double spend stayed prominent. As the authors of the paper demonstrated through their experiments, that the double spend was possible provided if the two conditions were satisfied.

Whenever a new node joins the network, it receives the data from its adjacent nodes that are previously in the network and they send the complete blockchain to the new node. There is a possibility of tampering with the delivery of blocks [14] which is discussed in the paper where it talks about attacker altering the information received by bitcoin nodes, and modifies their views of the ledger state, thereby allowing the adversary to considerably increase its mining advantage in the network, and to double spend transactions in spite of the current countermeasures adopted by bitcoin.

2.2 SECURITY

Several security threats loom over the bitcoin mining system, some of which are Selfish Mining [15] where in the attackers force the honest miners to abandon the blocks that they were working on by presenting a fake private chain to them, where they have manually added the fake blocks. Another eminent security threat is gaining 50% of the computing power in mining process. The paper [16] discusses about the probability of overcoming the 6-deep confirmed transaction is 50% even with 40% computing power. These threats could be countered by placing checkpoints, so that the blocks that precede the checkpoint can no longer be altered.

2.3 PRIVACY

Several research studies [17, 18, and 19] have hinted at potential privacy limitations with bitcoin transactions. The authors [18] did a Transaction Graph Analysis on bitcoin blockchain in 2012-13, wherein they isolated all the large transactions in the system, and discovered that almost all of them are closely related to single large transaction that took place in November 2010. Another such study was carried out in 2011 by Fergal, Reid et al. [17] wherein they constructed a network graph and tried to link as many known usernames on that graph, thereby revealing several static and dynamic properties of the network.

Dread Pirate Roberts (DPR) unknowingly leaked his bitcoin address on popular forum bitcointalk, which was later analyzed in the research carried out by Dorit Ron et al. [19] where they constructed a flow chart in order to understand the flow of bitcoins between the DPR's address and other addresses and trace the missing bitcoins from the network.

Bitcoin's design keeps all transactions in a public ledger. The sender and receiver for each transaction are identified only by cryptographic public key ids. This leads to a common misconception that it inherently provides anonymous use. The authors [20] explore the level of

anonymity in the Bitcoin system, their approach is to annotate the public transaction graph by linking bitcoin public keys to real people - either definitively or statistically.

In this paper we attempt at tying other bitcoin addresses to a user that belongs to him/her in comparison to the above. We construct the address graph to analyze the input and output of a transaction and the addresses that belong to the same user are tied together. Tying various addresses gives us an estimate of the overall balance in a user's wallet.

CHAPTER III

METHODOLOGY

3.1 TOOLS USED

This section includes the tools used for data collection:

3.1.1 BITCOIN CORE

Bitcoin Core is the main client for bitcoin. It was first developed by Satoshi Nakamoto and was released under the name Bitcoin, which was later changed to Bitcoin-Qt and ultimately to Bitcoin Core. The software acts as full client for nodes that are trying to connect to the Bitcoin network. The changes made to bitcoin core software reflect the changes made to the underlying bitcoin protocol. It offers various features such as transaction verification engine for full nodes, wallet (which is built in by default) for carrying out bitcoin transactions.

It doesn't allow trading of bitcoins for any other currency but generates QR code on user's behalf for receiving bitcoins. The software downloads the entire blockchain in order to validate the transactions. The wallet is bundled with a command line based daemon with JSON-RPC interface named bitcoind, which can be used for setting up a testing environment like Bitcoin. Bitcoin-cli is another program that is included alongside bitcoind and is used for sending RPC commands to bitcoind. Through those RPC commands you can extract information stored in the blockchain like hash of a particular transaction, raw information about a transaction, details of all the transactions stored inside a block etc.

3.2 IMPLEMENTATION

In this section we shall describe the several steps involved in revealing bitcoin user activity information by leveraging publically transaction information.

3.2.1 PRE-PROCESSING

Generally raw transactions are stored inside the blocks that are mined roughly every 10 minutes. They have to be extracted from the blockchain. As of June 14, 2017, approximately 472,000 blocks have been mined in the bitcoin blockchain. Each block contains hundreds of transactions

.

3.2.2 BLOCKCHAIN PARSING

The standard bitcoin client bitcoin-0.14.2[1] automatically downloads the whole blockchain in a P2P (Peer to Peer) fashion. The blockchain is stored as a distributed ledger in a ".dat" file format. The data from the blockchain was extracted through the RPC (Remote Procedure Call) channel provided alongside bitcoin client software with the help of RPC commands. The result from the RPC command is in JSON format which contains the transaction details stored inside a block in a raw format. A python script was written to extract the useful details from the raw transaction and store them in a file, which were required to construct the address graph. Figure 3.1 shows the screenshot of the information extracted from the blockchain.

TXID	Input Address	Value	Output Address	Time & Dates
363ec29e4fc43f97576423d0d522f5f0fc79c5c018c3a210c5644ab79a38041d		25.0301	1CjPR7Z5ZSyWk6WtXvSFgkptmpoi4UM9BC	2014-01-12 01:36:19
f264c5c36b624110201a27bd02883508d29dfe1f94975aa8cc652fb8bc1496a9	1AyHWFST4jhcdVs74Hd3pvCdafpktmiF8Y 154yacyVMuVoahSd1dmsz6PDS8MZzX3FzB 1KzXSypcJUdStaHfLtucBiaqEWDZ5PhBix 14FJfPxbXpBrTPnPcCS1gwWpTz7XVmSJeQ 15CdDBvSfr5JsnbqoSgnEdSFvtKwf1GoNj	299.0	13vZBq8ayCzficfP3uwm52bctv5mSXKfM1	2014-01-12 01:36:19
6e4bcc3380404f711362c493d700a472ee01468cfd1444d09e47ad34654accc9	1PCP6C9KuGgHzT9tF1UXYgHXXAmzFxkimA	6.0	1BHZT5KRdezCBCs5FrbmVVEv7JkhXYzXfn	2014-01-12 01:36:19
dc53815bcb8e66951b8ae965cbc2e7f90158d9295607526a3df615366c20265a	1Mw2HfCxJbhrzui2RNW2rASrcYECJHcJ2c	1.5 1.19538703	1CcRXBTE4LtnqTVCmUNwvS2CS1bwrNz9Ke 1Mw2HfCxJbhrzui2RNW2rASrcYECJHcJ2c	2014-01-12 01:36:19 2014-01-12 01:36:19
3bed03bb59c5f56ce25ec7dda9c84ed89bce89e28dae6096266dc3a8136fa54f	1JRuEYxqAQCUYE4VCPqeqsGFEDnRrCYnWU	1.0 4.998	19trGPXF4iqDPUDqdus7Rckd7eb9br8uZ4 1PX1mmqx4ZSrjEnKdxXF8BuHVXqXvu5GvN	2014-01-12 01:36:19 2014-01-12 01:36:19
c704c4b9b8b757e79496071854aaa51acab755480856c4d91f7f7c51e4a0aab8	1Ao3wfPrc5rGKE7cHUq7JRyGGcvqq3Sf11 1EnZBwTT4ymPnv3zhbSGEUQn7vZ5AkHTA8 1HEBCqBSJSYfekEyxMR7bwd9anFtHJgLu6	0.01131755 0.525	19d81913h85oHuQft1izygzMd4av15CFuD 13x3A6HhEcNVuMr3qgf2GixNtNrrPe8f71	2014-01-12 01:36:19 2014-01-12 01:36:19

Figure 3.1 Transaction Details

The first column is the TXID which is an identifier used to uniquely identify a particular transaction; specifically, the sha256 hash of the transaction. The second column is the input address/addresses that are participating in the transaction. The third column is the amount of BTC (bitcoins) transferred and the fourth and fifth column represent the address/addresses that is/are receiving the amount and the date and time when they first appeared in the blockchain.

3.3 WEB SCRAPING

In Bitcointalk[21] forum posts, users usually attach their bitcoin address as a signature for the contribution they make to the community, for example with a tutorial on how to setup a mining software. Figure 3.2 shows an example of such a post. They expect to receive tips from forum readers that find their post helpful. A Python package called Scrapy [22] was used to fetch and parse “Bitcointalk” forum pages. A spider was written that crawls the forum in a breadth-first manner looking for post signatures that might contain bitcoin addresses.

Figure 3.2 A typical user signature line that includes bitcoin address

The scraping code ran for ran for a total of 7 days starting 10/01/2016 till 10/07/2016 and scraped 1,460 unique users with their addresses. A total of 112,086 pages were crawled, after which the experiment was terminated manually, due to hardware issues. Scrapy takes a lot of space on the disk while keeping a track of previously visited links.

```
{"mb300sd": "1D7FJWRzeKa4SLmTznd3JpeNU13L1ErEco"}  "January 09, 2014, 10:24:23 PM"
"2016-10-03T14:38:45.047080"
{"awgilyas": "1AwangATDktxLw8e6VaDijw9AMWz5wy8bx"}  "December 05, 2013, 08:15:30 PM"
"2016-10-03T14:38:56.876711"
{"aaa801": "1325TrScK8jkiPuMEMxNf1VXHHfnR1QtgN"}  "March 23, 2012, 11:22:01 PM"
"2016-10-03T14:39:38.347585"
{"lodcrappo": "1PoRYaGS56ksQmK7XXLurW3B2zwCAE8PRc"}  "March 29, 2012, 03:50:35 PM"
"2016-10-03T14:39:43.884403"
{"N[e]wBie": "1ESZr887vTZqYtDuwwspn1jBaoRU9jMcv1"}  "March 23, 2014, 09:54:25 PM"
"2016-10-03T14:39:59.379305"
{"Criminal": "16VMabM7wQmsuBwGPJQguJGL3onbfX6eBF"}  "March 10, 2014, 04:57:30 AM"
"2016-10-03T14:40:26.174233"
{"jhansen858": "1DDpiEt36VTJsiJunyBc3XtG6CcSAnsQ4p"}  "March 21, 2014, 06:29:03 AM"
"2016-10-03T14:45:06.773516"
{"eduncan911": "16aoCMuo9XAGsPuKK9NL4xWr8B9fsSvMx"}  "February 07, 2014, 12:19:26 PM"
"2016-10-03T14:45:53.869509"
{"crudpuppy": "1FSAWqTe6geJwM8qyG7iG1TcJQykmWSEy8"}  "October 02, 2013, 04:20:54 PM"
"2016-10-03T14:49:02.924429"
{"shivansps": "1NEvmZTwLhqBgr3h7LwZap1sEHsyf3QFuv"}  "June 22, 2013, 09:58:07 PM"
"2016-10-03T14:50:09.572956"
{"Leon D": "1AR2FxnFK9cxZ4YbU6hn1hWW1N3JiCEjYk"}  "July 09, 2013, 06:37:42 PM"
"2016-10-03T14:50:15.355055"
```

Figure 3.3 Scraped user addresses and id's

Figure3.3 contains the name of the user on Bitcoin platform followed by his bitcoin ID, the date and time when he made the post, and underneath that is the time of the system when this username and bitcoin ID was extracted from the bitcointalk platform.

3.4 DATA PROCESSING

The address graph is constructed using the transaction records from the blockchain and the user addresses from the web scraping. The address graph contains the nodes that are denoted by public addresses of anonymous individuals and edges that represent a particular transaction between the source address and the target address. From the graph all the transactions in which each individual user's id participates can be grouped together to give an idea of the other id's owned by the individual and the balances in each one of them.

3.5 EVALUATION MEASURE

The heuristic is very simple and reliable. People who reveal their Bitcoin ID in public, for business purposes or donation purposes, their addresses can be looked up in the blockchain to see what all transaction they had been a part off. All other addresses that are participating in a transaction alongside that address as an input are grouped together as they belong to the same user because you can only control the addresses that belong to you. For example you can make payment from multiple credit cards, provided you own each one of them.

The other addresses that belong to the user are gathered and the balance in each individual address is calculated to give an estimate of the overall balance of an individual. For evaluation phase we contacted 20 users (out of 1460 that we scraped) randomly to verify the correctness of

our results. The remaining users couldn't be contacted due to time constraint, and it wasn't manually possible to query each individual.

For the verification process, 20 users were picked randomly. They were contacted over the bitcointalk platform manually, to authenticate that the addresses that were gathered, actually belonged to them. Out of 20 users 17 confirmed that the results were correct and that the addresses mentioned did belong to them. The remaining 3 users never replied back to our messages.

CHAPTER IV

EVALUATION

4.1 ADDRESS GRAPH

Once the transaction records are extracted from the blockchain, we construct an address graph that gives an intuition towards the flow of bitcoins between public key addresses over time. The address graph is a directed graph where the nodes represent the characteristics associated with a bitcoin transaction like input address, transaction id, block number (which block it belongs to) and output address while the directed edges represent the relationship between the nodes or entities like which transaction id "uses" this particular address, to which entity an address "belongs" to, which block "contains" this transaction id and the "input" and "output" addresses associated with a transaction.

Typical transactions in today's blockchain are multi-input/multi-output transactions. Both the source and target entities are trying to use new addresses for every subsequent transaction. Fig 4.1 depicts the address graph for the bitcointalk user "omegadraconis".

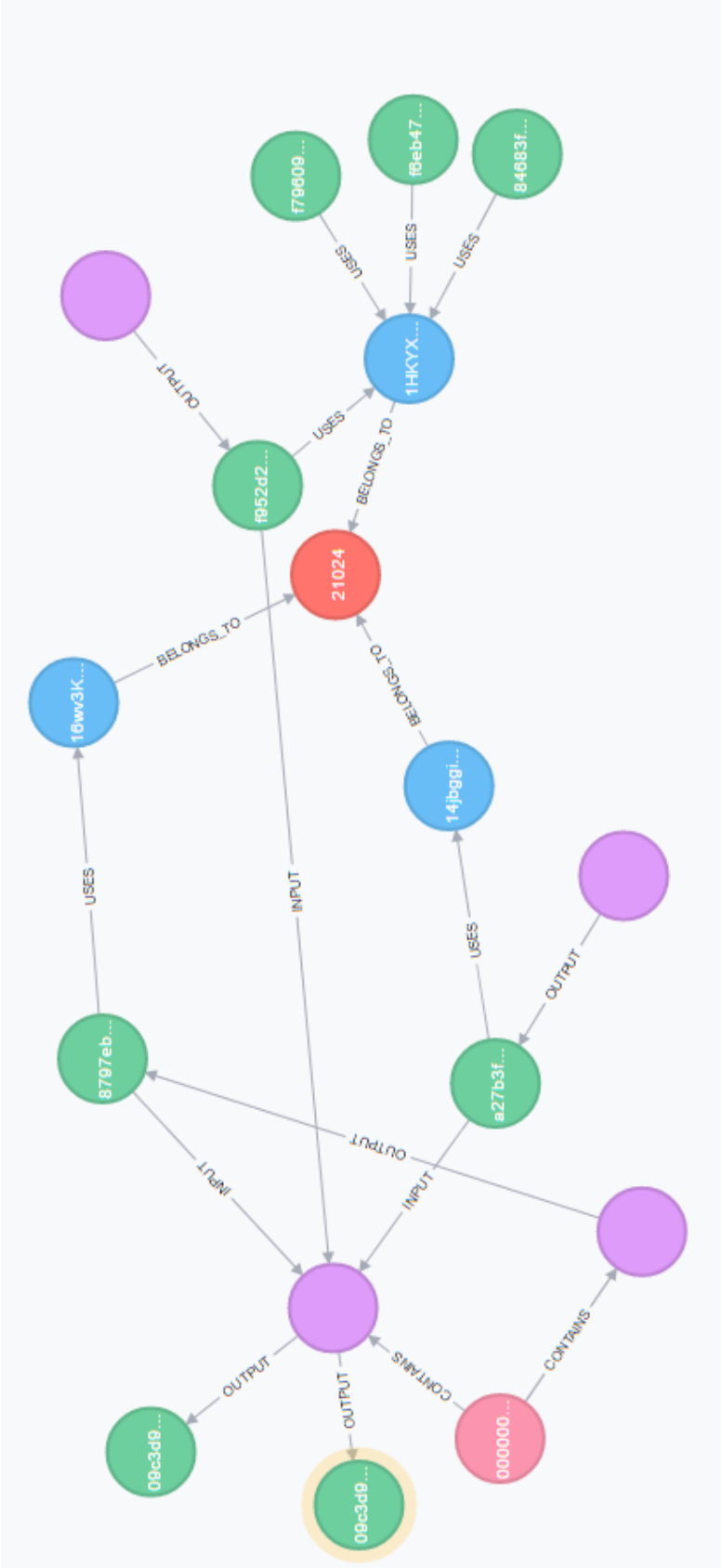

f79609....
f8eb47....
84683f....
USES
USES
USES
1HKYX....
OUTPUT
USES
f952d2....
BELONGS_TO
21024
BELONGS_TO
BELONGS_TO
16wv3K....
14jbggi....
USES
INPUT
USES
OUTPUT
8797eb....
a27b3f....
OUTPUT
INPUT
INPUT
OUTPUT
OUTPUT
CONTAINS
CONTAINS
09c3d9....
09c3d9....
000000....

Figure 4.1 Address Graph for "1HKYXgu9uLp8AQXabYrqbmAGqS73huNM7K"

The top right corner of the image shows the transaction id(shown in green circle) and tells us that the address associated with this transaction id is the recipient address (to whom the bitcoins are transferred) in the transaction (1HKYXgu9uLp8AQXabYrqbmAGqS73huNM7K), indicated with the help of "uses" relationship. The green color entity signifies that the address involved with it is either the sender or receiver in the transaction.

Subsequently the output of one transaction could be an input for some other transaction. As could be seen in the picture towards the top left corner, where the output of a transaction (shown in purple color) f952d2da7d8f0cf6b4761ad1fa64a78072554dcad15740f22d331d8194d0d2c7 (transaction id) is the input for some other transaction, shown in the left hand side of the image. The purple color entity represents the transaction and keeps a track of participating input and output transaction ids.

The bottom right corner of the image shows the block hash (represented with pink circle) which "contains" a link to every transaction id that has been recorded under that block at the mining stage. Block hash contains various properties like the height of the block or in other words the order in which it was added to the blockchain for example the first block that was entered into the blockchain has a height "1" in the blockchain, id of the block and also the timestamp recorded at the time of publishing.

The entity (represented with red circle) represents all the addresses that belong to a single person, as shown in the Figure 4.1. The addresses that are bound by the same entity number belong to the same person/group, as could be seen in the figure above, that the addresses (represented by blue

circle) are bounded by the same entity number“21024”. It reflects that all of these addresses belong to the same user “omegadraconis”.

Table 1 shows the list of other addresses that belong to the same user and the bitcoin balance in each one of them. Most of the addresses shown in Table 1 reflect a zero against them under the balance column, that’s mainly because every time an address participates in a transaction all the bitcoins that it contains are consumed irrespective of whether the amount of bitcoins to be transferred to another address is more or less than the bitcoins contained inside it. Once the desired amount is send to the other address, a new change address is generated to send the remaining balance.

Suppose if Alice wants to send 2.3 bitcoins to Bob and Alice has control over two addresses ‘a’ and ‘b’ that has 2 bitcoins and 1 bitcoin respectively. In order to transfer the desired bitcoins to Bob, both the addresses of Alice would participate in the transaction and a total of 3 bitcoins would be send. Since Alice had the set the amount to 2.3 bitcoins, the remaining 0.7 bitcoins should be returned as the change amount. The bitcoin script generates automatically a new change address and sends the remaining balance to the new address of the user.

Apart from the above mentioned reason, a user also has an option to transfer all the change contained in different addresses to one single address of his choice, so that he can have the overall amount in his wallet inside one address.

BITCOIN ADDRESS AND BALANCE	
Number Of Addresses : 9	
ADDRESS	BALANCE
1HKYXgu9uLp8AQXabYrqbmAGqS73huNM7K	0.00001000

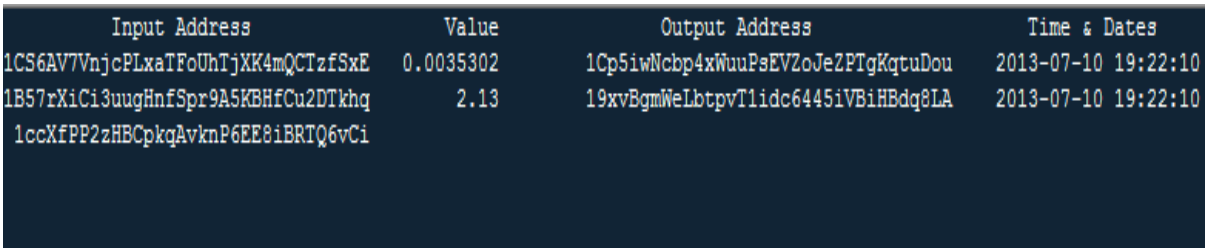
1HDeGjSDN4BX6Xwch2PbBVjcV8BAGwGdsx	0.00000000
1GTzrz2Bwyj9W8oLXAdKXuquqX1Dhq34Cz	0.00000000
14jbggiCtvbENi91jZQQKgueSgmJraS1hH	0.00000000
1BXYbKdtw35wJJfX6j9QWs3z1MC6ZAJPwX	0.00000000
16wv3KGFnK1GGi3ZUVFdxKUTftv1DUQFs3	0.00000000
1EMuF3u3nUo1TwrKCs6FRZozFjwUzG88Lp	0.00000000
1xZEajuHNnLTvEvBYNK5ATLoUR7NN5ctz	0.00000000
1LUcEQaEa8ePQuhxWNWkQsiYsgKGVKmrU8	0.00000000

Table 1 List of bitcoin addresses and balances in them for user omegadraconis

The above table reflects on the balances on each individual address of the user, which when combined together can give the overall balance of a user at any particular time. For our experiment we recorded the data from the beginning of blockchain till 14th June, 2017 approximately 472,000 blocks and constructed the address graph for the same. The balances shown above in the table for the respective address are accurate as of end date specified for the data gathering.

Whenever a person makes an account on any wallet, by default the option for receiving the change on a different bitcoin address is enabled by the wallet to increase the privacy of the user, in order to make it difficult for the others to know the balance of that person. So whenever that person makes a transaction and if there are some leftover bitcoins then the difference is sent to a new address. The person has an option of disabling this feature and selecting no privacy mode in order to receive change on the same address. He also has the option to accumulate all the change from the addresses back to the old address or a new address depending on his/her will.

He can also make use of these addresses that contain the change from previous transactions, to make a new transaction alongside his main address. This heuristic has been used in this research to find other addresses of a user that appeared alongside his/her already known address, collected from the forum “bitcointalk”.

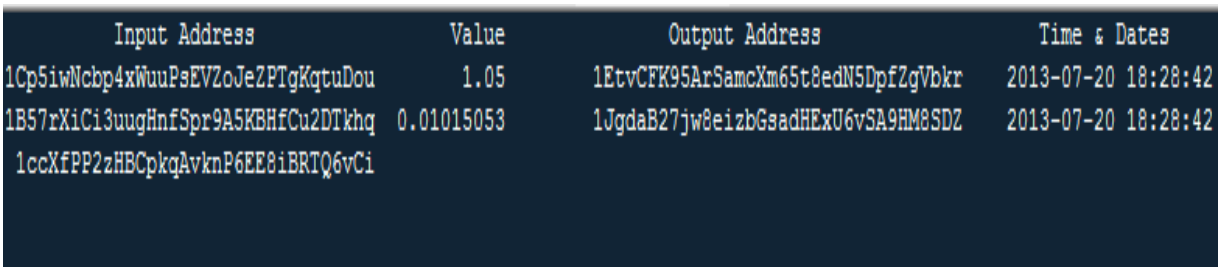


Input Address	Value	Output Address	Time & Dates
1CS6AV7VnjcPLxaTFoUhTjXK4mQCTzfSxE	0.0035302	1Cp5iwNcbp4xWuuPsEVZoJeZPTgKqtuDou	2013-07-10 19:22:10
1B57rXiCi3uugHnfSpr9A5KBHfCu2DTkhq	2.13	19xvBgmWeLbtpvT1idc6445iVBiHBdq8LA	2013-07-10 19:22:10
1ccXfPP2zHBCpkqAvknP6EE8iBRTQ6vCi			

Figure 4.2 Transaction of bitcointalk user “Trongersoll”

The above picture shows the transaction made by bitcointalk user “Trongersoll”, who owns the address “1B57rXiCi3uugHnfSpr9A5KBHfCu2DTkhq”. As could be seen in the picture there are other addresses that participate in the transaction as well, all of which are also controlled by the same user, as you could only control the addresses that are owned by you.

The output address in the above transaction “1Cp5iwNcbp4xWuuPsEVZoJeZPTgKqtuDou” also belongs to the same user, shown in the Figure 4.3 below.



Input Address	Value	Output Address	Time & Dates
1Cp5iwNcbp4xWuuPsEVZoJeZPTgKqtuDou	1.05	1EtvCFK95ArSamcXm65t8edN5DpfZgVbkr	2013-07-20 18:28:42
1B57rXiCi3uugHnfSpr9A5KBHfCu2DTkhq	0.01015053	1JgdaB27jw8eizbGsadHExU6vSA9HM8SDZ	2013-07-20 18:28:42
1ccXfPP2zHBCpkqAvknP6EE8iBRTQ6vCi			

Figure 4.3 Transaction involving change address of user "Trongersoll"

The change address appears alongside other addresses that are owned by the same user, which implies that the change address also belongs to the same user, which forms the basis for our heuristics which we stated earlier. Table 2 below shows the remaining addresses owned by the user and the balance in each one of them.

BITCOIN ADDRESS AND BALANCE	
Number of Addresses : 7	
ADDRESS	BALANCE
1B57rXiCi3uugHnfSpr9A5KBHfCu2DTkhq	8.17464323
1ccXfPP2zHBCpkqAvknP6EE8iBRTQ6vCi	3.51182210
1CS6AV7VnjcPLxaTFoUhTjXK4mQCTzfSxE	0.00189577
1Cp5iwNcbp4xWuuPsEVZoJeZPTgKqtuDou	0.00000000
1HEMT43fkj3RztLsfbfac4EKxvfNPjDakk	0.00000000
1LQRDoGsjLgm7mNa3z3ikwhfruZv3fe6cD	0.00000000
18caWMT2nuvbTPdMQcrpcuuR8mKpgEZKQ7	0.00000000

Table 2 List of bitcoin addresses and balances in them for user Trongersoll

For this study 1460 users with their respective bitcoin addresses were scraped from bitcointalk forum and the address graph was generated connecting their other addresses to the already known address. The balance on each address was gathered from blockchain.info website by querying the API of the website. The graph below shows the number of addresses collected per user.

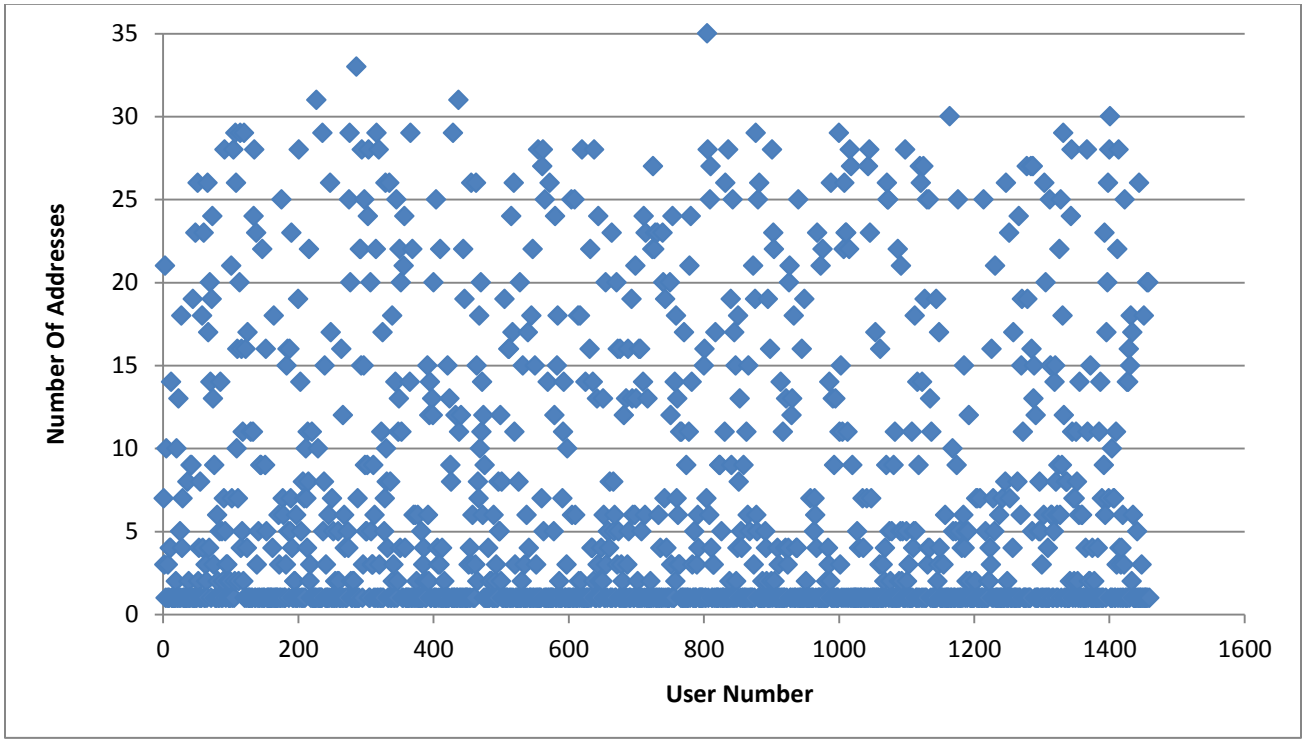


Fig 4.4 Number of Addresses per user

For the verification process, 20 users were picked randomly. They were contacted over the bitcointalk platform manually, to authenticate that the addresses that were gathered, actually belonged to them. Out of 20 users 17 confirmed that the results were correct and that the

addresses mentioned did belong to them. The remaining 3 users never replied back to our messages.

4.2 FINDINGS

In this section we analyzed the transactions made by user over time, tried to gain an insight as to how they are spending their bitcoins over the internet. Whom they are transacting with, which services they are accessing in exchange for their bitcoins, which sites they are utilizing their bitcoins on etc.

We gathered few known bitcoin addresses that are available publicly, and tried to find a link between them and the users that we have gathered from bitcointalk. The table below shows the list of some famous addresses and their contemporary owners.

ADDRESS	OWNER
1dice4J1mFEvVuFqD14HzdViHFGi9h4Pp	SatoshiDICE
1F1tAaz5x1HUXrCNLbtMDqcw6o5GNn4xqX	Silkroad Seized Coins
1HB5XMLmzFVj8ALj6mfBsbifRoD4miY36v	WikiLeaks

Table 3 List of famous bitcoin addresses and their owners

4.2.1 SATOSHI DICE

Satoshi Dice (sometimes stylized as SatoshiDICE) is a "blockchain-based betting game" operating since 2012. We collected all the data about the transactions that were ever made by the address “1dice4J1mFEvVuFqD14HzdViHFGi9h4Pp”, and compared with the addresses of the users that we had collected.

We found few users who transacted with SatoshiDICE frequently, the below table shows the list of those users alongside their transactions frequencies to SatoshiDICE.

USER	FREQUENCY
Tunafish	1
marto74	4
Aruca	47

Table 4 Transaction Frequencies of users with SatoshiDICE

The total balance for this SatoshiDICE address as of 6th July 2017 is 0.00079398 BTC. It has completed over 98,900 transactions.

Input Address	Value	Output Address	Time & Dates
12DNdacCtUZ99qcP74FwchaCPzeDL9Voff	0.01	1dice4J1mFEvVuFqD14HzdViHFGi9h4Pp	2013-03-23 16:56:53
	0.00082141	12DNdacCtUZ99qcP74FwchaCPzeDL9Voff	2013-03-23 16:56:53

Figure 4.5 Transaction between SatoshiDICE and user marto74

4.2.2 WIKILEAKS

WikiLeaks is a non-profit international body that leaks private information, news and classified media in the public which are tipped off by anonymous sources. The website started accepting bitcoins as donation to their project starting 2011. We analyzed all the donations that have been made to the website from the beginning and found that few users from our list had made secret donation to the project.

Below is the list of the users and their transaction id's of the transactions made to wikileaks.

USER	TRANSACTION ID
forrestv	bee135af448eadb304d69ce7fef8ca0f252c6e77b09a8760136a96294ce1f33c
Reikoku	f9a1f4e4758bf78f39ae3aae097b103362ad578fca871017e915c83dafd13a17
kicker049	ba4ed1dd0acddeaf5c94f1820143083e22eb2864fae341e9e39a8bf365933f3b
Lifeboat	f2494c6f1c61042870fb965e1191428b3c6bedf61fbdad30cadcf226bd2f36d9

Table 5 Transaction hashes made between user and Wikileaks

The total balance in the Wikileaks account as of 6th July 2017 is 6.62062537 BTC. It has completed over 26,000 transactions.

Input Address	Value	Output Address	Time & Dates
1J1zegkNSbwX4smvTdoHSanUfwvXFeuV23	3.0	1HB5XMLmzFVj8ALj6mfBsbifRoD4miY36v	2012-07-12 16:31:00

Figure 4.6 Transaction made to Wikileaks by forrestv

4.2.3 SILKROAD SEIZED COINS

Silk Road was an online black market and the first modern dark net market, best known as a platform for selling illegal drugs. The website was shut down by FBI in later half of 2013. We tracked the transactions made to Silkroad by customers to purchase those illegal drugs, and found one of the users from our list to be one of those customers.

Below is the transaction detail that was made to Silkroad by user "zachamo".

USER	TRANSACTION ID
Zachamo	361c89602b3562d208332d4cdbe9020e80018e671968435323498f8fa4c2d588

Table 6 Transaction ID made between user and Silkroad

CHAPTER V

CONCLUSION

We constructed an address graph which gave us a picture of the flow of bitcoins in the network, through which we tracked the activities of bitcointalk users with the help of their Bitcoin ID that we scraped from their respective posts on the bitcoin forum. We not only discovered other addresses of the user but also estimated the overall balance in each user's wallet. This shows that bitcoin transaction network is not entirely anonymous, and that using an appropriate network representation, it is possible to associate many Bitcoin IDs with each other, with the help of some external identifying information.

Furthermore we were able to find a direct link between bitcointalk user and Silk Road owner Dread Pirate Roberts (DPR). We were also able to successfully find transactions hashes between forum users and gambling site SatoshiDICE and whistleblower WikiLeaks implying that they may have dealt with, supported or interacted with such entities. With appropriate tools, the activity of known users can be observed in detail.

REFERENCES

[1] S. Nakamoto. “Bitcoin: A Peer-to-Peer Electronic Cash System”, 2008.

[2] C. Dwork and M. Naor. “Pricing via Processing or Combatting Junk Mail”. In *Proceedings of the 12th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO’92)*, pages 139–147. Springer, 1992.

[3] A. Back. “Hashcash – A Denial of Service Counter-Measure”, http://citeseer.nj.nec.com/back02hashcash.html Accessed 25th Aug 2003s.

[4] Bitcoin Block Half, “Bitcoin Block Reward Halving Countdown”, 2017, Available: http://www.bitcoinblockhalf.com/

[5] T. Sander and A. “Ta-Shma. Auditable, anonymous electronic cash”. In *CRYPTO*, 1999

[6] Blaise Gassend, Dwaine Clarke, Marten van Dijk, Srinivas Devadas, Ed Suh. “Caches and Merkle Trees for Efficient Memory Authentication”. In *Proceedings of the 9th High Performance Computer Architecture Symposium (HPCA'03.* Anaheim, 2003.

[7] P. Rogaway. “Nonce-based symmetric encryption”. In FSE 2004, volume 3017 of LNCS. Springer.

[8] Villasenor, John. "Secure Bitcoin Storage: A Q&A With Three Bitcoin Company CEOs". Forbes. 26 April, 2014.

[9] A. Narayanan and V. Shmatikov. “De-anonymizing Social Networks”. In *Proceedings of the 30th Symposium on Security and Privacy*, pages 173–187. IEEE, 2009.

[10] L. Backstrom, C. Dwork, and J. Kleinberg. “Wherefore Art Thou r3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography”. In *Proceedings of the 16th International Conference on World Wide Web*, pages 181–190. ACM, 2007.

[11] D. Kaminsky. “Black Ops of TCP/IP Presentation”. Black Hat, Chaos Communication Camp, 2011.

[12] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Stefan Savage. "A Fistful of Bitcoins: Characterizing Payments Among Men with No Names". *IMC'13* , October , Barcelona, Spain.

[13] G.O. Karame, E. Androulaki, and S. Capkun. "Two bitcoins at the price of one? double-spending attacks on fast payments in bitcoin". In *Proceedings of Conference on Computer and Communication Security*, Zurich, 2012.

[14] Arthur Gervais Hubert Ritzdorf Ghassan O. Karame Srdjan Capkun. Tampering with the Delivery of Blocks and Transactions in Bitcoin. CCS'15, Denver, Colorado, USA.

[15] Ittay Eyal and Emin Gün Sirer, "Majority is not Enough: Bitcoin Mining is Vulnerable", In *Financial Cryptography and Data Security*, 2014.

[16] Yogesh Malhotra, "Bitcoin Protocol: Model of 'Cryptographic Proof' Based Global Crypto-Currency & Electronic Payments System", December 4, 2013.

[17] Fergal Reid and Martin Harrigan. "An Analysis of Anonymity in the Bitcoin System". arXiv, 2011.

[18] Dorit Ron and Adi Shamir. "Quantitative analysis of the full bitcoin transaction graph". Cryptology ePrint Archive, Report 2012/584, 2012. http://eprint.iacr.org/.

[19] Dorit Ron and Adi Shamir. How Did Dread Pirate Roberts Acquire and Protect His Bitcoin Wealth?. In BITCOIN '14: *Proceedings of the 1st Workshop on Bitcoin Research*, Barbados, 2014.

[20] Michael Fleder, Michael S. Kester, Sudeep Pillai. "Bitcoin Transaction Graph Analysis". arXiv, 2015.

[21] https://blockchain.info

[22] https://scrapy.org/

VITA

Nikhil Jain

Candidate for the Degree of

Master of Science

Thesis: STUDY OF BITCOIN ADDRESS GRAPH: LINKING ADDRESSES AND ESTIMATING OVERALL BALANCE OF A USER

Major Field: Computer Science

Biographical:

Education:

Completed the requirements for the Master of Science/Arts in your major at Oklahoma State University, Stillwater, Oklahoma in July, 2017.

Completed the requirements for the Bachelor of Science/Arts in your major at University of Petroleum and Energy Studies, Dehradun, Uttarakhand in 2015.

Experience:

Graduate Teaching Assistant Computer Science Department, Oklahoma State University, OK: C/C++, Spring 2017