Modeling Suspicious Transactions in the Bitcoin Blockchain

Chris Shurtleff, James Cate, Kevin Gardner, and Devanshu Agrawal

***Abstract*— We propose exploring various methods for detecting unusual and potentially illicit transactions within the bitcoin cryptocurrency. To identify suspicious transactions, we will model the blockchain as several seperate but related unidirectional, sparse graphs. Our end goal will be to provide a useful resource for easily determining how unusual a particular transaction is, to aid companies in compliance with regulatory norms.**

I. INTRODUCTION TO BITCOIN

Bitcoin is a virtual currency, unsanctioned by any government, and powered by cryptography. Bitcoin relies on a decentralized ledger, called the blockchain, maintained concurrently by a network of independently-owned nodes. Each node is a computer that runs bitcoin client software, contains the full blockchain, and propagates transactions to other nodes on the network. Potential transactions are verified for legitimacy and added to the end of the ledger in groups of valid transactions called a block. The process for compiling transactions into a block is computationally intensive, so the node responsible for adding a block of transactions, termed miner, is rewarded with a small amount of bitcoin to incentivize participation. The blockchain is public, trending, and threatens the classic conception of currency, therefore it is an interesting target for analysis.

II. CURRENT RESEARCH

Bitcoin transactions are in principle anonymous. But bitcoin addresses sometimes appear in conjunction with a user name on websites such as public forums. Such websites sometimes also contain conversations of bitcoin transactions from which useful information may be inferred. Data of this nature has been scraped from such websites and used to annotate the bitcoin block chain with real user information [Fleder et al.(2015)Fleder, Kester, and Pillai]. Although the extent of annotation is small compared to the blockchain itself, its use in combination with the rich graphical structure of the bitcoin transactions network itself has led to interesting insights. For example, the annotated bitcoin blockchain allows for the construction of a bitcoin transactions graph whose nodes represent concrete users. Implementation of the PageRank algorithm on this user graph leads to the detection of interesting users who should perhaps be flagged for future investigation [Fleder et al.(2015)Fleder, Kester, and Pillai].

Even in the absence of real user information, the bitcoin blockchain is still a treasure trove of data ripe for analysis thanks to its rich structure, completeness, and public availability. Statistical analysis of the bitcoin transactions graph and its features has led to insights into the behavior of users and transactions; it was found, for example, that most large transactions in 2012 could be traced back to one single transaction in 2010 and that the subgraph of such large transactions exhibited peculiar features– perhaps an attempt to conceal or obscure suspicious activity [Ron and Shamir(2012)]. This example motivates the idea that even though bitcoin transactions are in principle anonymous, it could still be possible to detect suspicious activity by revealing anomalous structures in the bitcoin transactions graph. For instance, it has been shown that certain local features in the bitcoin transactions graph (e.g., degree and clustering coefficient) can allow us to cluster users and transactions and thus to detect suspicious outliers that should be flagged for further investigation [Pham and Lee(2016)], [Zambre and Shah(2013)].

The core structure of bitcoin transactions is that they make reference to previous transactions and therefore allow bitcoins to be traced. We believe that local feature extraction from the bitcoin transactions graph fails to fully capture this essential structure; for our project, we plan to compare transactions based on the flow of bitcoins common to both transactions.

III. PROPOSAL

Bitcoin, being decentralized and pseudo-anonymous, lends itself to criminal activity which inevitably requires money laundering. Our project will focus on ways to automatically flag suspicious transactions with our stretch goal being to identify nodes that have access to dirty money. We will pursue several strategies to flag transactions centered around modelling the blockchain using graph theory. We will consider a graph mapping bitcoin addresses to vertices and transactions to edges and vice versa. Building on this model, we propose using Fleders [Fleder et al.(2015)Fleder, Kester, and Pillai] Union Find algorithm to cluster addresses into likely individuals. Our potential transaction-identification-strategies within the graph include:

- Determining the degree of relationship between two transactions
- Identifying miners and analyzing the structure of their transactions
- Tracing specific bitcoins back to the original miner
- Scraping identifying information from the web and associating that with known address clusters

IV. METHODS

We will obtain the bitcoin blockchain data either by using a bitcoin client, bitcoin-0.8.5, or by downloading a prepared

*This work was not supported by any organization

bitcoin data set from the Stanford Network Analysis Project. In the latter case, the data is a text file in which each line records a transaction. We will use python to parse the blockchain and will construct the transactions graph using the NetworkX python package. We will apply the union-find algorithm to group bitcoin addresses that are likely to belong to a common user and will develop algorithms based on network theory to analyze the bitcoin flow through the transactions graph among users.

Time permitting, we will use the scrapy python package to obtain real user information from public forums for annotating our transactions graph.

Finally, we are considering the use of computing clusters to aid in scaling our analysis to the entire bitcoin blockchain.

V. PROJECT TIMELINE

TABLE I
INITIAL TIMELINE OF MILESTONE DUE DATES

October 13	Prepare limited transaction sample for development
October 27	Prepare and begin exploration of graph
November 17	Miner and Pool Structure
November 17	PageRank analysis
November 20	External data scraping (if stretch is met)
November 20	Scaled blockchain (if stretch is met)
November 24	Visualization and Interpretation
December 5	Final Report

REFERENCES

[Fleder et al.(2015)Fleder, Kester, and Pillai] Michael Fleder, Michael S. Kester, and Sudeep Pillai. Bitcoin transaction graph analysis. *CoRR*, abs/1502.01657, 2015.

[Pham and Lee(2016)] Thai Pham and Steven Lee. Anomaly detection in the bitcoin system - A network perspective. *CoRR*, abs/1611.03942, 2016.

[Ron and Shamir(2012)] Dorit Ron and Adi Shamir. Quantitative analysis of the full bitcoin transaction graph. Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Israel, 2012.

[Zambre and Shah(2013)] Deepak Zambre and Ajey Shah. Analysis of bitcoin network dataset for fraud. 2013.