Exploring Entity Behavior on the Bitcoin Blockchain

Petra Isenberg, Christoph Kinkeldey, Jean-Daniel Fekete

Exploring Entity Behavior on the Bitcoin Blockchain

Petra Isenberg*
Inria, Université Paris Saclay

Christoph Kinkeldey†
Inria, Université Paris Saclay

Jean-Daniel Fekete‡
Inria, Université Paris Saclay

Abstract

We contribute a visual exploration system for analyzing the behavior of individual entities exchanging Bitcoins. Bitcoin is a cryptocurrency, popular for allowing pseudonymous financial transactions. The Bitcoin blockchain is the public ledger of the Bitcoin system holding data on millions of individual transactions between pseudonymous addresses. These addresses belong to individual *entities* such as people, services, or enterprises. Understanding how the Bitcoin system is used, however, is difficult because it is unclear which addresses belong to the same entities. Our tool addresses this problem by clustering addresses and displaying transaction detail for individual entities.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Screen design

1 Introduction

Bitcoin is a decentralized cryptocurrency that allows for digital payments to be made without the control of a central authority. Bitcoins can be transferred between pseudonymous *addresses* in the form of public key hashes—this means that no personally identifiable information is requested or stored about entities involved in a transfer. Bitcoin's bookkeeping system is a public evolving ledger called the *blockchain* that stores transaction information visible to anyone. Since privacy is important to Bitcoin users, they are encouraged to generate new addresses each time they receive Bitcoin payments so that their entire payment history cannot be easily tracked.

When a single entity is represented by multiple addresses, it is very difficult to study how Bitcoin is used in general or how its use compares to traditional currency systems. For this reason, clustering algorithms have been proposed [16] to group addresses that likely belong to the same entity on the network – being a person, an enterprise, or a service. Simply put, the clustering works on the assumption that multiple input addresses in a transaction all belong to the same entity. This heuristic has been shown to work well in practice as a transaction has to be signed using private keys that match all inputs and, while doable, it is cumbersome to sign a transaction with private keys belonging to different entities.

While researchers are interested in understanding Bitcoin use and even try to attack the pseudonymous nature of Bitcoin, only few visual analytics tools exist that help users or hobby analysts to understand better how they are themselves or how others are represented pseudonymously on the blockchain. We are working on a publicly available tool that represents the transaction history of a specific entity on the network based on a given input address.

2 Related Work

A number of startup companies offer data and services to analyze the Bitcoin blockchain. A common business model is to offer access to extracted information from the blockchain for analysis (e. g., [10]). This is often complemented by services such as the detection of illegal activity like money laundering [7, 8], a goal shared with the visual analytics tool BitConeView by Battista et al. [1]. The funding success of some of these startups reflects the growing importance of blockchain analysis tools. Most existing tools and websites focus on providing simple charts showing descriptive statistics over time such as the Bitcoin market price, information on blocks and transactions [4, 5] or graphs showing connections between sending and receiving addresses [8]. In addition, a few Bitcoin-related visualizations exist that visualize transactions in realtime [2, 9, 12, 13] or show information on the creation of new Bitcoins [3, 6].

Our work is related to these existing projects in that we also aim to provide public access to blockchain data. We complement existing approaches, however, by providing an entity-based view on individual transaction data. We extend our own previous work on BitConduite [11] that aggregates entity-based behavior information but does not provide the detail of individual transactions. Our tool provides views on the Bitcoin blockchain that are not currently available in any other analysis software.

3 System Design

Our system for analyzing entity-based transaction is designed to allow Bitcoin users and casual blockchain analysts to visually track their own or others' transactions based on a single input address.

3.1 Data Processing

We extracted our input data from the Bitcoin Core client and stored it in a MongoDB database. We make use of Reid and Harrigan's clustering heuristic [16] to combine addresses for individual entities. We store the resulting clusters together with other aggregate information in a MonetDB [14] database. As of 06/2017 the whole blockchain contains about 120GB of data and 284,000 transactions. About 600,000 unique addresses are used in transactions daily. As clustering this amount of data takes a very long time, our development currently uses only a two-year subset of the data between 01/2009 and the end of 2010.

3.2 Visual Analytics Interface

Fig. 1 gives an overview of our tool using an input address involved in a transaction with Satoshi Nakamoto, Bitcoin's mysterious founder [15]. The person(s) behind Nakamoto's name have to this date not been identified and therefore analyzing transactions with him/her is interesting to many. Other input addresses can be entered either by pasting an address in the textbox (Fig. 1 A) or choosing one of the orange option buttons.

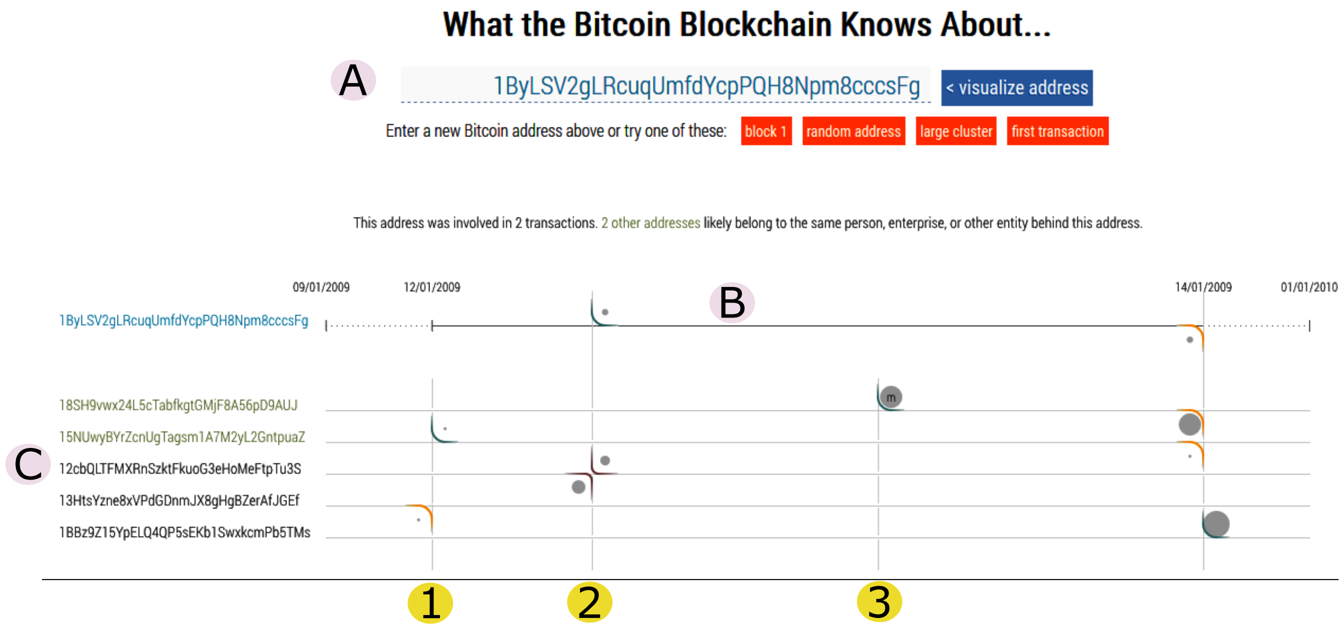
Next, the visualization shows an activity timeline (Fig. 1 B) for the chosen address and all other addresses belonging to the same entity. The solid portion of the timeline represents the period of time in which the entity was active while the endpoints connected by a dotted line show the temporal context of the whole dataset. In Fig. 1 we loaded data from Jan. 9, 2009–Jan. 1, 2010 according to Central European time. We can see that this entity was active between the 12th and 14 th of January.

Each of the entity's addresses as well as external output addresses in any of the entity's transactions are represented in rows below the starting timeline. Green-colored addresses belong to the same entity as the input address, while black external addresses belong to others. Address C in Fig. 1 is known to belong to Nakamoto.

[*] e-mail: petra.isenberg@inria.fr

[†] e-mail:christoph.kinkeldey@inria.fr

[‡] e-mail:jean-daniel.fekete@inria.fr



What the Bitcoin Blockchain Knows About...

Enter a new Bitcoin address above or try one of these:

This address was involved in 2 transactions. 2 other addresses likely belong to the same person, enterprise, or other entity behind this address.

A
1ByLSV2gLRcuqUmfdYcpPQH8Npm8cccsFg
< visualize address
block 1
random address
large cluster
first transaction
09/01/2009
12/01/2009
14/01/2009
01/01/2010
B
1ByLSV2gLRcuqUmfdYcpPQH8Npm8cccsFg
18SH9vwx24L5cTabfkgtGMjF8A56pD9AUJ
15NUwyBYrZcnUgTagsm1A7M2yL2GntpuaZ
C
12cbQLTFMXRnSzktFkuoG3eHoMeFtpTu3S
13HtsYzne8xVPdGDnmJX8gHgBZerAfJGEf
1BBz9Z15YpELQ4QP5sEKb1SwxkcmPb5TMs
m
1
2
3

Figure 1: Overview of our Bitcoin Blockchain Entity Explorer. The data shows information on an input address (A) involved in a Bitcoin transaction (2) with Satoshi Nakamoto (Bitcoin's founder). This entity was involved in 4 transactions (vertical lines) and also mined Bitcoins (3).

Each transaction on the timeline is represented by a vertical line connecting to horizontal lines emanating from the displayed addresses. When an address is involved in a particular transaction, glyphs are drawn at the intersection of transaction and address lines. Transaction 1 in Fig. 1 involves Bitcoins sent from address `13HTs...` as indicated by an orange in-glyph. One of the entity's addresses (`15NU...`) receives the Bitcoins as indicated by a blue out-glyph. The size of the circle in each glyph represents the amount this address received or sent. An address serving both as a sender and receiver is connected by a red in-out-glyph such as Nakamoto's address involved in Transaction 2. Some out-glyphs such as the one in Transaction 3 include an "m" standing for coins newly created through a process called mining (akin to printing money in a regular currency). This interface communicates how an entity's addresses are connected internally and to external addresses.

3.3 Interaction

The interface invites users to casually explore the Bitcoin information space through pivoting operations. These operations allow new input addresses to be chosen from any of the displayed addresses. Once an address is clicked on, the visualization shows data for the new input address. In addition, detail on demand on transactions is displayed on hover over the transaction line or any drawn glyphs.

4 Challenges and Future Work

One of the main challenges with the current design is scalability. For the two years covered by our current data snapshot, the largest entity already has 901 addresses. We expect the cluster sizes to grow rapidly especially for data from recent years where the process of creating new addresses has been extremely simplified in Bitcoin management software. We are therefore currently implementing a multi-scale data exploration approach where large entities are first displayed in an aggregated manner. We envision these aggregated clusters to be opened up into sub-clusters grouped by similarity. We will also use focus+context techniques on the timeline to allow for more detailed exploration of specific transaction clusters coupled with filtering to only the involved addresses.

Acknowledgments

This research was partially supported by Labex DigiCosme (project ANR-11-LABEX-0045-DIGICOSME) operated by ANR as part of the program "Investissement d'Avenir" Idex Paris-Saclay (ANR-11-IDEX-0003-02)

References

[1] G. D. Battista, V. D. Donato, M. Patrignani, M. Pizzonia, V. Roselli, and R. Tamassia. Bitconeview: Visualization of flows in the bitcoin transaction graph. In *Proc. Symposium on Visualization for Cyber Security (VizSec)*, pp. 1–8, Oct 2015. doi: 10.1109/VIZSEC.2015.7312773

[2] Bitbonkers.com. Website. `bitbonkers.com`, visited 06/2017.

[3] Bitnodes. Website. `https://bitnodes.21.co/`, visited 06/2017.

[4] Blockchain.info. Website. `https://blockchain.info/charts`, visited 06/2017.

[5] Blockr.io. Website. `http://blockr.io/`, visited 06/2017.

[6] Blocks.Wizb.it. Website. `https://blocks.wizb.it/`, visited 06/2017.

[7] Chainanalysis. Website. `https://www.chainalysis.com/`, visited 06/2017.

[8] Elliptic.co. Website. `https://www.elliptic.co/`, visited 06/2017.

[9] L. Hendriks. Website. `http://bitcoin.interaqt.nl/`, visited 06/2017.

[10] Kaiko.com. Website. `https://www.kaiko.com`, visited 06/2017.

[11] C. Kinkeldey, J.-D. Fekete, and P. Isenberg. BitConduite: Visualizing and Analyzing Activity on the Bitcoin Network. In *Eurographics Conference on Visualization (EuroVis), Posters Track*, 2017.

[12] M. Laumeister. Website. `http://www.bitlisten.com/`, visited 06/2017.

[13] D. McGinn, D. Birch, D. Akroyd, M. Molina-Solana, Y. Guo, and W. J. Knottenbelt. Visualizing dynamic bitcoin transaction patterns. *Big Data*, 2(4):109–119, 2017. doi: 10.1089/big.2015.0056

[14] MonetDB. Database. `http://www.monetdb.org/`.

[15] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. Published online., 2008. `http://bitcoin.org/bitcoin.pdf`.

[16] F. Reid and M. Harrigan. *An Analysis of Anonymity in the Bitcoin System*, pp. 197–223. Springer New York, New York, NY, 2013. doi: 10.1007/978-1-4614-4139-7_10