Summary of **Bitcoin Transaction Graph Analysis** by **Michael Fleder, Michael S. Kester, Sudeep Pillai** published on **February 6, 2015**

Introduction: Bitcoin has gained popularity over a period of time in terms of trading electronically using cryptocurrency. It uses a public distributed ledger to keep track of all the transactions in Bitcoin network. In Bitcoin blockchain system, users are identified based on their cryptographic public ids. This feature made users think that their anonymity is preserved. However, some recent studies prove that it certainly has some privacy concerns. In this paper, authors try to break anonymity in Bitcoin systems using graph analysis. The authors first try to link public keys to actual user's names and then create a users graph and come up with a summary of users activity which can help any attacker to break the privacy.

Methodology

In the threat model used for research, it is assumed that attacker has access to all the information available online and user can overhear transactions which may not be 100% accurate. Steps to create graph are as below:

1) Scraper is used to fetch information such as Bitcoin addresses from public forums available on web which is shared by users willingly or unwillingly. Usernames available in public forums were then linked with their Cryptographic Bitcoin addresses.

2) That information is then linked to the transactions in Bitcoin.

Implementation

The implementation includes Preprocessing of data in Bicoin Blockchain and Transaction fingerprinting.

1) Preprocessing of Data: Bitcoin Blockchain is downloaded and its data is parsed using a tool called Armory to retrieve information required to build a graph.

2) Web Scraping: A python package called Scrappy is used to fetch information from web forums via a crawler which fetches information from bitcointalk.org in breadth-first manner. Fetched information is verified with data parsed in step 1. The result of this step is the identification of large number of users. Code was run for less than 30 hours and a total of 222 users were identified with their overall 2404 addresses.

3) Transaction fingerprinting: Matching the roughly overheard transactions with actual transactions in Bitcoin blockchain. The amount is first converted using daily market price from Bitcoin. The overheard amount is compared with an approximation of +- $1 in the blockchain and time of the transactions is approximated to +- 5 minutes.

4) Graph analysis: Metrics evaluated from graph analysis are: most important nodes in the network, the assets of these nodes, the movement of assets between nodes, nodes having high number of transactions.

5) PageRank: PageRank algorithm is used to find the some of the interesting users based on their activities and the transactions made by them.

6) User De-anonymization: Identities of some of the interesting users were revealed.

Results

The authors were able to link the transactions associated with SatoshiDICE and Wikileaks. They were also able to link users in forum with nodes belonging to SIlk Roads. It is hence proved in the paper that the system is not entirely anonymous.