PERSPECTIVES ON THE DETECTION OF ANOMALOUS BEHAVIOR IN BLOCKCHAIN NETWORKS

Hunter Kippen hmk64@drexel.edu
Andrew Wiggins afw42@drexel.edu
Ashish Shrestha as3828@drexel.edu
Minhea Maris mmm475@drexel.edu
Scott Shevrin sis442@drexel.edu

Abstract: Anonymity in digital currencies like Bitcoin has made it a hotbed for malicious actors as illicit transactions are easy, secure and practically untraceable. This paper explores on why digital currencies are used by malicious entities, and explore detection techniques that could be employed to sort out anomalous transactions using machine learning techniques, primarily the ones investigated by Hirschman et. al. in the paper titled - "Unsupervised Approaches to Detecting Anomalous Behavior in the Bitcoin Transaction Network".

INTRODUCTION

Since 2009, Bitcoin and similar blockchain technologies have been at the forefront of decentralized currency systems. As with Bitcoin [1], and Ethereum [2], the public ledgers of these technologies provide the necessary trust in the system. This is similar to that of traditional financial institutions, but with some key distinctions. In current public blockchain networks, users are authenticated using asymmetric cryptography. Effectively each account is associated with a public key address rather than a user's actual information.

Recent research has found that the pseudo-anonymity afforded to users by these networks is enabling anomalous behavior on these networks that would not normally be present in the SWIFT network employed by the large financial institutions for electronic fund transfers [3]. This anomalous behavior includes money laundering, transaction mixing, and similar activities that would normally get flagged on the SWIFT network.

Stopping these anomalous behaviours is essential for cryptocurrencies to be seen as legitimate alternatives to traditional fiat currency. To stop such anomalies from occuring, they must be detected first. Since the blockchains of most large cryptocurrencies are public, it is relatively simple to collect large amounts of data from these networks. The transactions between addresses can be translated to edges on a graph, while the addresses themselves can be nodes. Since there is no current ground truth data on what constitutes a 'normal' user on a blockchain network, unsupervised learning techniques must be used to separate users into categories, where more scrutiny can be placed on outliers [4].

ANOMALOUS BEHAVIOUR IN THE BLOCKCHAIN

Illegal activities like drug dealing, human trafficking, and money laundering, require safe and untraceable means of transaction between malicious parties. Traditional stores of value like cash and gold were the de facto standards for these transactions. But physical cash and gold are not easy to transport and store without raising suspicion. Banks are centralized, bank accounts are personally identifiable, and all transactions are traceable. With the advent of blockchain

based store of value, beginning with the inception of Bitcoin, anonymous transaction without require physical storage and exchange is now possible. In 2010, centralized entities, Visa and Mastercard, blocked donation payments to Wikileaks. Wikileaks since then have accept donations by Bitcoin by publishing their Bitcoin address. [5]

Bitcoin was not built to be truly anonymous, rather anonymity is the consequence of Bitcoin's dependency on public addresses and private keys for transactions, and almost all the other cryptocurrencies following it. But since the ledger itself is public and transactions are broadcasted to all nodes, Bitcoin is completely deterministic and uncensored. Although public addresses are not directly associated with an individual, Bitcoin is the most traceable currency till date. [6]

The public ledger a.k.a. the blockchain, ensures that every transaction made on the network is preserved forever. Block explorer services provides tools to investigate every transaction on the chain. Complex tracing algorithms can be developed to track movement of assets on the blockchain. Continuing from the previous mentioned detail of Wikileaks, anyone can explore the ledger and determine that on December of 2013 3000 bitcoins were sent to several other Bitcoin addresses, followed by other similar transactions. [7] Eventually, when a Bitcoin address performs transaction with a known entity the actor might expose their identity.
To remove the stench from the "dirty" crypto-coins, mixing services provide a mechanism for pseudo-anonymity. Mixing services, as the name specifics, mixes crypto-coins by performing a series of random transactions thereby obfuscating the origin and destination of transaction similar to how the Tor network performs source and destination obfuscation. This makes it difficult for tracking entities to determine the actual destination address of transaction.

Machine learning algorithms might potentially be able to connect the origin and destination addresses that underwent mixing services. Coin mixing services might leave traces of complex but identifiable patterns for machine learning algorithms. Furthermore, the behavioral characteristics of blockchain addresses owned by coin mixing services should be determinably different from that of human owned account like frequency of transactions and amount contained in each transaction.

TRANSACTION MIXING

Transaction mixing is a technique used to anonymize the coin owned by a certain person. This is realized by sending the coin from the individuals requiring anonymity to random addresses and getting getting it back the same amount, but from different sources. Thus, the coins you used to own will now be associated with someone else, making it confusing to trace where the money come from.

Although the process itself is useful for keeping anonymity, mixing services can be used for money laundering or criminal activities. Although the data used in the paper only used unencrypted transactions, the anonymity can increase even more using encryption, encryption padding to avoid size-based identification, sending dummy packets and randomizing the sending time [8], but the process will never be impossible to trace back. This would make it harder to analyze the data.

BITCOIN DATASET

The dataset used for the paper [4] is no longer available at the reference specified, however a detailed description of the data is available. The data stores the transaction key, the public keys of the users , the date and value of the transactions. The dataset associates each key to a user, however, some transaction have multiple public key as input. This means that the authorizing entity had access to both private key, thus, they can be associated to the same user.

Finer details are missing from the dataset, such as bitcoins transferred and the addresses they were transferred to, however those details can be queried from the bitcoin blockchain using the transaction keys and public keys available. These details help associate public kes to users and run the clustering algorithms proposed later in the paper.

K-MEANS CLUSTERING

K-means is a classic machine learning clustering algorithm. In general, the goal of clustering algorithms is to partition unlabeled data into groups, or clusters. This is a type of unsupervised learning, since there are no ground truth labels for the data. Rather than seeing how well the system can classify the data into existing classes, the class groupings are discovered by the algorithm.

In the K-means algorithm, in addition to the input data, a value k must also be input which represents the number of clusters the data should be partitioned into. The K-means algorithm is defined formally as follows. Let $X = \{x_1, x_2, ..., x_n\}$ be the set of n points that are to be clustered. Let $C = \{c_1, c_2, ..., c_k\}$ be the k clusters of points. Let μ_i be the mean, or centroid, of cluster c_i. The objective of the K-means algorithm is to find clusters C that minimize the loss function L. The loss function is defined as the sum over all clusters of the sum of squared differences between each point and its cluster's mean:

$$L = \sum_{i=1}^{k} \sum_{x_j \in c_i} \left\| x_j - \mu_i \right\|^2$$

While using this Euclidean distance is the most common distance metric in K-means, other distance metrics may also be used.

With no direct method to arrive at the global minimum of the loss function, the only way to ensure the global minimum is reached is to apply brute force and compute every combination of clusters. Since this is excessively computationally expensive, approaches to computing a decent clustering with K-means focus instead on finding a local minimum to the function. This can be done somewhat efficiently by utilizing the following iterative algorithm which will converge to a local minimum of the loss function.

First, the algorithm is initialized by randomly creating a partition of k clusters to serve as the starting clusters and these clusters' centroids are computed. Next, the following two steps (A and B) are repeated until the algorithm converges. (A) The points are assigned to new clusters, where each point is assigned to the cluster with the minimum distance between the point and the cluster centroid. (B) New cluster centroids are computed based on the updated clusters. Finally, after alternating steps A and B, the algorithm finally ends once the clusters do not change. The resulting clustering is the output of the algorithm. [8]

In the Hirshman paper [4], K-means is used as a basic clustering technique to identify anomalous groups of points. The K-means algorithm is applied to the transaction data with an input value of $k = 5$. This value of k was determined experimentally, since it was the amount of clusters that provided the smallest overall loss value over 1,000 test trials. After partitioning the data into 5 clusters, the authors focused on the smallest clusters. Since the points belonging to these minority clusters were not absorbed into any larger groups, these points are identified as candidate anomalous transactions.

ROLX CLUSTERING

Another method for clustering the blockchain data was also used by Hirshman et. al [4]. This method was specifically optimized for data that was organized into a graph structure. This method, called RolX, which is short for Role extraction, was designed to separate graph nodes into various 'roles' automatically [9]. These roles would be based on learned structural similarities between nodes. These could include number of in-edges vs out-edges, how many cycles the node is a part of, or other such metrics. The algorithm, according to its creators, is flexible, and scalable, so it can fit largely any problem that would require its use.

The algorithm uses a feature matrix factorization from an $n \times f$ matrix into two nonnegative, $n \times r$ and $r \times f$ matrices, where n is the number of nodes, f is the number of features, and r is a specific low rank found through minimization using KL divergence. After factorization, r becomes the number of roles found in the graph.

Hirshman et. al, used RolX to validate their clusterings found by the K-means algorithm. However, due to the presence of zero valued features in their data, they could not find r using the minimization techniques outlined by RolX's creators. This is likely due to the fact that the transaction network outlined in their dataset was sparse. Most users only send transactions to a few different places, and these hubs are rarely also interconnected. As such, they used the Frobenius Norm of each column of their matrix to calculate the factorization error. The Frobenius Norm is effectively the L-2 norm for vectors but done over the entire matrix. For an MxN matrix, the equation is:

$$||A||_F = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} |a_{ij}|^2}$$

Hirshman et. al. found that the minimal error occurred when there were seven roles associated with their data.

This alternative method worked because the primary reason for using KL divergence for minimization was to ensure all roles were necessary, when there is no need for such guarantees when the only interest in the data are the outlying clusters that may contain abnormal behaviors. The clustering from RolX provided similar data to that of K-Means. There were outlying clusters of data that showed hubs with large variance in either transaction amounts. More specifically, there were two roles that were found that contained these high variance transactions. One role weighted the incoming transaction variance more heavily, while the other weighted the outgoing transaction variance more heavily. There was also large amounts of overlap between these two roles and that of the small hubs found using K-means.

FINDINGS

Although there are no details of transactions by individual public keys there is still enough data to notice distinct patterns. Transactions can be traced from one source to a singular destination through many binary intermediate sources and transactions. This can be done to obfuscate the true owners of the coins and make it near impossible to track the money by the time the money reaches the destination.

Hirshman et al. demonstrate this principle through a bitcoin laundry service readily available for the public, and track their own coins through each transactions. When they examined the outlying clusters of data found in both K-Means and RolX, they found that the clusters with the fewest members overlapped significantly. These clusters contained users with high transaction variance, meaning they were sending or receiving inconsistent amounts of coin. When they examined these users more closely, it was found that they were participating in a mixing service [4].

CONCLUSION

In summation, Hirshman et al. have demonstrated that it is possible to detect malicious behavior by identifying anomalous transaction clusters in the Bitcoin transaction network. Since the publishing of this paper, cryptocurrency and blockchain technologies have moved away from UTXO-based networks in favor of those with smart contract behavior, such as Ethereum. Future work in the detection of malicious behavior in blockchain will require the extension of these techniques to smart-contract-based networks. Additionally, it is possible that the clustering techniques used in [4] may not be extensible to the Ethereum network. Therefore, we suggest that additional research focus on applying deep learning architectures to this clustering problem, as research across various domains have demonstrated their power in generalization.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System", 2009.

[2] V. Buterin, "A Next-Generation Smart Contract and Decentralized Application Platform", 2013.

[3] S. Jovicic and Q. Tan, "Machine Learning For Money Laundering Detection In The Blockchain Financial Transaction System", *Journal of Fundamental and Applied Sciences*, vol. 10, no. 4, pp. 376-381, 2018.

[4] J. Hirshman, Y. Huang and S. Macke, "Unsupervised Approaches to Detecting Anomalous Behavior in the Bitcoin Transaction Network", 2013.

[5] BohannonMar, John, et al. "Why Criminals Can't Hide behind Bitcoin." *Sciencemag | AAAS*, American Association for the Advancement of Science, 9 Dec. 2017, www.sciencemag.org/news/2016/03/why-criminals-cant-hide-behind-bitcoin.

[6] Young, Joseph. "Michael Perklin: Bitcoin Is One of the Most Traceable Currencies on the Planet." *Binary District Journal - Unfiltered Access to the Minds behind the Tech.*, 28 Dec. 2017, journal.binarydistrict.com/michael-perklin-bitcoin-is-one-of-most-traceable-currencies-on-the-planet/.

[7] Cox, Joseph. "Where Did WikiLeaks' $25 Million Bitcoin Fortune Go?" *The Daily Beast*, The Daily Beast Company, 28 Dec. 2017, www.thedailybeast.com/where-did-wikileaks-dollar25-million-bitcoin-fortune-go

[8] Bitcoin Forum. https://bitcointalk.org/index.php?topic=241.0

[9] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters,* vol. 31, *(8),* pp. 651-666, 2010.

[10] K. Henderson *et al*, "RolX: Structural role extraction & mining in large graphs," in 2012, . DOI: 10.1145/2339530.2339723.