Detection of Blockchain Transactions Used in Blockchain Mixer of Coin Join Type

Artem A. Maksutov [1], Maxim S. Alexeev, Natalia O. Fedorova, Daniil A. Andreev
Department of Computer Systems and Technologies
National Research Nuclear University "MEPhI"
Moscow, Russian Federation
[1]AAMaksutov@mephi.ru

Abstract — **This paper is devoted to both the detection of transactions that can participate in money laundering schemes and the possibility of modification of payment systems based on blockchain technology in the interests of AML/CFT. In the process of work, the research of transaction anonymization method with the use of a decentralized approach based on Coin Join transactions is performed. Also, methods of tracking such transactions, the possibility of their deanonymization and the ability to determine the relationships between users and their transactions are examined. The results show that tracking Coin Join transactions is a feasible task that allows to determine the fact of user participation in the creation of a transaction, which gives additional advantages in the process of detecting transactions involved in money laundering schemes.**

Keywords — blockchain; bitcoin; blockchain mixer

I. INTRODUCTION

Bitcoin was invented in 2008 with the publication of an article called "Bitcoin. The Peer-to-Peer system of Electronic Money," from the author under the pseudonym Satoshi Nakamoto. Nakamoto combined several previous inventions, such as b-money and HashCash, and on this basis created a fully decentralised electronic cash system that would be independent of the Central office for issuing funds and verifying transactions. A key innovation was to use a distributed computing system (the so-called "proof of work" algorithm) to conduct global "elections" every 10 minutes, which allowed the decentralised network to arrive in a state of transaction consensus. This approach elegantly solved the issue of double spending when one currency could be spent twice. Previously, the problem of double spending was a weak point of digital currencies, which was solved with the help of classical centralised clearing. The Bitcoin network itself began its existence in 2009 from the first implementation of Satoshi Nakamoto. Subsequently, the code has been revised and rewritten by other programmers. The network of distributed computing, which provides the system with security and reliability, has increased exponentially in size and currently exceeds the capacity of the largest supercomputers. The total market value of Bitcoin is approximately between $ 5 billion and $ 10 billion depending on the current exchange rate. The largest transaction carried out through the network to date was such a volume of $ 150 million. The transfer was effected almost instantaneously, and almost nothing is its owner is not worth it. Satoshi Nakamoto disappeared from public view in April 2011, shifting the responsibility for the development of the system on the shoulders of a group of volunteers. Who was behind the identity of the virtual Creator of Bitcoin, is still unknown. However, neither Satoshi Nakamoto nor anyone else has control over the Bitcoin system, which operates by transparent mathematical principles. The invention itself was innovative and has already advanced science in the field of distributed computing, Economics, and econometrics. Satoshi Nakamoto's invention of the unsolved problem is a distributed efficient computation, a solution to the previously known " problem of Byzantine generals."In short, the problem is trying to agree on the course of action by exchanging compromised information to communication channels. By unreliable Solution and Satoshi even potentially Nakamoto, uses the concept of" proof of work " to reach consensus without a Central trusted centre and represents a breakthrough in the science of distributed computing that is widely used outside of cryptocurrencies. This principle can be used to achieve consensus among decentralised networks, to prove the fairness of elections, lotteries, asset registers, digital notarization, etc.

II. FORMULATION OF THE PROBLEM

Despite the large number of advantages of Bitcoin, this cryptocurrency, like the rest, is not perfect. The main disadvantage of bitcoin is that all transactions are publicly available to each member of the network, and, therefore, anyone can track the necessary transaction and use the information for various purposes. So despite the fact that the addresses in the bitcoin network do not identify a specific user, when withdrawing the amount of money out through different electronic payment systems and exchanges, the address in the bitcoin network can be compared with a physical or legal entity person. Some studies on this topic demonstrate how basic heuristics can be used to classify Bitcoin addresses that are likely to belong to the same user. Also, some studies show that it is possible to determine the relationship between Bitcoin address and IP address.

To eliminate this drawback, active users of the network, the developers began to come up with algorithms that allow you to anonymize transactions - obfuscate paths, mixing addresses of different recipients and senders. This method reduces the ability to track a particular user's transaction chain.

Services that provide such an opportunity are called bitcoin mixers.

978-1-7281-0339-6/19/$31.00 ©2019 IEEE

However, such services are also not safe, because they deal directly with users ' money. There are a large number of cases of fraud with the use of such services.

The CoinJoin algorithm developed by Gregory Maxwell comes to the rescue. The algorithm does not make changes to the Bitcoin Protocol and ensures that even a malicious mixer-service will not be able to steal bitcoins, however, there is a significant drawback in It: the mixer-service still needs to have the proper level of user trust, because the service itself has information about the connection of transactions with specific users.

Despite all the benefits of such services, their services can be used not only by honest users of the network but also by criminals and organisations. These services can be used to perpetrate serious crimes, such as trafficking in arms, drugs and even human beings. Therefore, the possibility of complete anonymisation can lead to very serious consequences. To detect such crimes may need a tool to de-anonymize transactions processed mixer.

III. SOLUTION

For the analysis of the transaction is necessary to build two graphs: the transactions graph and users graph.

A transactions graph is a stream of bitcoins between transactions over time. Each vertex of this graph plays the role of a transaction. Each edge of this graph is directional and acts as a link between the output of the source transaction and the input of the recipient transaction. Each edge also includes the BTC value passed in the output and the transaction time.

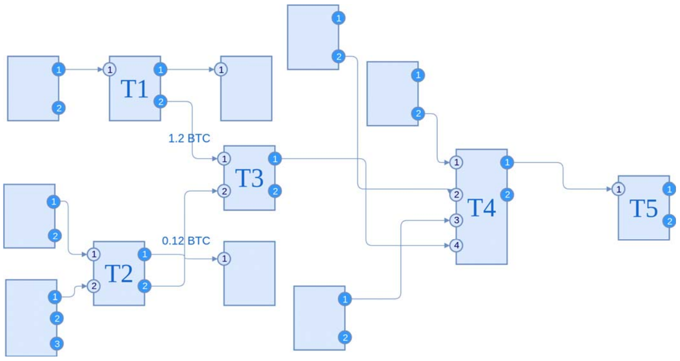

Fig. 2. Transaction graph

Each vertex of this graph plays the role of a transaction. Each edge of this graph is directional and acts as a link between the output of the source transaction and the input of the recipient transaction. Each edge also includes the BTC value passed in the output and the transaction time.

This graph has no cycles and no multi-edges. It is a directed acyclic graph.

The user graph describes the interaction between users over time. Each vertex of the graph corresponds to one user, and each directed edge between the sender and the recipient is an input-output pair of one transaction.

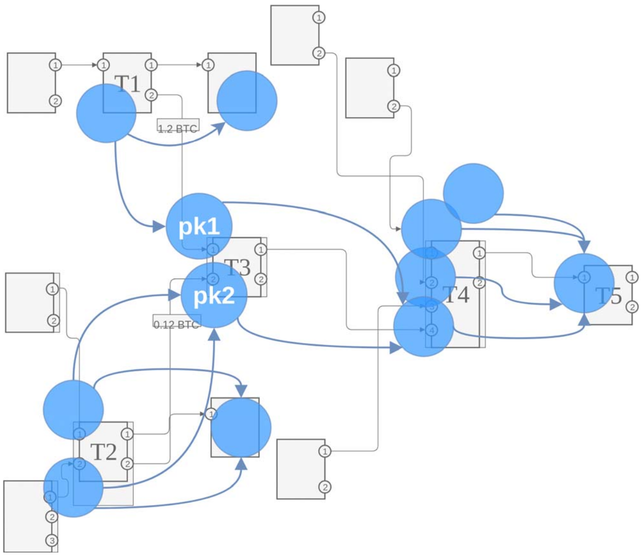

Fig. 2. Users graph

The user graph describes the interaction between users over time.

Each vertex of the graph corresponds to one user, and each directed edge between the sender and the recipient is an input-output pair of one transaction.

You need to take a preliminary step before you can design a user graph. Suppose that the graph is imperfect at first, in the sense that each vertex represents a public key rather than a specific user.

To bring the graph to a perfect view, you need to combine all the vertices that correspond to the public keys of one user. The difficulty is that public keys are a mechanism in the Bitcoin system created to provide anonymity: the public can see that someone (identified by a public-key) is sending an amount to someone else (identified by another public-key), but without information linking the transaction to anyone. Generally speaking, it is a good practice for a bitcoin recipient to generate a new public key each time they send money so that the transactions sent to them are not linked.

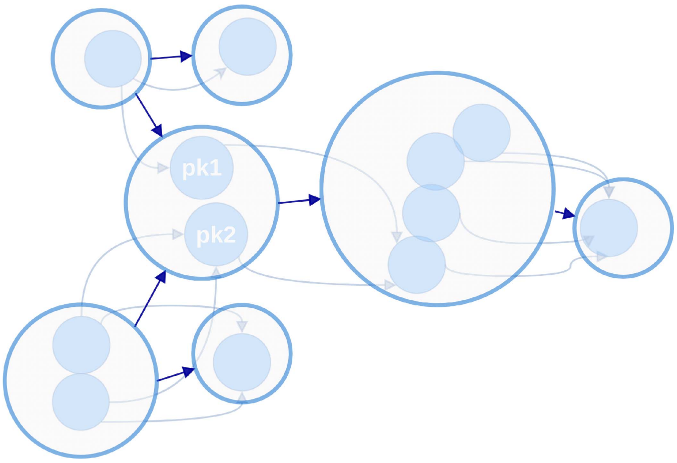

Fig. 3. Perfect users graph

For anonymity analyse, a mathematical model was developed that describes CoinJoin transactions. It excludes the Commission charged by miners for including a transaction in the block to simplify the model. The results obtained during the analysis of this model can be taken as the lower bound of the estimation of the complexity of de-anonymization of transaction data. In practice, the analysis of CoinJoin transactions will require more resources.

CoinJoin transaction T = (O, I, v) consists of inputs I = {i_1, ..., i_n} and outputs O = {o_1, ..., o_n}. Since the model does not include commissions, the sum of bitcoin values at the inputs and outputs must match:

$$\sum_{i\in I} v(i) = \sum_{o\in O} v(o)$$

The CoinJoin transaction consists of subtransactions tk, each, in turn, contains inputs Ik and outputs Ok respectively. For heuristics usage, it is necessary to break the CoinJoin transaction to its sub-transactions. However, it is unknown how many subtransactions a single CoinJoin transaction contains. It is only known that the sum of the values of the transmitted bitcoins at the inputs and outputs must coincide and that there can be no situation when the same output is in two subtransactions. Therefore, there must be at least one way to partition all inputs and outputs so that each subset of inputs has only one corresponding subset of outputs.

Each CoinJoin transaction has at least one valid mapping, where I' = I and O' = O, i.e. consists of only one subtransaction. Such a mapping is unlikely, except only if it was an ordinary transaction and not a CoinJoin transaction. Each mapping that contains more than one subtransaction can be used to build a new mapping that is a join of those subtransactions.

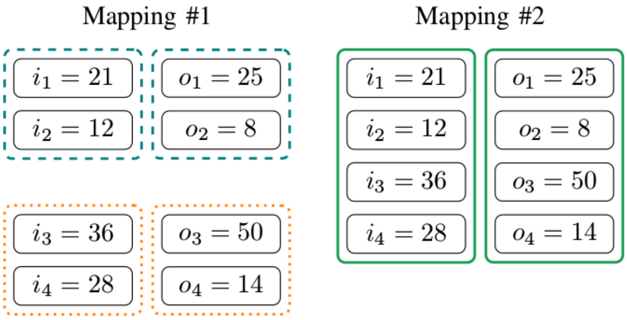

Fig. 4. Coin-join transaction example with 2 subtransactions

Based on the definition of the mapping, we can determine the probability $p_{II}(i_1, i_2)$ that two inputs belong to the same subtransaction as the number of mappings in which the inputs i_1 and i_2 are elements of the corresponding subset divided by the total number of mappings:

$$i_1, i_2 \in I, p_{II}(i_1, i_2) = \frac{|\{M = (I', O', m) \in M' | \exists I \in I' : i_1, i_2 \in I\}|}{|M'|}$$

The probability of pOO that two outputs belong to the same subtransaction is determined in the same way.

IV. RESULTS

For tests, was realised a simulation of the bitcoin network via the program.

For transactions tracking, all generated transactions were presented as a graph. All CoinJoin transactions were specially marked to be able to verify the correctness of tracking.

Figures 5 and 6 show the graphs of simulated transactions and the tracking of Coin-join transactions in them.

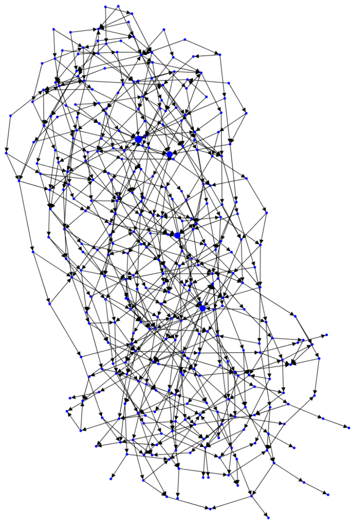

Fig. 5. Generated bitcoin model with Coin-join transaction

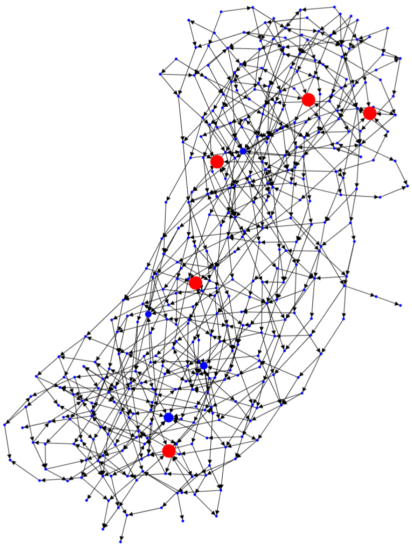

Fig. 6. Generated bitcoin net model with marked Coin-join transaction

V. CONCLUSION

The results show that tracking Coin Join transactions, as well as ordinary transactions, is a feasible task. Developed approach allows determining the fact of user participation in the creation of Coin Join transactions, which gives additional advantages in the process of detecting transactions involved in money laundering schemes.

ACKNOWLEDGMENT

Research was held within the framework of the Program of improving the competitiveness of the National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), as well as in the framework of the priority areas grant program of the Russian Science Foundation “Fundamental and exploratory studies by individual research groups”.

REFERENCES

[1] Maksutov, A.A., Simonenko, A.V., Shmakov, I.S., Classifiers based on Bayesian neural networks // Proceedings of the 2017 IEEE Russia Section Young Researchers in Electrical and Electronic Engineering Conference, ElConRus 2017, p. 700 - 703

[2] Maksutov, A.A., Goryushkin, P.N., Gerasimov, A.A., Orlov, A.A., PRNG assessment tests based on neural networks // Proceedings of the 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2018, p. 339 - 341

[3] Kriesel. D. A Brief Introduction to Neural Networks. (2007).
Available at: http://www.dkriesel.com/ (accesses 1 October 2016)
4th ed. Cambridge University Press, 2010. 480 p.

[4] Alex Graves. Practical Variational Inference for Neural Networks. Conference “Neural Information Processing Systems 24”. Spain, 2011. pp. 2348-24