UDC 004.421.2:519.178 DOI: 10.18698/0236-3933-2022-2-18-25

# BITCOIN CRYPTOCURRENCY ADDRESS GROUPING METHODS

**N.S. Belova** virtseva@list.ru
**I.P. Ivanov** ivanov@bmstu.ru

**Bauman Moscow State Technical University, Moscow, Russian Federation**

**Abstract**

Nowadays, Bitcoin cryptocurrency is an alternative means of payment for purchases in many areas of our lives. However, fraudsters trying to seize cryptocurrency funds and often attack Bitcoin users. In this regard, methods are being developed to determine reliability of the transfer and prevent the loss of funds by the user. Data on all transactions in the Bitcoin network is publicly available, but does not contain any information about the user, except for the cryptocurrency wallet address, and the user is able to create a new transfer address for each transaction. To check reliability of a potential transfer recipient, algorithms for classifying the Bitcoin users are being developed. For classification, it is necessary to join Bitcoin addresses into groups related to the same user. As a rule, address grouping methods are based on the heuristic of combining transaction inputs and the heuristic of determining the recipient address in the transaction. However, such methods provide inaccurate and incomplete results, which leads to development of new approaches having their own advantages and disadvantages, which limits their scope. Data structure of the Bitcoin cryptocurrency blockchain is considered, as well as comparison of existing approaches to address grouping is provided

**Keywords**

*Bitcoin transaction analysis, Bitcoin transaction structure, Bitcoin address grouping algorithms, Bitcoin address clustering*

Received 01.10.2021
Accepted 18.10.2021


**Introduction.** Since its introduction, Bitcoin became an alternative means of payment in many areas, which opened new opportunities to fraudsters trying to seize the funds of their victims. All cryptocurrency transactions are available to anyone making it possible to analyze the Bitcoin transaction data to improve security in using cryptocurrency. However, Bitcoin network does not contain any information about the user except for the address, to which the cryptocurrency funds were transferred. In addition, the user for each new transaction is able to generate a new final transaction address. In this regard,

an urgent task is to select groups of addresses belonging to one user. Classified data makes the further analysis possible to determine reliability of the Bitcoin network user transaction, as well as to find the point of withdrawal of funds received by the fraudsters.

To solve this problem, a large number of algorithms were developed that differ in the volume and novelty of data extracted from the Bitcoin blockchain, as well as in the set of distinguished classes and in the set of data collected for testing and training. The use of incomplete amount of data leads to limitations on applicability of these approaches and affects quality of the results obtained. This paper presents main approaches to solving the problem of address grouping and comparing them. The address grouping algorithms are directly related to the Bitcoin transaction internal structure, so the first section describes the Bitcoin cryptocurrency blockchain data structure. The second section considers ways to group addresses belonging to the same user.

**Bitcoin transaction structure.** To understand the address grouping methods, it is necessary to envision, how Bitcoin network transactions are structured. Bitcoin network data is represented by a chain of blocks, where each subsequent block is connected to the previous one (Fig. 1). Blocks contain a header and a list of transactions, by which Bitcoin users transfer cryptocurrency funds to each other [1]. Block header contains data about the current and previous blocks necessary to maintain the blockchain and ensure its security. The list of transactions contains one transaction with the transfer fee to the block creator (coinbase transaction) and a list of transactions of the other Bitcoin users.

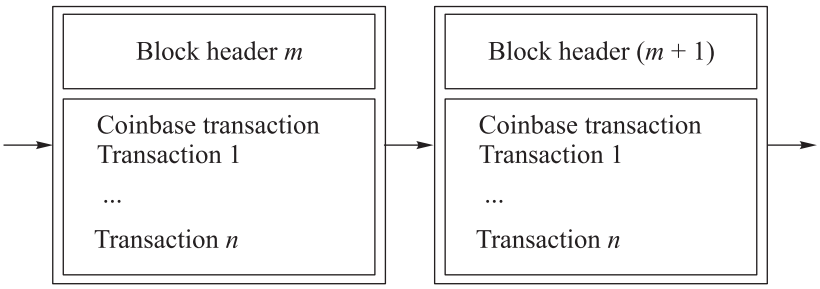


**Fig. 1.** Schematic representation of the Bitcoin cryptocurrency blockchain

To add a transaction, the user needs to create an electronic wallet, which is a pair of generated keys: open and private. Based on the open key, addresses are generated used to transfer Bitcoins between the users [2]. This transfer is carried out using a transaction. Transaction in the Bitcoin network includes a set of inputs and outputs (Fig. 2). Each output contains the amount of trans-

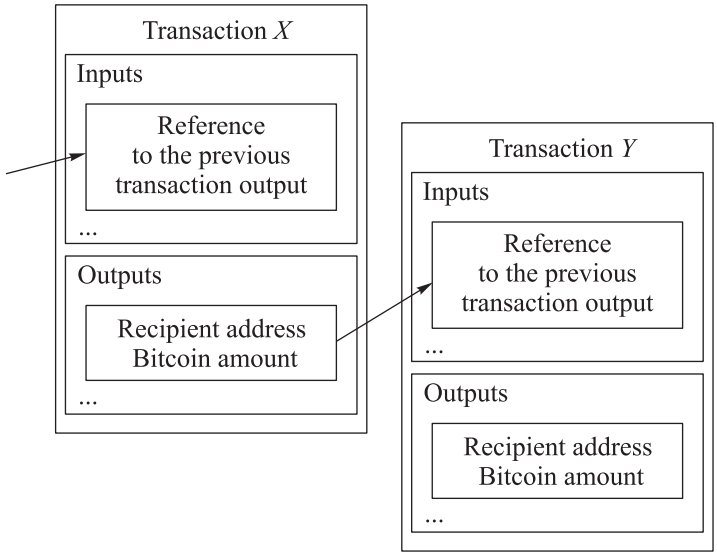


**Fig. 2.** Schematic representation of the Bitcoin cryptocurrency transaction structure

ferred cryptocurrency funds, as well as the address of these funds recipient. Each input is a reference to the previous transaction output or uses Bitcoins not involved in transactions previously. Difference between the amount of Bitcoins in the transaction inputs and outputs is the fee for adding transaction to the Bitcoin network. It should be noted that it is possible in one transaction to spend Bitcoins from the inputs only in full. Therefore, if it is necessary to transfer only part of the funds, transaction is added with an output containing a change returned to the owner of this transaction input funds. In addition, each block contains one transaction transferring the fee of the remaining transactions as remuneration to the user, who created the block [3, 4].

Protection against searching for transaction made by a single user is inherent in the very structure of the Bitcoin cryptocurrency transaction. When preparing a transaction, the user is able to use a new address created based on the open key of his electronic wallet. Thus, new addresses could be introduced in each pair of transactions of a single user, and many owners of cryptocurrency are using this option [5]. In this regard, the task arises to find the ways to group addresses belonging to the same user.

**Methods for address grouping.** Users of the Bitcoin cryptocurrency are able to generate a new final transfer address in each transaction based on the open key of their electronic wallet. Thus, all transfers of a single user would be unrelated to each other in the Bitcoin blockchain. This significantly reduces possibility to classify Bitcoin addresses and, accordingly, to determine reliability

of any funds transfer. To solve this problem, methods are being developed to group addresses belonging to the same user.

The authors of [6] suggest using two heuristics to find Bitcoin addresses belonging to the same Bitcoin user:

– heuristics based on transaction inputs. All inputs of one transaction should belong to one user (Fig. 3);

– heuristics based on transaction outputs. New address is created for each transaction, where the user would receive the change from the funds transfer. Respectively, if one address of the two addresses in the transaction outputs already appeared in the blockchain before, and the second appears for the first time, then it is safe to assume that the new address is the change address, i.e., it belongs to the same user, as the transaction input addresses.

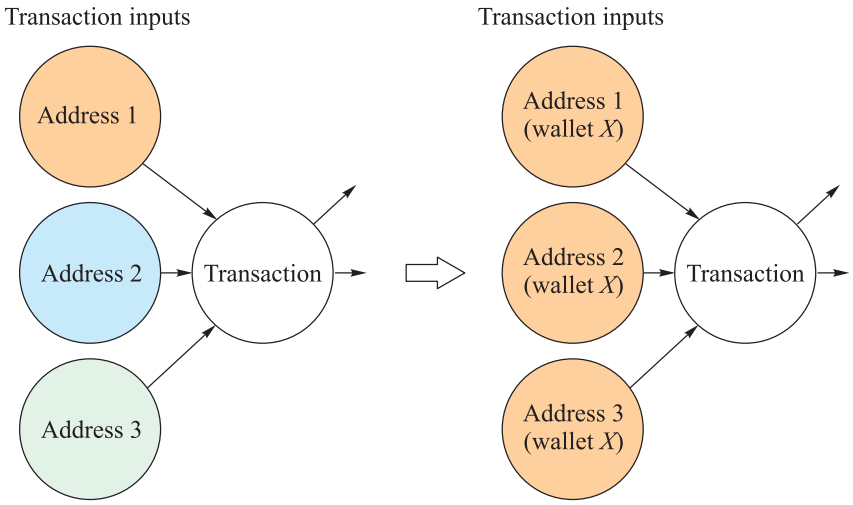


**Fig. 3.** Joining into group different Bitcoin addresses owned by the same user based on transaction input joining heuristic

However, introduction of these heuristics nowadays would lead to erroneous grouping of addresses belonging to different users [7]. This is due to the fact that services appeared making it possible for two users to create one transaction for funds transfer specifying their addresses as transaction inputs and using multiple outputs as funds transfer and change endpoints [8]. Similarly, the second heuristic results in an error, if funds are transferred to more than one user. If in 2013 such a situation could be rare, now such an assumption is erroneous [8]. However, heuristic of the transaction input combination is still used today in address grouping as one of the most appropriate way [9].

Heuristic version for obtaining the change address in a transaction based on three conditions is provided in [10].

1. Transaction should not be of the coinbase type (remuneration to the user, who created the block).

2. Address of the first output appeared in the blockchain for the first time, and the second output was encountered before.

3. There is no address among the outputs, which is among the inputs of the same transaction.

If all three conditions are met, the new address in the transaction outputs is proposed to be attributed to the same user, who owns the addresses from its inputs. The same approach is used in [8]. However, this approach could also lead to an error, if the user himself creates a transaction with two outputs, one of which is his change address already used before, but was not introduced in the inputs of the same transaction.

Another heuristic modification based on outputs is proposed in [11]. Addresses are identified being used as the one-time change addresses, when the following transaction conditions are met:

– transaction contains two outputs;

– the number of inputs is not equal to two;

– both addresses from the transaction outputs do not belong to its inputs;

– one of the addresses in the outputs was not used in transactions before and contains four or more digits in the fractional part of the Bitcoins number;

– another transaction output was already used before and was not defined as the change address.

The authors confirm that significant improvement in the quality of allocated address groups was achieved by adding condition on the fractional part of the Bitcoins number. However, using all conditions still does not guarantee that the change address would be determined correctly for all transactions.

Already known data on address ownership is used in the address grouping [12]. Work [11] proposes to define pairs of classes, to which an address could not belong simultaneously. For example, it was found that addresses of gambling services could simultaneously belong to exchangers, but are incompatible with miners. Accordingly, a condition is added in grouping that addresses belonging to incompatible pairs could not be grouped [11]. These patterns make it possible to reduce the number of incorrectly grouped addresses. However, there are exceptions that do not completely allow avoiding errors in address grouping.

Methods proposed make it possible to group addresses based on assumptions about typical behavior of most users and on the fact that users tend to reuse their addresses [13]. In this regard, new services are being created that complicate solution of the task. For example, work [14] proposes a method for mixing the Bitcoins. Such services are called the mixers. They perform what is called

the Bitcoin mixing by creating a series of transactions between different Bitcoin addresses and dividing these transactions by the time of creation, amount of funds transfer and fee. As a result of a series of transfers, initial amount of the mixer user returns to his own Bitcoin address (minus fee for the used transactions), but the final addresses no longer have any obvious connection in the transaction chains with the initial user addresses. In addition, mixers create additional transactions with inputs owned by multiple users to make it more difficult to find the Bitcoins ultimate recipients in the transaction history, which could also lead to an error in using the input combination heuristic [15].

Therefore, users, if they desire, acquire opportunity to create transactions in such a way that it becomes impossible to group their addresses in the transaction chains. At the same time, such opportunities are being improved, which requires development of new approaches in the Bitcoin address grouping.

**Conclusion.** Overview of the ways to group addresses belonging to one Bitcoin network user is provided. Advantages and disadvantages of the considered approaches are revealed. Most address grouping heuristics rely on the fact that users tend to reuse their addresses, and also that in a transaction, it is possible by a combination of features to identify the address, where the change is transferred. Some approaches take into account the user behavior statistics based on addresses, which owners are known.

Disadvantage of most heuristics lies in fairly high probability of the different users address grouping into one group, which introduces restrictions on the use of such approaches in further analysis of the address groups. Besides, new methods to transfer Bitcoins are constantly being developed making it more difficult to group the user addresses, which requires additional analysis of the user behavior and development of new approaches to grouping the Bitcoin user cryptocurrency wallet addresses. Detailed consideration of approaches to group the addresses taking into account the use of mixers by users in funds transfer remains outside the scope of the current work. The use of mixers is becoming increasingly popular making this an area for further research.

## REFERENCES

[1] Crosby M., Nachiappan P.P., Sanjeev V., et al. Blockchain technology: beyond Bitcoin. *AIR*, 2016, no. 2, pp. 6–19.

[2] Herrera–Joancomartí J. Research and challenges on Bitcoin anonymity. In: Data privacy management, autonomous spontaneous security, and security assurance. Nature Switzerland AG, Springer, 2015, pp. 3–16.
DOI: https://doi.org/10.1007/978-3-319-17016-9_1

[3] Maesa D.D.F., Marino A., Ricci L. Uncovering the Bitcoin blockchain: an analysis of the full users graph. *IEEE DSAA*, 2016, pp. 537–546.
DOI: https://doi.org/10.1109/DSAA.2016.52

[4] Zohar A. Bitcoin: under the hood. *Commun. ACM*, 2015, vol. 58, no. 9, pp. 104–113. DOI: https://doi.org/10.1145/2701411

[5] Conti M., Kumar E.S., Lal C., et al. A survey on security and privacy issues of Bitcoin. *IEEE Commun. Surveys Tuts*., 2018, vol. 20, no. 4, pp. 3416–3452.
DOI: https://doi.org/10.1109/COMST.2018.2842460

[6] Androulaki E., Karame G.O., Roeschlin M., et al. Evaluating user privacy in Bitcoin. *Int. Conf. on Financial Cryptography and Data Security*. Berlin, Heidelberg, Springer, 2013, vol. 7859, pp. 34–51. DOI: https://doi.org/10.1007/978-3-642-39884-1_4

[7] Spagnuolo M., Maggi F., Zanero S. Bitiodine: Extracting intelligence from the Bitcoin network. *Int. Conf. on Financial Cryptography and Data Security*. Berlin, Heidelberg, Springer, 2014, pp. 457–468. DOI: https://doi.org/10.1007/978-3-662-45472-5_29

[8] Tasca P., Hayes A., Liu S. The evolution of the Bitcoin economy: extracting and analyzing the network of payment relationships. *J. Risk Finance*, 2018, vol. 19, no. 2, pp. 94–126. DOI: http://dx.doi.org/10.1108/JRF-03-2017-0059

[9] Harrigan M., Fretter C. The unreasonable effectiveness of address clustering. *Proc. IEEE UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld*, 2016, pp. 368–373.
DOI: https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld. 2016.0071

[10] Meiklejohn S., Pomarole M., Jordan G., et al. A fistful of Bitcoins: characterizing payments among men with no names. *Proc. IMC*, 2013, pp. 127–140.
DOI: https://doi.org/10.1145/2504730.2504747

[11] Ermilov D., Panov M., Yanovich Y. Automatic Bitcoin address clustering. *Proc. IEEE ICMLA*, 2017, pp. 461–466. DOI: https://doi.org/10.1109/ICMLA.2017.0-118

[12] Chawathe S.S. Clustering blockchain data. In: Clustering methods for big data analytics. Nature Switzerland AG, Springer, 2019, pp. 43–72.
DOI: https://doi.org/10.1007/978-3-319-97864-2_3

[13] Harrigan M., Fretter C. The unreasonable effectiveness of address clustering. *Proc. IEEE UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld*, 2016, pp. 368–373.
DOI: https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld. 2016.0071

[14] Valenta L., Rowan B. Blindcoin: blinded, accountable mixes for Bitcoin. *Int. Conf. on Financial Cryptography and Data Security*. Berlin, Heidelberg, Springer, 2015, pp. 112–126. DOI: https://doi.org/10.1007/978-3-662-48051-9_9

[15] Fleder M., Kester M.S., Pillai S. Bitcoin transaction graph analysis. *arXiv preprint arXiv:1502.01657*, 2015. DOI: https://doi.org/10.48550/arXiv.1502.01657

**Belova N.S.** — Assist. Lecturer, Department of Theory of Informatics and Computer Technologies, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

**Ivanov I.P.** — Dr. Sc. (Eng.), Professor, Head of the Department of Theory of Informatics and Computer Technologies, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

**Please cite this article as:**

Belova N.S., Ivanov I.P. Bitcoin cryptocurrency address grouping methods. *Herald of the Bauman Moscow State Technical University, Series Instrument Engineering*, 2022, no. 2 (139), pp. 18–25.
DOI: https://doi.org/10.18698/0236-3933-2022-2-18-25