A survey on public blockchain-based networks: structural differences and address clustering methods

Hye-yeong Shin
dep. of computer engineering
Keimyung University
Daegu, Republic of Korea
yeoung@stu.kmu.ac.kr

Meryam Essaid
dep. of computer engineering
Keimyung University
Daegu, Republic of Korea
maryama.essaid@kmu.ac.kr

Sejin Park
dep. of computer engineering
Keimyung University
Daegu, Republic of Korea
baksejin@kmu.ac.kr

Hongtaek Ju
dep. of computer engineering
Keimyung University
Daegu, Republic of Korea
juht@kmu.ac.kr

***Abstract*—Bitcoin is the most representative UTXO-based blockchain platform, and many studies have been conducted related to it. However, account-based blockchains such as Ethereum are not yet profoundly analyzed. There is an urgent need to track all cryptocurrency transactions involved with illegal activities to deanonymize and identify malicious users. To link users' accounts to real identities in both networks, we first need to examine the differences between Ethereum and Bitcoin to propose an efficient deanonymizing method. Therefore, this paper compares and analyzes the wallet address clustering method of Bitcoin and Ethereum.**

Keywords—Public Blockchain, Bitcoin, Ethereum, Wallet, account, deanonymizing users

I. Introduction

Blockchain anonymous protects the privacy of users. However, there is a problem that some malicious users can use cryptocurrencies for illegal activities by taking advantage of the anonymous aspect [2-4]. The Deanonymization of users is necessary to find out who is participating in the Dark Web. Unfortunately, it is impossible to deanonymize the user using only the information recorded in the Blockchain. Therefore it is necessary to collect additional data on the Dark Web where illegal trades(or transactions) are conducted. After collecting data related to illegal trades, the user can be tracked and deanonymized using some clustering methods.

Bitcoin is a representative blockchain platform [1]. In [5, 6, 34], a preliminary study was conducted to cluster addresses managed by the same Wallet using Bitcoin transaction data. Also, various studies [7-13] conducted user deanonymization using the methods presented in [5, 6] and data collected from SNS and Dark Web. Many studies on user deanonymization have been done in UTXO-based blockchain platforms [13-23] such as Bitcoin and Monero, and Zcash.

Deanonymization methods are needed not only in UTXO blockchain but also in account-based Blockchain. However, In the Account-based Blockchain like Ethereum, the method used to deanonymize users is quite different. Therefore, this paper examines how to account clustering in UTXO-based Blockchain before proceeding with research on deanonymizing Ethereum account users. After investigating research related to methods for account clustering, users' deanonymization can be performed using the methods used for clustering and the data collected from SNS and dark web. In this paper, we referred to the Ethereum account address as the wallet address to unify the terms of Bitcoin and Ethereum.

The structure of this paper is as follows. Sections 2 and 3 explain the basic concept of Bitcoin and Ethereum, respectively. Furthermore, section 4 describes the difference between Bitcoin and Ethereum. While, section 5 examines the method of clustering wallet addresses between Bitcoin and Ethereum, and section 6 compares and analyzes the clustering methods. Finally, section 7 concludes this paper and discusses future research directions.

II. Basic Concept Of Bitcoin

Bitcoin is the first cryptocurrency platform built using blockchain technology [1]. Bitcoin operates in P2P Network and consists of equal nodes. Each node validates the block containing the transaction(tx) according to the PoW consensus algorithm. The verified block is connected to the blockchain ledger, and the miner who created the block receives mining rewards.

A. Block

Bitcoin block records block header(version, Hash of previous Block header, Merkle root, Timestamp, Bits, Nonce), nTx, and tx ID list.

All newly mined blocks include the hash value of the previous block header, generating a chain of blocks linked back to the genesis block (first mined block and the root of the ledger). This value acts as a hash pointer to the previous block. By linking blocks back to the genesis block, all records in the ledger cannot be changed, even if the hash value of the header changes when the tx value in the block is forged.

The Merkle Tree(or Merkle Trie) is a data structure designed to efficiently explore whether a particular tx is included in the block. It contains information that summarizes the hash values and the Witness hash values of all txs included in the block. The top node of the Merkle tree is called the Merkle root, and this value is recorded in the block header. Txs can be verified quickly using Merkle Tree, and double-spending can be checked.

B. *Wallet*

Bitcoin Wallet manages private keys, public keys, and UTXO(Unspent Transaction Output). Bitcoin wallets are logical objects, creating and using private keys and public key pairs as needed.

The private key generates a public key with the ECC(Elli ptic curve cryptography) [25] . It generates a wallet address using the public key and SHA-256 hash algorithm. The wallet address created in this way can be used in Bitcoin txs. Moreover, UTXO from the wallet addresses generated from the same private key can be used in one tx.

Since the public key cannot be reversed with the public key hash, the UTXO managed by the corresponding Wallet cannot be used if the private key is lost. Therefore, the management of the Wallet's private key is critical.

C. *Transaction*

The tx is a record of the transfer of ownership of BTC recorded in UTXO. Typically, the sum of the values of UTXO managed by Bitcoin wallets is said to be the balance, but the concept of the balance does not exist in Bitcoin.

Since all txs generated in Bitcoin are made based on the contents recorded in UTXO, UTXO can be thought of as the basic unit of bitcoin txs. When someone wants to generate a tx, they can use one or more UTXOs managed by the Wallet.

III. Basic Concept Of Ethereum

Ethereum is a blockchain application platform that supports smart contract functions based on blockchain technology [24]. Ethereum allows cryptocurrency txs, and contracts can be made between parties that cannot be trusted. Ethereum operates on a P2P network and consists of nodes that generate and validate blocks according to the PoW consensus algorithm. The verified block is connected to the existing Blockchain, and the miner who created the block will receive ETC(Ether) as a mining reward.

A. *Block*

Ethereum block records block header, Uncle Block Hash List, and tx ID list. The block header consists of a total of 15 components.

B. *Account*

Accounts are the executor and fundamental unit of all txs on the Ethereum platform. All accounts are given addresses as unduplicated identifiers, which are called account addresses or wallet addresses.

Information such as the balance of an account, the number of txs, and the account type are called states. The overall state of Ethereum means the status information of all accounts present in Ethereum and is stored and managed as a Merkle Patricia tree.

There are two types of accounts in Ethereum.

a) EOA(Externally Owned Account): EOA is an account used by general users of the Ethereum platform. EOA is managed as a private key, can be used to generate digitally signed txs.

b) CA(Contract Account): CA is an account that is created when a smart contract is distributed on the Blockchain and acts as a pointer to the contract. Even if there are contracts that perform the same operation, CA has its uniqueness.

C. *Transaction*

In Ethereum, some txs are digitally signed with EOA's private key and Contract creation tx that are created when smart contracts are distributed on the Ethereum network. However, this paper focuses on explaining only the txs generated by EOA.

Txs created by EOA are payment txs and invocation txs. The payment tx is to transfer ETH from one EOA to another EOA, and the invocation tx is that the EOA calls/executes a specific function of the smart contract.

The payment/invocation tx generated by EOA is encoded with RLP(Recursive Length Prefix) and converted to a hash value through the cryptographic hash algorithm Keccak-256. The hash value is digitally signed with EOA's private key through the ECDSA algorithm [27].

The Nonce which is in tx data is the number of txs created by the sending wallet address and is not duplicated and increases by 1 in sequence. In Ethereum, the double-spending problem is solved with the nonce value of the tx.

IV. Comparison Of Bitcoin And Ethereum

Blockchain platforms can be compared based on the characteristics of Bitcoin, and Ethereum as discussed in sections 2 and 3. This section compares the fundamental concepts of Bitcoin and Ethereum. Bitcoin and Ethereum have something in common as they are both cryptocurrency platforms created based on blockchain technology. However, unlike Bitcoin, Ethereum supports smart contracts to provide more advanced Blockchain-based services to users.

A. *Wallet address*

Bitcoin is based on UTXO, and there is no concept of balance. Therefore, it is impossible to intuitively know the sum of UTXOs held in the Bitcoin wallet address. If someone wants to send a BTC, one or more input addresses and output addresses can be included in the generated tx. However, the input address must be wallet address(es) managed with the same private key(s).

On the other hand, Ethereum account-based txs are performed, and wallet address information are stored as state information. Status information allows us to intuitively check the balance of the wallet address and other information. In Ethereum, if someone wants to send ETH to another wallet address, only one input address and output address can be included in the tx. However, in this case, since the reuse rate of the wallet address increases, so the anonymity of users may not be preserved.

Of course, multiple inputs or output addresses can be written through smart contracts. In general, it is possible to overuse the same address.

B. *Double spending attack*

The two blockchain platforms also have different ways to solve the double-spending attack.

In Bitcoin, the user who made the tx creates a digital signature. The person who transmitted the tx verifies the digital signature. For tx verification, ScriptSig of the created tx and ScriptPubKey of UTXO are used.

The full node searches UTXO information in the UTXO set to verify UTXO. If the corresponding UTXO information is not found, this is considered a double-spending. When tx information is found, it goes through a verification process and waits to be included in the block. Furthermore, if the tx is included in the block, the corresponding UTXO information is deleted from the UTXO set.

Ethereum solves the double-spending attack by using the nonce value of the tx. The nonce value is the number of txs that occurred in the sending wallet address that created the tx. In other words, the value of Nonce is the statistical value of txs sent from the sending address and it can be found using getTransactionCount. If the difference between the result value of getTransactionCount and the nonce value of the current tx is 1, the tx is valid. However, suppose the difference between the values has a value of 0 or negative. In that case, the tx is considered as a double-spending. In Ethereum, it can be seen that txs are processed sequentially according to the value of the Nonce.

V. Wallet Address Clustering Study And Method

This section points out some related work regarding the wallet addresses clustering in Bitcoin and Ethereum, focusing on studies using txs data to cluster addresses.

A. Bitcoin

Reid, F. et al.[5] have analyzed the data of the bitcoin tx inputs and proceeded to cluster wallet addresses by collecting historical tx data from the bitcoin ledger. In this study, the authors assumed that the used tx's input addresses are managed with the same private key. Furthermore, [5] mentioned the limitation of user anonymity and argued that it is necessary to secure it.

Ron, D. et al. [6] have analyzed all tx activities of specific addresses that are assumed to be managed in the same Wallet using the method presented in [5], by describing the statistical characteristics of txs in the Bitcoin network until 2012, when the experiment was conducted. They have also extracted wallet addresses that have never sent BTC to other addresses and the total UTXO (7,019,100BTC) from those addresses. Also, the results of the wallet address that generated the most txs were presented. [6] It is challenging to track only the tx flow of a tx without using a specific method.

The method presented in [5, 6] is a fundamental research method on Bitcoin wallet address clustering. Hye-young et al. [10] conducted a study to analyze the financial activities of Satoshi Nakamoto using the methods suggested in [5, 6]. They have analyzed txal data in blocks 0 to 653,000 estimated that Satoshi had at least 1,011 wallet addresses and 20,143.438 BTC.

While in [12], authors have scrapped bitcoin addresses disclosed on social media, web, etc., and analyzed txs occurring at those addresses through BitIodine, a wallet address clustering tool developed by Spagnuolo et al. [33]. BitIodine is a wallet address clustering tool developed by adding the change wallet address to the method used in [5, 6]. [12] proceeded to cluster wallet addresses through BitIodine, and then proceeded to deanonymize the collected bitcoin address information.

In addition to the method used in [5, 6, 12], many other studies were also conducted to extract the characteristics of tx data related to the wallet address. Kanemura, K et al. [8] have proposed a method of analyzing wallet addresses and txs used in illegal txs and identifying DNM (DarkNet Market) addresses. In [8], authors have collected more than 200,000 bitcoin wallet addresses by web crawling and extracted 73 features through tx analysis related to the collected wallet addresses. In order to identify/classify the Bitcoin wallet address with the extracted features, data was trained using a supervised classifier. In addition, a voting-based system was proposed to classify wallet addresses through learning results. [8] confirmed that DNM addresses paid higher fees than other users and found that this plays an essential role in identifying DNM addresses andnon-DNM addresses. [8] confirmed the classification accuracy equivalent to 81% through the proposed Majoriyu voting-based voting method. The coin-mixing service was mentioned as the reason for 19% of classification inaccuracy.

B. Ethereum

Klusman et al. [29] have attempted to apply all of the de-anonymization techniques in Bitcoin [12, 32, 33] to research Ethereum de-anonymization. [32] conducted the study under the premise that the node that first propagated the tx was the person who created the tx. It attempted to connect to all nodes in the Bitcoin network and conducted a study to deanonymize the Bitcoin client by linking the IP address and the wallet address. In order to proceed with the research method proposed in [32], fixed node information is required. However, in Ethereum, research cannot be conducted similarly because the neighboring nodes connected to the node change periodically. The method proposed in [32] is quite challenging since it consumes many resources in order to keep connections established with all nodes. In other words, it was found that it was not possible to conduct because the P2P operation method and resource consumption were high.

In addition, as in [12], an attempt was made to analyze the tx by collecting address information on SNS and the web, but it was not possible to proceed due to the difference in the tx structure. The study proposed in [29] confirmed that the de-anonymization technology applied in Bitcoin is not applied to Ethereum due to the difference in the P2P operation method and tx composition.

Linoy et al. [30] have conducted a study on deanonymizing a smart contract's users and CA address using the Stylemetry method. The study is used to identify smart contracts related to illegal txs on Ethereum. However, no studies related to EOA were included.

Béres et al. [31] have collected wallet address information recorded in ENS (Ethereum Name Service), Ethereum wallet address, and user information recorded in Twitter and Darknet. [31] conducted graph representation learning based on Ethereum's payment/call tx data. In addition, the wallet address features were extracted from the daily activities and tx fees of the wallet address. [31] studied a method of clustering wallet addresses owned by the same user based on the graph representation learning result and the extracted data.

VI. Compare and analyze the method for Wallet

In this section, we compare and analyze the research methods investigated in section 5.

In Bitcoin, many studies on the clustering of wallet addresses have been conducted based on data stored in txs. It was possible to deanonymize users of addresses clustered with data collected from the web and tx data [8]. However, unlike bitcoin, Ethereum has difficulty clustering wallet addresses based on data recorded in accounts and txs. Because in the case of bitcoin, wallet addresses managed by the same private key can be listed in the vin of the tx. Contrary to Bitcoin, in Ethereum, it is difficult to proceed in the same way because only one sending address can be written in one tx. For this reason, for wallet address clustering in Ethereum, a new method was proposed in [31] to analyze and extract the features of wallet addresses and tx data.

VII. Conclusion And Future Work

De-anonymous studies on blockchain users are very important for tracking users who used cryptocurrency for illegal transactions. In UTXO-based blockchains such as Bitcoin, there have been many studies on user deanonymization based on address clustering that used the transaction data recorded in the ledger. In addition, by extracting features from wallet address clustering data and data collected from the darknet and SNS, users could be tracked and deanonymized.

However, Only a few attempts tried to cluster address and deanonymize used in account-based blockchains such as Ethereum. We can confirm that it is challenging to cluster wallet addresses using only the data recorded in the transaction through the fundamental characteristics of Ethereum [27]. Future research intends to cluster wallet addresses by extracting and analyzing each Ethereum wallet address's features and conducting users deanonymization study by collecting data related to illegal trades.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07050380) and Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No.2021-0-00484, Core Technologies for Hybrid P2P Network-based Blockchain Services).

References

[1] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system," 2008.

[2] Foley, S., Karlsen, J. R., Putniņš, T. J. "Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?," The Review of Financial Studies, Vol. 32, No. 5, pp. 1798-1853, May-April, 2019.

[3] Paquet-Clouston, M., Haslhofer, B., Dupont, B. "Ransomware payments in the bitcoin ecosystem," Journal of Cybersecurity, Vol. 5, No. 1, 2019.

[4] N. Christin, "Traveling the silk road: A measurement analysis of a large anonymous online marketplace," in Proceedings of ACM International Conference on World Wide Web, pp. 213–224, 2013.

[5] Reid, F., Harrigan, M. "An analysis of anonymity in the bitcoin system," Security and privacy in social networks, Springer, pp. 197-223, 2012.

[6] Ron, D., Shamir, A. "Quantitative analysis of the full bitcoin transaction graph," International Conference on Financial Cryptography and Data Security, Springer, pp. 6-24, April, 2013.

[7] Meiklejohn, S., Pomarole, M., Jordan, G., et al. "A fistful of bitcoins: characterizing payments among men with no names," In Proceedings of the 2013 conference on Internet measurement conference, pp. 127–140, October, 2013.

[8] Kanemura, K., Toyoda, K., Ohtsuki, T. "Identification of darknet markets' bitcoin addresses by voting per-address classification results," 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), IEEE, pp. 154-158, May, 2019.

[9] Sejin Jeong, Nohyun Kwak, Brent Byunghoon Kang, "A Study of Bitcoin Transaction Tracking Method through Illegal Community," Journal of The Korea Institute of Information Security & Cryptology, Vol. 28, No. 3, pp. 717-727, 2018.

[10] Hye-yeong Shin, Meryam Essaid, Hongtaek Ju, "Estimating wallets and bitcoins owned by Satoshi using Hash-rate and Transactions analysis", Asia-Pacific Network Operations and Management Symposium (APNOMS 2020), Sep. 2020.

[11] Hye-yeong Shin, Hongtaek Ju, "A study on Transaction Tracking and Analysis through Bitcoin Transaction Data," master's thesis, Keimyung University, Daegu, 2021.

[12] Al Jawaheri, H., Al Sabah, M., Boshmaf, Y., et al., "Deanonymizing tor hidden service users through bitcoin transactions analysis," Computers & Security, 89, 101684, 2020

[13] Biryukov, A., Tikhomirov, S., "Security and privacy of mobile wallet users in bitcoin, dash, monero, and zcash," Pervasive and Mobile Computing, 59:101030, 2019.

[14] Alonso, Kurt M., "Zero to monero," 2020.

[15] Monero, https://www.getmonero.org/, cited 2021 March. 27.

[16] Chervinkski, Joao Otávio Massari, Kreutz, D., "Floodxmr: Low-cost transaction flooding attack with monero's bulletproof protocol," IACR Cryptology ePrint Archive, 2019:455, 2019.

[17] Möser, M., Soska, K., Heilman, E., et al. "An empirical analysis of traceability in the monero blockchain," arXiv preprint arXiv:1704.04299, 2017.

[18] Li, Y., Yang, G., Susilo, W., et al. "Traceable monero: Anonymous cryptocurrency with enhanced accountability," IEEE Transactions on Dependable and Secure Computing, 2019.

[19] Zcash, https://z.cash/technology/, cited 2021 March. 27.

[20] Biryukov, A., Feher, D., "Deanonymization of hidden transactions in zcash," University of Luxembourg, 2018.

[21] Biryukov, A., Feher, D., Vitto, G., "Privacy aspects and subliminal channels in zcash," In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, pp. 1813-1830, 2019.

[22] Tramer, F., Boneh, D., Paterson, K. G., "Ping and reject: The impact of side-channels on zcash privacy," 2019.

[23] Zhang, Z., Li, W., Liu, H., et al. "A refined analysis of zcash anonymity," IEEE Access 8: 31845-31853, 2020.

[24] Buterin, V., "Ethereum: A next-generation smart contract and decentralized application platform," URL https://github.com/ethereum/wiki/wiki/% 5BEnglish% 5D-White-Paper, 7, 2014.

[25] "SEC2 ver2", http://www.secg.org/sec2-v2.pdf, cited 2021 March. 29.

[26] "Nakamoto Satoshi's Bitcoin Wallet address", https://www.blockchain.com/btc/address/1A1zP1eP5QGefi2DMPTfTL5SLmv7DivfNa, cited 2021 Marh. 31.

[27] "ECDSA Algorithms", http://bit.ly/2r0HhGB, cited 2021 Marh. 31.

[28] Sompolinsky, Y., Zohar, A., "Accelerating bitcoin's transaction processing," fast money grows on trees, not chains, 2013. URL https://eprint.iacr.org/, 2013.

[29] Klusman, R., Dijkhuizen, T., "Deanonymisation in ethereum using existing methods for bitcoin," 2018.

[30] Linoy, S., Stakhanova, N., Matyukhina, A., "Exploring Ethereum's Blockchain Anonymity Using Smart Contract Code Attribution," In 2019 15th International Conference on Network and Service Management (CNSM), IEEE, pp. 1-9, 2019

[31] Béres, F., Seres, I. A., Benczúr, A., et al. "Blockchain is Watching You: Profiling and Deanonymizing Ethereum Users," arXiv preprint arXiv:2005.14051, 2020.

[32] Biryukov, A., Khovratovich, D., Pustogarov, I., "Deanonymisation of clients in Bitcoin P2P network," Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, pp. 15-29, 2014.

[33] Spagnuolo, M., Maggi, F., Zanero, S. "Bitiodine: Extracting intelligence from the bitcoin network," International conference on financial cryptography and data security. pp. 457-468, Springer, Berlin, Heidelberg, 2014.

[34] Essaid, Meryam, et al. "Mapping Out Bitcoin's Pseudonymous actors, " 2020 International Conference on Information Networking (ICOIN). IEEE, pp. 802-806, 2020.