2023 8th International Conference on Information and Network Technologies (ICINT) | 979-8-3503-0145-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICINT58947.2023.00019

An EOA Identity Tracing System (AITS) on Ethereum Blockchain

Sawanya Rattanabunno, Warodom Werapun, Jakapan Suaboot, Tanakorn Karode, Maneenate Puongmanee
College of Computing
Prince of Songkla University
Kathu, Phuket 83120, Thailand
{s6430621003, warodom.w, jakapan.su, s6230622001, maneenate.p}@phuket.psu.ac.th

***Abstract*—Ethereum, one of the most popular cryptocurrencies, allows for anonymous transactions and is frequently used for money laundering and scams. Although advanced scammers are hard to trace as they use sophisticated coin-mixing techniques, we argue that many generic scams only involve amateurs who manually mix transactions to avoid police detection. To counter this, centralized exchanges require users to go through Know-Your-Customer (KYC) processes before exchanging tokens for fiat currency. This paper extends the anti-money laundering further by proposing *An EOA Identity Tracing System (AITS)*, which traces the flow of crypto tokens from the thief wallet to the exchange and backtracking from the KYC's identities to the thief's real identity. The proposed AITS also aids investigators with the token-transferring graph that is useful for off-chain investigation. The experimental results on the 1,045 thieves' transactions recorded over 290 days reveal behaviors that the scammers used to evade police detection.**

Keywords—identity trace, blockchain, romance, scam, Ethereum

I. Introduction

Bitcoin [1] is a decentralized digital currency that has gained popularity over the last decade. It allows for direct transactions between users without a middleman, such as banks or brokers. Bitcoin has two defining features: transparency, as all transactions are recorded publicly on a decentralized ledger called the blockchain; and anonymity, which refers to the ability of users to keep their identity private while participating in a transaction. The anonymity aspect of Bitcoin comes from the fact that users are identified by pseudonyms, also known as addresses, rather than by their real-world identities [2]. The addresses are generated from users' public keys and do not include any personal information. Cybercriminals use this anonymous property of a cryptocurrency to hide their identities when committing illegal actions. In fact, cryptocurrency is widely used in money laundering [3][4].

In 2015, Ethereum was launched with the smart contract functionality that allows programmers to create accounts and tokens. The accounts are categorized into (i) externally owned accounts (EOA), which is a type of account that is controlled by private keys; and (ii) contract account, which uses the smart contract code to control. Meanwhile, Ethereum also supports many types of tokens, some are cryptocurrencies, while others are not for use as a means of payment. Unfortunately, some famous tokens such as ETH, BNB, and USDT, are tempting to scammers [5][6] as they can exploit the anonymity property for malicious purposes. Although transactions of scammers are known, it is difficult or sometimes impossible to reveal the identities of scammers.

In order to thwart misused behaviors, most centralized exchanges, including Binance[1], require users to establish a customer's identity using Know-Your-Customer (KYC)[2] processes before swapping any token to fiat currency, i.e., a currency that is issued and backed by the government. However, the sophisticated scammer could still mix many transactions involving non-KYC wallets before cashing out.

Fig. 1 depicts a process of tracing a scammer's real identity. Say, the scammer manipulated the victim to get his tokens either by hacking or using social engineering tricks. The process begins with the address of the victim wallet ①, which was used to transfer tokens to the thief wallet ②. In step ③, the scammer used tokens mixing technique involving transactions of several non-KYC wallets to evade the investigation. After that, ④, many KYC wallets will be used to cash out the fiat currency from the exchange ⑤. In step ⑥, the investigator collects all real users' identities from the KYC wallets. A step ⑦ is the most difficult part. Here, an investigator needs to gather information from various data sources (i.e., both on-chain and off-chain). If enough evidence is obtained, the real identity of the thief can be revealed in step ⑧.

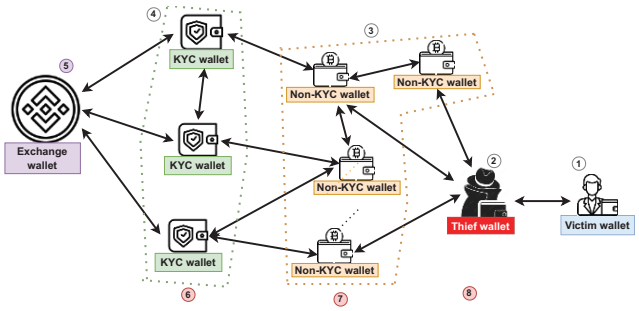

Fig. 1. A process of tracing attacker's identity.

In fact, aggregation of off-chain information, such as IP addresses, blog posts, and CCTV footage, serves as complementary data to on-chain information, particularly in the context of tracing the origin of a scammer through backtracking from KYC wallets to his/her address. Chainalysis [7] devises a tool to investigate a thief's identity by gathering both on-chain and off-chain data, including insider data from government and police intelligence. Unfortunately, this product

1. https://www.binance.com/en
2. https://en.wikipedia.org/wiki/Know_your_customer

979-8-3503-0145-8/23/$31.00 ©2023 IEEE
DOI 10.1109/ICINT58947.2023.00019

is commercial, and with the privacy data protection act (PDPA), all algorithms and information used are confidential. Additionally, the price of the full-feature Chainalysis platform is expensive. Hence, not many regions have enough funds for this platform.

In reality, a huge number of scams involve amateur scammers. Take a romance scam case for an example, a thief uses the romance scam [8][9] strategy to trick the victim's trust. Then, the thief asks for a small amount of investment and pays back with real profits. When the victim trusts the thief either because of love or a small amount of profit returned, the victim will invest a large amount of money. The thief then uses cryptocurrency for money transferring, including peer-to-peer trading. After the thief gets enough money, he/she disconnects from the victim and is gone forever. Most of the time, the victim knows only the crypto wallet address of the scammer. In fact, many victims have very limited knowledge about cryptocurrency, hence it is a popular tool for scammers. As a result, for such cases, the thief wallet address is the only evidence the police officer receives from the victim.

Therefore, in this paper, we propose an approach to help the police officer identifies the real identity of the thief by tracing from the thief wallet address (i.e., EOA) that committed the scam. Our proposed work helps the investigator to visualize the flow of the crypto tokens from the thief's wallet to the exchange and backtrack from the KYC's identities to the thief's real identity. Contributions of this work are summarized as follows: (i) we propose a system to find addresses of wallets between the thief and the exchange wallet, which are useful for tracing the thief's real identity; and (ii) we implement software to visualize thief and exchange wallets relationship and investigate token transferring behaviors of thieves over the Ethereum platform.

This paper is organized as follows. Section 2 presents the background and related work. Section 3 describes our proposed system. The experiment result is illustrated in Section 4. Eventually, we conclude our work in Section 5.

II. Background and Related Work

A. Cryptocurrencies

Cryptocurrencies are a form of digital currencies that utilize cryptography and blockchain to function. Blockchain is a decentralized ledger that records all transactions. Bitcoin [1] is the first digital currency that operates without central authority. It was invented by Satoshi Nakamoto in 2008. It is considered a decentralized system among other digital currencies. There are several benefits to using cryptocurrency as a form of payment or investment. Many cryptocurrencies offer great benefits such as decentralization, security, borderless, lower transaction costs, immutable, accessibility, and anonymity. One interesting feature of cryptocurrency is privacy, specifically, users can create wallets without revealing their identities. This feature benefits a user who has a privacy constraint, e.g., donating a large amount of money. However, thieves (i.e., scammers) take advantage of the privacy to ask their victims to send them money using cryptocurrency wallets, that leads to blockchain crime [10].

B. Money laundering Using Cryptocurrencies

The dark web [11] is an online marketplace that sells illegal products and services, as well as information obtained from illegal activities. The dark web is accessed through special browsers, and it plays a significant role in the cybercrime world. Bitcoin is a commonly used cryptocurrency in the dark web marketplaces because it offers a high degree of flexibility and anonymity. It allows users to conceal both their activities and identities. K. Soska, and N. Christin [12] found that the black market on the dark web is growing, with a long-term analysis of 35 markets over two years detailing the expansion of this market. Their study also estimates that the black market on the dark web will generate $100M USD annually and reach a new revenue record by 2021 with a total of $2.1 billion in cryptocurrencies. Due to the anonymity property, cryptocurrencies can indeed be used as a tool to do money laundering, for instance, Bitcoin mixing services [15][16]. Even though most exchanges impose the KYC process, malicious users can still mix a number of transactions with many unverified parties. Tornado.cash [13] offers a mixing transaction service for privacy on the Ethereum blockchain. However, to prevent criminal activity, mixing several transactions such as Tornado.cash service is blacklisted by U.S. Department of the Treasury.

C. Related Work

Fortunately, due to the ban on the mixing or tumbling service, the scammer must manually mix his/her coins before cashing out. However, not many thieves are smart enough to create a complex mixing. In fact, some of them are not aware that the process of converting the cryptocurrency to fiat could reveal their real identities. Specifically, when they cash out their token to fiat with not many mixing transactions or leave other off-chain pieces of evidence. Hence, catching this type of thief is still possible, and the investigation cost is not too expensive.

Several researchers investigate anti-money laundering challenges on Bitcoin mixing or tumbling services, such as [14][15]. J. Seo, M. Park, H. Oh and K. Lee [16] suggest a method to detect the abnormal coin mixing event. Chainalysis experts [7] have followed the path of the money that was taken during the hack on an exchange. They have traced it to places where the stolen funds were converted to fiat or other digital currencies. D. Goldsmith, K. Grauer, and Y. Shmalo [17] look at six groups of Bitcoin transactions that are believed to be connected to two well-known hacking groups. Methods for using graphical [18] representations to examine blockchain transactions using a graph-based visualization technique. H. Kanezashi, T. Suzumura, X. Liu, and T. Hirofuchi [19] use a combination of different types of graph neural networks to identify fraudulent activities on the Ethereum blockchain.

P. Li, H. Xu, and T. Ma [20] propose an identity tracing scheme and evaluates it using a simple and efficient proof method. Their scheme is appropriate for blockchain systems that utilize public/private key pairs and would not disrupt the anonymity feature of the original system. However, their work is only a mathematically proven concept, and the real implementation is not available.

III. Proposed Solution

This section presents An EOA Identity Tracing System (AITS), which is designed and implemented to trace Ethereum blockchain transactions in order to identify the real thief's identity. Details of components and implementations are discussed below.

A. AITS Architecture

Fig. 2 illustrates the architecture of AITS. Basically, it consists of backend and frontend, API, Etherscan, and databases (i.e., SQL and graph). Programming APIs of each component are as follows:

(i). Backend and frontend: *Nestjs* is used for the backend-server to obtain blockchain information from *Etherscan*, whereas *Nextjs* is chosen for the web frontend to visualize graphical illustrations to users and to support REST API queries.

(ii). API: the proposed algorithm is implemented in the *block-tracer-api*. This component is responsible for querying data from *Etherscan*, caching the raw blockchain data in the SQL database, generating graph models, storing the graph models in the graph database.

(iii). Etherscan: Ethereum is a decentralized, open-source blockchain with smart contract functionality, used to request transactions from blockchain networks via Etherscan APIs with related wallet addresses from transaction history.

(iv). Databases: PostgreSQL is used for the SQL database, while *neo4j* is used for storing the generated graph models. Specifically, neo4j is used to pull graph data consisting of user wallet nodes and their respective relationships from the Etherscan.

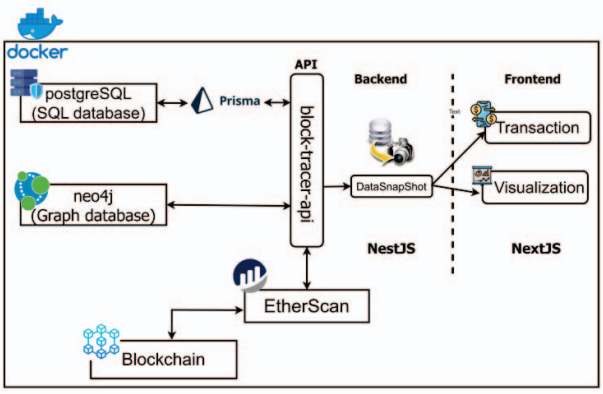

Fig. 2. AITS architecture.

B. Reaching Exchange DFS Algorithm

This section describes the core engine of AITS, named *Reaching Exchange DFS* algorithm. First, AITS queries transactions T_D that are related to the thief wallet address from the Ethereum blockchain network and stores them in the SQL database. Formally, $T_D \in \{\alpha, \beta_i\}$, where α denotes transactions that contain a thief wallet address, and β defines transactions that contain α or β_i. Here, $2 \leq i \leq D$ represents the number of hops from the thief wallet address, which limits at maximum D hops or reaching one of the exchange wallet addresses. To select all transactions T_D at maximum D hops from the thief wallet address, from Ethereum, we propose *Reaching Exchange DFS* algorithm as follows.

Algorithm 1: Reaching Exchange DFS (*RE_DFS*)
Input: *G*: The wallet graph stored in an adjacency list
Output: All wallet addresses inside *G* in the DFS order until reaching exchange wallets or the maximum depth *D*.

```
Function RE_DFS(wallet, search_level):
  if wallet.visit = true or
     wallet.address = exchange.address or
     search_level > maximum_depth then
              return;
  end
  print(wallet.address);
  wallet.visit <= true;
  for next_wallet ∈ G[wallet].neighbors() do
       RE_DFS(next_wallet, search_level+1);
  end
end
```

The output wallet graph *G* is stored in the *neo4j* graph database, which is useful for generating the graph illustration for investigators or other users. The transaction queries and visualization of AITS are implemented using *NextJS*. These web features will link to the backend API, called *block-tracer-api* to search in the graph database and SQL database. To make searching more efficient, AITS schedules *data-snapshot* by querying blockchain transactions via *Etherscan* APIs regarding an initial thief wallet to any wallet and following the next transactions until reaching one of the exchange wallets or the maximum searching depth level. Then, storing them in the databases.

Pulling transactions via *EtherScan* is faster than directly reading from a block in the blockchain or searching for a transaction relating to a wallet address because *EtherScan* aids in filtering transactions with an initial wallet (i.e., with a specific source address). However, a free account in *EtherScan* has a condition about API rate limit, we must cache some previous transaction searching before re-synchronize it again as a checkpoint. AITS uses *reaching exchange depth-first-search* (Algorithm 1) to traverse all destination wallets (i.e., graph nodes) until reaching exchange wallets or stopping at the maximum depth level.

All fetched transactions related to the thief wallet address are stored in a database (i.e., PostgreSQL). They are later put into the *neo4j* graph database, for the purpose of faster and simpler gathering all nodes between a thief wallet address and Binance (i.e., exchange).

After AITS gets the thief wallet address, called Thief. It fetches all related transactions, as shown in Fig 3. The Thief (initial wallet address) is linked to different wallets (i.e., 0x1999, 0x17c0f8, 0xd0046) until reaching the Binance wallet. AITS extracts all related wallet path that can be used to identify user identity starting from the exchange.

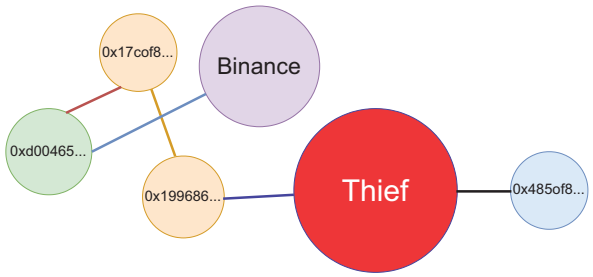

Fig. 3. AITS transfer graph visualization (Frontend).

IV. Evaluation and Experimental Results

This section discusses the experimental method used to evaluate our proposed approach and experimental results. Details are given below.

A. Experiment Settings

We gathered thief wallet addresses from the local regional police service, consisting of 10 wallet addresses. We put these wallets in the system and searched until the exchange wallet address was found in a transaction. We obtained all related centralized exchange wallets collected from *EtherScan* in AITS database. The maximum number of search hops was set to 4 to prevent the search results explosion issue, so-called NP-complete problem, as the memory consumption will grow exponentially.

Since this project was a collaboration with the local regional police service, the outcome from this research, i.e., wallet addresses that participated in the centralized exchanges, was given to the police. Hence, the police can contact the exchanges to acquire data on individuals' identities who cashed out from exchanges. However, this paper does not cover the process of the police investigation to trace back to the thief's real identity using information gathered from the off-chain.

B. Experimental Results

We analyzed related transactions gathered from AITS using 4 hops limit, and found 4 out of total 10 thieves, which were connected to an exchange. Fig 4 depicts behavior of the thieves transferring USDT during the period of 290 days.

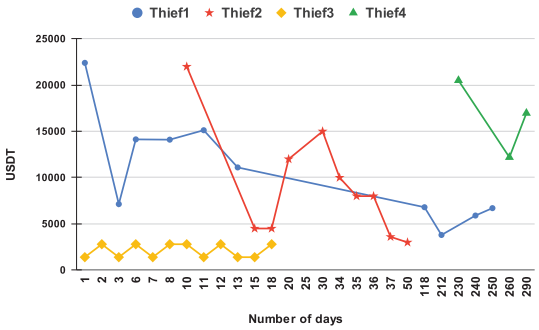

Fig. 4. The number of days thieves used to transfer the amount of USDT after the fraud.

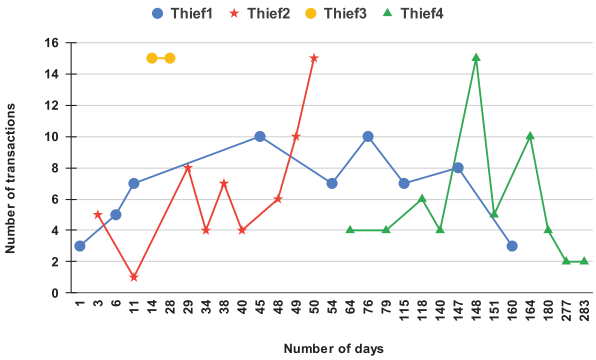

Fig. 5. The number of transactions in a day thieves transferred money.

We observed 3 different patterns of thieves transferring the stolen USDT. The first group was Thief1 and Thief2. They behaved similarly transferring large amounts of USDT at the beginning and decreasing the amount after that. The second group was Thief3, who transferred small USDT every day over a short period of around 3 weeks. The third group was Thief4, who transferred a large amount of USDT over a short period but postponed his/her activity for a very long period, around 8 months after the fraud. This result gave the police insight into the different types of behavior of the thieves.

Fig. 5 compares the number of transactions in a day the scammers transferred the USDT to an exchange. We sorted transactions of token amounts descending and selected the top 15 transactions for each thief wallet. Then we examined how many transactions were used to transfer in a day. Thief1 used 3 transactions (with 22,400 USDT, sent less transactions with high token amounts) on the first date. He/she took 160 days to distribute his/her tokens. On the other hand, Thief2 used 5 transactions to transfer 22,000 USDT. Thief3 used 15 transactions to transfer approximately 2,800 USDT. Thief4 used 4 transactions to transfer 20,505 USDT. Additionally, Thief2, Thief3, and Thief4 waited for 49, 14, and 148 days before doing 15 transactions, respectively. Although thieves took several days to move their assets to other wallets, they often transferred a high token amount at the beginning. The reason could be that some of them may be needing to get the money, or they simply were not aware of the police observation.

Fig 6 demonstrates the number of USDT and hops thieves used in our dataset. We took a sampling of 400 wallet addresses from a total of 1,045 transactions related to 4 thief wallet addresses that transfer tokens at 1, 2, 3, and 4 hops to see how many hops the scammers used to transfer most of their money. According to our observation, a total of 22,400 USDT was transferred using 4 hops, 14,000 USDT using 3 hops, and 7,100 USDT using 2 hops. There was no direct transfer observed. Intuitively, when thieves transferred a large amount of USDT, they tried to distribute it using more hops before reaching an exchange to evade detection from the police.

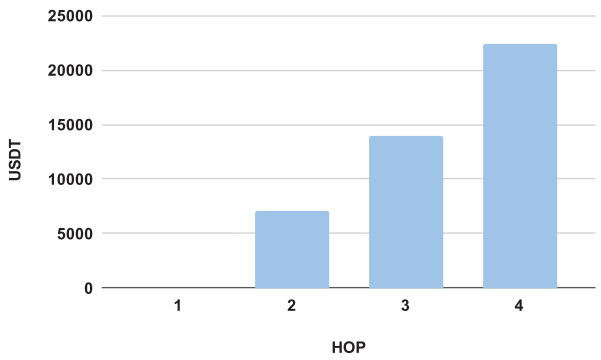

Fig. 6. The number of USDT and hops.

V. Conclusion

This paper presents a transaction path examination using tracing algorithms to obtain information between a thief wallet and exchanges wallets. Since most exchanges impose the KYC processes to prevent fraudulent activities, the KYC accounts can be used in this work to backtrack the real identity of the thief. The main contribution of this work is not only the proposed tracing system but also the discovery of the actual crypto-scammer behaviors. For instance, how the thieves tried to spread their USDT by transferring through multiple intermediaries before sending the funds to exchanges to avoid law enforcement monitoring. In future work, artificial intelligence technology will be examined to be used in analyzing data and identify patterns of behaviors to predict potential fraudulent transactions.

Acknowledgment

The authors would like to thank College of Computing, Prince of Songkla University for supporting research (Grant No. COC6304156S).

References

[1] S. Nakamoto, “A peer-to-peer electronic cash system,” Bitcoin.–URL: https://bitcoin. org/bitcoin. pdf, 2008.

[2] K. Toyoda, P. T. Mathiopoulos, and T. Ohtsuki, “A novel methodology for hyip operators’ Bitcoin addresses identification,” IEEE Access, vol. 7, pp. 74 835–74 848, 2019.

[3] C. Brenig, R. Accorsi, and G. Muller, “Economic analysis of cryptocurrency backed money laundering.” in ECIS, 2015.

[4] M. Moser, R. Bohme, and D. Breuker, “An inquiry into money laundering tools in the Bitcoin ecosystem,” in 2013 APWG eCrime Researchers Summit. IEEE, 2013, pp. 1–14.

[5] E. Badawi, G. -V. Jourdan, G. Bochmann and I. -V. Onut, "An automatic detection and analysis of the Bitcoin generator scam," 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Genoa, Italy, 2020, pp. 407-416, doi: 10.1109/EuroSPW51379.2020.00061.

[6] B. Hammi, S. Zeadally, Y. C. E. Adja, M. D. Giudice and J. Nebhen, "Blockchain-based solution for detecting and preventing fake check scams," in IEEE Transactions on Engineering Management, vol. 69, no. 6, pp. 3710-3725, Dec. 2022, doi: 10.1109/TEM.2021.3087112.

[7] Chainalysis, “Chainalysis reactor,” Online available: https://www.chainalysis.com/, Accessed: Dec 2021.

[8] G. Suarez-Tangil, M. Edwards, C. Peersman, G. Stringhini, A. Rashid and M. Whitty, "Automatically dismantling online dating fraud," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1128-1137, 2020, doi: 10.1109/TIFS.2019.2930479.

[9] S. Al-Rousan, A. Abuhussein, F. Alsubaei, O. Kahveci, H. Farra, and S. Shiva, "Social-guard: Detecting scammers in online dating", 2020 IEEE International Conference on Electro Information Technology (EIT), pp.416-422, 2020

[10] R. Phillips and H. Wilder, "Tracing cryptocurrency scams: Clustering replicated advance-fee and phishing websites," 2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Toronto, ON, Canada, 2020, pp. 1-8, doi: 10.1109/ICBC48266.2020.9169433.

[11] K. Shubhdeep and S. Randhawa. “Dark Web: A web of crimes.” Wireless Personal Communications, 112 (2020): 2131-2158.

[12] K. Soska, and N. Christin, Measuring the longitudinal evolution of the online anonymous marketplace ecosystem, Proceedings of the 24th USENIX Security Symposium, 2015, 33–48.

[13] R. Semenov, A. Pertsev, and R. Storm, Tornado cash, Dec 2019, Online available: https://en.wikipedia.org/wiki/ Tornado_Cash

[14] S. Mabunda, "Cryptocurrency: The new face of cyber money laundering," 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 2018, pp. 1-6, doi: 10.1109/ICABCD.2018.8465467.

[15] J. Crawford and Y. Guan, "Knowing your Bitcoin customer: Money laundering in the Bitcoin economy," 2020 13th International Conference on Systematic Approaches to Digital Forensic Engineering (SADFE), New York, NY, USA, 2020, pp. 38-45, doi: 10.1109/SADFE51007.2020.00013.

[16] J. Seo, M. Park, H. Oh and K. Lee, "Money laundering in the Bitcoin network: Perspective of mixing services," 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2018, pp. 1403-1405, doi: 10.1109/ICTC.2018.8539548.

[17] D. Goldsmith, K. Grauer, and Y. Shmalo, Analyzing hack subnetworks in the Bitcoin transaction graph. Appl Netw Sci 5, 22 (2020). https://doi.org/10.1007/s41109-020-00261-7

[18] J. S. Tharani, E. Y. A. Charles, Z. Hóu, M. Palaniswami and V. Muthukkumarasamy, "Graph based visualisation techniques for analysis of blockchain transactions," 2021 IEEE 46th Conference on Local Computer Networks (LCN), Edmonton, AB, Canada, 2021, pp. 427-430, doi: 10.1109/LCN52139.2021.9524878.

[19] H. Kanezashi, T. Suzumura, X. Liu, and T. Hirofuchi, Ethereum fraud detection with heterogeneous Graph Neural Networks, abs/2203.12363, arXiv, CoRR, https://doi.org/10.48550/arXiv.2203.12363, 2022

[20] P. Li, H. Xu, and T. Ma, An efficient identity tracing scheme for blockchain-based systems, Information Sciences, Volume 561, 2021, pp. 130-140, 0020-0255, https://doi.org/10.1016/j.ins.2021.01.081.