Journal of Cyber Security
DOI: 10.32604/jcs.2022.038401
Article

Phishing Scam Detection on Ethereum via Mining Trading Information

Yanyu Chen[1] **and Zhangjie Fu**[1,2,*]

[1]Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, China
[2]The State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, Shaanxi, China
*Corresponding Author: Zhangjie Fu. Email: wwwfzj@126.com
Received: 01 July 2022; Accepted: 30 September 2022

Abstract: As a typical representative of web 2.0, Ethereum has significantly boosted the development of blockchain finance. However, due to the anonymity and financial attributes of Ethereum, the number of phishing scams is increasing rapidly and causing massive losses, which poses a serious threat to blockchain financial security. Phishing scam address identification enables to detect phishing scam addresses and alerts users to reduce losses. However, there are three primary challenges in phishing scam address recognition task: 1) the lack of publicly available large datasets of phishing scam address transactions; 2) the use of multi-order transaction information requires a large number of queries and computations; and 3) the extraction of phishing scam address features relies on machine learning methods excessively, which leads to the loss of practical meaning and is harmful to the research of phishing scam addresses. This paper proposes a systematic phishing scam address recognition scheme, to simultaneously overcome the three challenges in phishing scam address recognition. In this paper, a systematic phishing scam address recognition scheme is proposed to addresses these issues. Specifically, due to the insufficient number of address tagged in the existing publicly available Ethereum phishing scam address transaction dataset, we first construct a transaction dataset involving over 10000 tagged addresses. To the best of our knowledge, this is the largest dataset of tagged addresses for Ethereum phishing scam detection. Then, we design a new heuristic rule to implement feature extraction of address nodes by analysing the traditional financial involved accounts combined with information specific to Ethernet transactions. After that, a novel adaptive feature importance filtering method is designed to adaptively adjust the filtering threshold based on the final classification results, which reduce the feature dimensionality while ensuring a certain detection performance. Finally, random forest is used to classify whether the addresses is a phishing scam address or not. Extensive experiments on real Ethereum datasets show that our approach (98.89% Precision, 98.35% Recall, 98.62% F1) achieves state-of-the-art performance.

Keywords: Blockchain; identity inference; Ethereum; node classification

1 Introduction

The era of digital cryptocurrencies began with the publication of Satoshi Nakamoto's paper entitled "Bitcoin: A Peer-to-Peer Electronic Cash System" [1]. Since then, cryptocurrencies have begun to flourish and have even become legal tender in El Salvador [2]. According to coinmarketcap.com, there are now over 20,000 cryptocurrencies with a market capitalisation of almost a trillion dollars. These cryptocurrencies are generated using blockchain technology. In general, blockchain can be thought of as a secure and reliable tamper-evident ledger achieved through a consensus mechanism [3]. It is because of its tamper-evident nature that blockchain technology is widely used in areas such as transaction processing [4], drug tracking [5] and credit management systems [6].

But Ethereum [7] and Bitcoin [1], the most widely known representatives of blockchain technology applications, have long been present in illegal activities because of their anonymity and financial attributes. A secure, trustworthy and legal blockchain environment is an important foundation for the healthy development of blockchain technology and countries are also paying more and more attention to blockchain regulation [8].

Phishing scams have received a lot of attention due to their widespread existence, the large number of victims and the huge amount of money lost. Phishing scam identification has become an important research topic in the blockchain ecosystem as a technique to detect phishing scams and maintain blockchain security [9], and has attracted a lot of attention from researchers. According to a report by Chainalysis [10], phishing scams were once again the most prevalent form of cryptocurrency crime in 2021, taking over $7.7 billion worth of cryptocurrency, with the Ponzi scheme Finiko netting over $1.1 billion. These show that detecting and preventing phishing scams is already a pressing issue in the blockchain ecosystem.

Traditional phishing scams obtain fiat currency by constructing fake platforms, obtaining passwords, telecommunication scams, etc. [11–14]. Then, they transfer the money obtained through illegal means such as underground money changers for profit. Existing phishing scams on the blockchain obtain cryptocurrencies by constructing fake platforms to obtain private keys, promising high returns, etc. Then, they transfer the fraudulent money directly to profit by means of coin mixing, etc. [10].

Although the means of acquisition are different and the target of acquisition is different, both are essentially the transfer of funds to the account. Therefore, we hope to draw on the traditional financial research on illegal account fund transfer to achieve the detection of phishing scams on the blockchain.

Since Ethereum is one of the most popular blockchain platforms, with 14.8 transactions per second and 700,000 active addresses per day [15], we use Ethereum as an example to demonstrate the effectiveness of our approach. Specifically, we construct a transaction dataset with over 10,000 tagged addresses, of which 3,634 are phishing scam address tags. Then, based on the features derived from the analysis of traditional financial fraud accounts and Ethereum trading information, we designed and constructed the features of each node and used an automatic feature culling method to remove unimportant features while ensuring the stability of the metrics. Finally, extensive experiments were conducted on real-world datasets to validate the model in comparison with other methods, to assess the performance of the model under different parameters, and to discuss the effectiveness of these features.

In summary, we have made the following key contributions:

- We constructed a transaction dataset with over 10,000 tagged addresses, of which 3,634 were phishing scam address tags.

- We designed 149 node features based on traditional financial account research, combined with the features of Ethereum transactions and used a designed feature filtering scheme to remove unimportant features while ensuring the stability of the indicators.
- We designed one that requires only first-order transactions of the addresses in question to implement phishing scam address detection, and validated its effectiveness on real Ethereum data. The method outperforms existing methods in all performance metrics (98.89% Precision, 98.35% Recall, 98.62% F1) and can be part of a user's cryptocurrency wallet functionality as a function to alert the user of potential risks when interacting with unfamiliar accounts.

The remainder of this paper is as follows. Section 2 describes the background of the research and related work. Section 3 describes the overall design. Section 4 describes the experiments and Section 5 describes the conclusion of this paper.

2 Background and Related Work

Blockchain technology is the underlying core technology of cryptocurrencies [16] such as Bitcoin and Ethereum. Blockchain technology is a comprehensive technology that combines cryptography, distributed storage, cryptography and other technologies. The architecture generally consists of a data layer, a network layer, a consensus layer, an incentive layer, a contract layer and an application layer [17]. The development of blockchain technology is now generally considered to have reached the third generation [18]. Blockchain 1.0 is represented by Bitcoin, whose basic function is to achieve a de-neutralised digital currency. Blockchain 2.0 is represented by Ethereum, whose key feature is the addition of smart contract functionality, which enables the writing of automatically running programs on the blockchain. Blockchain 3.0 is still in development and lacks a unified consensus. While there are various descriptions of blockchain technology, a blockchain is essentially an immutable distributed ledger. It has become widely known in recent years as the value of cryptocurrencies has risen and fallen wildly. While cryptocurrencies are only a small part of the applications of blockchain technology, its immutable nature gives it a much wider scope for development. In recent years, more and more researchers have taken advantage of its properties to use it in areas such as drug storage and transportation [19], financial transaction ledgers [20], and identity and copyright authentication [21]. Governments are also investigating it for its immense practical value [22–24].

With the boom in blockchain technology, the security of the blockchain ecosystem has also received a lot of attention. Among these, phishing scams have received significant attention from researchers due to their widespread existence and the huge losses they cause. One of the earliest studies of phishing scams in the blockchain was conducted by Vasek et al. [25] studied phishing scams in bitcoin in 2015, where they found 192 phishing scams and determined that at least 13,000 users had been defrauded out of over $11 million. Since then, more and more researchers have been working on phishing scams in the blockchain and on ways to detect phishing scams in the blockchain.

Existing methods for detecting phishing scams in Ethereum are mainly divided into two types. The former focuses mainly on the study of the behaviour of the phishing scam address itself for the design of features. Chen et al. [26] extracted features from the transaction amount and transaction time information of the second-order inner neighbours of the target node, designed to extract 219-dimensional features, and then used an integrated algorithm based on lightGBM for classification. The latter mainly uses graph embedding to obtain structural information. Wu et al. [27] designed the trans2vec algorithm, designed biased sampling based on transaction amount and timestamp for the sampling process, and then used SVM for classification.

However, existing works do not go far enough into the study of phishing scams on Ethereum and do not make sufficient use of information about conducting Ethereum transactions. In addition, researchers have placed more emphasis on multi-order transaction information acquisition, with greater demand for data, and have not fully mined first-order neighbourhood information.

3 Methods

The aim of our research is to detect phishing scam address on Ethereum. In this section we present our overall solution, divided into the following sections: 1) the completely new dataset we constructed, 2) the phishing scam detection task we definited, 3) the feature extraction method we designed, 4) the classifier we chose. The overall structure is shown in Fig. 1.

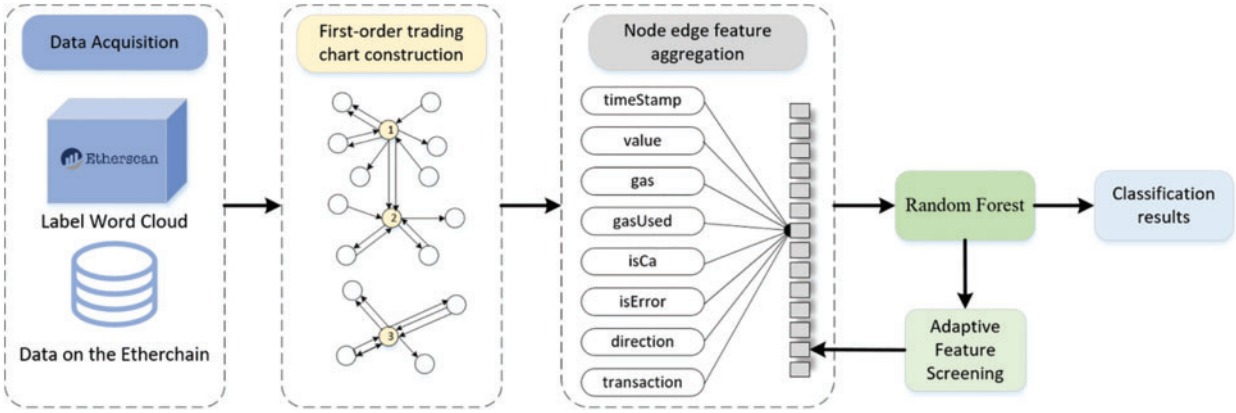

Figure 1: The overall structure

3.1 Dataset

We obtained over 10,000 tagged addresses from the Label Word Cloud at Etherscan, of which 3,634 were phishing scam address tags. To the best of our knowledge, this is the largest dataset of tagged addresses for Ethereum phishing scam detection. Although our current research on phishing scams is still focused on first-order neighbours, given that existing schemes largely require the use of second-order transactions, we build the dataset for all marked addresses second-order transactions. We hope to contribute a new public dataset for phishing scam address detection. At the same time, when we researched information on traditional illegal activities involving bank accounts, we found that the relevant data is generally not publicly available due to security and privacy issues. However, the phishing scam address detection dataset we provide does not have this problem as all transaction data is in the public blockchain environment, so we hope to be able to provide data to support research into the behaviour of traditional financial bank accounts.

3.2 Problem Definition

In this paper, because the Ethereum transaction network is naturally graph relational, we treat Ethereum addresses as nodes and transactions as edges. The Ethereum phishing scam address detection problem is considered as a node classification problem. Let directed graph $G = (V, E, X, Y, C)$, where V denotes the set of nodes, E denotes the set of edges, X denotes the edge attributes, Y denotes the addresses labels corresponding to the nodes, and C denotes if a node is a contract addresses. In G, $X \in R^{|E| \times S}$ where S is the size of the information available in the transaction. Each transaction

contains 5 available messages, namely timestamp (trading timestamp), value (transaction amount), gas (gas limit), gasUsed (final consumption of gas) and isError (whether the transaction was successful). $Y \in R^{|V|}$ contains 3 labels, 1 for a phishing scam address, 0 for a non-phishing scam address, and 2 for an unknown address. $C \in R^{|V|}$ contains 2 possibilities, 1 for a CA (Contact Account) and 0 for an EOA (Externally Owned Accounts).

The main objective of this work is to design feature fusion for nodes by analyzing the user features exhibited by illegal activities in traditional financial bank accounts in conjunction with the features of Ethereum transactions. The node fusion features $X_E \in R^{|V| \times d}$, where d is the dimensionality of the feature representation, are designed to characterize the behaviour of phishing scam addresses on Ethereum while using only the first-order transactions of the addresses in question. At the same time, the feature adaptive filtering module is used to remove unimportant features and to ensure that the metrics are stable. The obtained node features are finally applied to the downstream Ethereum phishing scam address classification task to achieve the detection of phishing scam addresses.

3.3 Feature Extraction and Filtering

We first looked at the information that each Ethereum transaction record can provide. In addition to the five pieces of information that can be obtained directly from the transaction, the output address can be obtained by querying the Ethernet on-chain data to see if the output address is a contract address (out_ca). Of these six pieces of information, timeStamp and value are commonly used in existing phishing scam address detection, while gas, gasUsed, out_ca, and isError are pieces of information that existing detection methods ignore. We consider timeStamp and value to be a direct match to the transaction time and amount in the transaction history of a traditional financial bank account. Gas and gasUsed are related to the complexity of the operation, the user's usage habits and the platform used, which matches the different scenarios that may arise from the use of different banks in traditional financial activities. We use out_ca to determine the different categories of transfer out target accounts, which matches the different account identities in traditional financial activities. We use isError to distinguish between successful and unsuccessful transactions that interfere with various different data.

We designed features based on the following transaction features exhibited by illegal activities in traditional financial bank accounts [28–30]: 1) decentralised transfer of funds in and out or centralised transfer in and out; 2) complex flow of funds; 3) concealment of unusual transaction methods; 4) diverse forms and large number of accounts involved. For characteristic 1, we designed features such as the in_degree, out_degree, the ratio of in_degree to total degree, and the ratio of out_degree to total degree to assess the transfer in and out. For features 2 and 3, we designed features such as maximum, minimum and average of gas and gasused to assess the complexity of the money flow operation and the different trading methods. For characteristic 4, we count the number of different output identities and their maximum, minimum and difference values in different situations to assess the variety of account forms. In the feature design, we considered comparison of validity with features extracted from existing manual feature methods using the information timeStamp and value, and ultimately constructed 149 dimensional features per address, as detailed in Appendix A.

We designed the automatic feature culling method to analyse the validity of each feature in the implementation and remove invalid features to achieve dimensionality reduction while evaluating the validity of previously neglected trading information *vs.* that used in the original manual feature method. We use the Gini_index of each feature to determine its importance, calculated as follows:

$$Gini\,(D) = 1 - \sum_{k=1}^{|y|} p_k^2 \tag{1}$$

$$Gini\_index\,(D, a) = \sum_{v=1}^{V} \frac{|D^v|}{D} Gini\,(D^v) \tag{2}$$

where D is the overall dataset, p_k is the probability of occurrence of the category k, a is a feature in the sample, D^v is the sub-dataset delineated according to the feature a, and the smaller the *Gini_index* of a means a higher degree of internal certainty of D^v. We rank the features according to their importance and retain the top $Q \times q$ features of the remaining features Q at a time until there is a drop in performance metrics over p, where q, p are hyperparameters, q represents the feature sieve factor and p represents the performance threshold.

3.4 CLassifier

Classifiers are used to perform downstream classification tasks giving results while also providing important metrics for feature selection. Many well-established and effective classification algorithms are available, such as Graph Convolutional Network (GCN), Graph Attention Networks (GAT), Support Vector Machine (SVM), Logistic Regression (LR) and Random Forest (RF). In the phishing scam detection problem, we found RF to be more effective. So we choose RF, which uses decision trees as the basic unit for classification by integrating a large number of decision trees, as our classifier. RF uses the idea of Bagging:

(1) Multiple, replayed partial training samples are removed from the training set to become a new training set.
(2) Using the new training set, multiple sub-models are trained.
(3) Using a voting method, the classification category of the sub-model with the most votes is determined as the final category.

We use decision trees for combined judgement of design features to achieve detection of phishing scam addresses.

4 Experiments

4.1 Method Comparison

To verify that our proposed method is more suitable for phishing scam detection tasks, we compare the existing performance advanced phishing scam models trans2vec and DElightGBM. We also compare with GCN, GAT, SVM, LR and other methods. The results are shown in Table 1.

Table 1: Method comparison

	Precision	Recall	F1
trans2vec	0.87	0.89	0.88
DElightGBM	0.9735	0.9601	0.9668
Our-GCN	0.9023	0.9815	0.9402
Our-GAT	0.9181	0.9730	0.9447
Our-SVM	0.9693	0.8973	0.9319
Our-LR	0.9661	0.9447	0.9553
Our-RF	0.9889	0.9835	0.9862

Since we performed the detection of phishing scam addresses, we believe that higher recall is more important. It can be seen that among these detection methods, trans2vec and SVM perform poorly, LR, DElightGBM and GAT have some performance, while our proposed Random Forest-based one exceeds 98% in all metrics. This implies that the proposed scheme is fully capable of practical application.

4.2 Feature Screening and Results Analysis

To ensure that the performance metrics of the method are stable while removing unimportant features, we make q = 0.9. We keep the top 90% of features continuously for each round where the difference between the metrics does not exceed p. For p taking 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, we conducted the following experiments to verify the performance of the scheme. The specific results are shown in Table 2.

Table 2: Effect of taking different values of p on feature screening for q = 0.9

	Precision	Recall	F1	The number of the features
$p = 0.001$	0.9889	0.9835	0.9862	149
$p = 0.002$	**0.9876**	**0.9848**	**0.9862**	**120**
$p = 0.003$	0.9876	0.9848	0.9862	120
$p = 0.004$	**0.9876**	**0.9848**	**0.9862**	**120**
$p = 0.005$	0.9863	0.9890	0.9876	28
$p = 0.006$	**0.9806**	**0.9766**	**0.9786**	**11**
$p = 0.007$	0.9806	0.9766	0.9786	11
$p = 0.008$	**0.9806**	**0.9766**	**0.9786**	**11**

As can be seen from Table 2, when p is taken from 0.001 to 0.004, it is not possible to filter the features effectively, limited by the high performance requirements. When p is relaxed to 0.005, the number of features decreases significantly and all performance indicators remain good. When p is greater than 0.006, the screening again remains smooth.

We have dedicated our analysis to the 8 most important features. The importance of these 8 features is specified in Fig. 2.

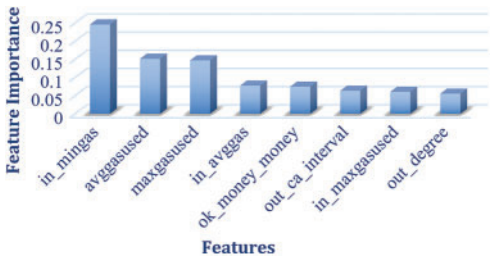

Figure 2: Feature importance

As shown in Fig. 2, we can see that the Ethereum transaction information gas and gasused contributed significantly to the classification results. The four most important features for the

classification results are all related to gas and gasused. This is a good indication that our mining of Ethereum transaction information is meaningful and successful.

5 Conclusion

In this work, our goal is to implement the detection of phishing scam addresses in Etherscan. First, we collect labels of phishing scam and non-phishing scam addresses and their transactions from etherscan, constructing a dataset of transactions with the highest number of tagged addresses so far. Then, we design features based on the following transaction features exhibited by illegal activities in traditional financial bank accounts. Finally, we use the adaptive feature censors to ensure that the performance metrics of the method are stable while implementing the removal of unimportant features, and evaluate each feature by feature importance ranking.

Extensive experiments have shown that our method achieves state-of-the-art performance, achieving high detection performance for low-dimensional features. In the future, we will progress to analyse the specific meaning of the features exhibited by phishing scam addresses. In addition, the data will be made public at a later stage in order to facilitate research in the field of phishing scam address detection.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Decentralized Business Review, 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf.

[2] E. Vázquez, "The technical fix: Bitcoin in El Salvador," *South Atlantic Quarterly*, vol. 121, no. 3, pp. 600–611, 2022.

[3] D. Yaga, P. Mell, N. Roby and K. Scarfone, "Blockchain technology overview," arXiv preprint arXiv:1906.11078, 2019.

[4] A. Adiyanto and R. Febrianto, "Authentication of transaction process in E-marketplace based on blockchain? Technology," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 2, no. 1, pp. 68–74, 2020.

[5] L. Bell, W. J. Buchanan, J. Cameron and O. Lo, "Applications of blockchain within healthcare," *Blockchain in Healthcare*, vol. 1, no. 2018, pp. 1–7, 2018.

[6] M. J. Ashley and M. S. Johnson, "Establishing a secure, transparent, and autonomous blockchain of custody for renewable energy credits and carbon credits," *IEEE Engineering Management Review*, vol. 46, no. 4, pp. 100–102, 2018.

[7] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, vol. 151, no. 2014, pp. 1–32, 2014.

[8] Y. Guo and C. Liang, "Blockchain application and outlook in the banking industry," *Financial Innovation*, vol. 2, no. 1, pp. 1–12, 2016.

[9] Z. H. Yuan, Q. Yuan and J. J. Wu, "Phishing detection on ethereum via learning representation of transaction subgraphs," in *Proc. the Int. Conf. on Blockchain and Trustworthy Systems*, Singapore, pp. 178–191, 2020.

[10] Chainalysis. The Chainalysis 2022 Geography of Cryptocurrency Report, 2022. https://go.chainalysis.com/geography-of-crypto-2022-report.html.

[11] B. Parmar, "Protecting against spear-phishing," *Computer Fraud & Security*, vol. 2012, no. 1, pp. 8–11, 2012.

[12] S. Kumar, A. Faizan, A. Viinikainen and T. Hamalainen, "Mlspd-machine learning based spam and phishing detection," in *Proc. the Int. Conf. on Computational Social Networks*, Shanghai, China, pp. 510–522, 2018.
[13] K. Dunham, "Phishing isn't so sophisticated: Scary!" *Information Security Journal*, vol. 13, no. 2, pp. 2–7, 2004.
[14] C. M. Kelley, K. W. Hong, C. B. Mayhorn and E. Murphy-Hill, "Something smells phishy: Exploring definitions, consequences, and reactions to phishing," in *Proc. the Human Factors and Ergonomics Society Annual Meeting*, Los Angeles, USA, pp. 2108–2112, 2012.
[15] H. Chen, M. Pendleton, L. Njilla and S. Xu, "A survey on ethereum systems security: Vulnerabilities, attacks, and defenses," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–43, 2020.
[16] A. Nasir, K. Shaukat, K. I. Khan, I. A. Hameed, T. M. Alam *et al.,* "What is core and what future holds for blockchain technologies and cryptocurrencies: A bibliometric analysis," *IEEE Access*, vol. 9, pp. 989–1004, 2020.
[17] X. Han, Y. Yuan and F. Y. Wang, "Security problems on blockchain: The state of the art and future trends," *Acta Automatica Sinica*, vol. 45, no. 1, pp. 206–225, 2019.
[18] A. AL-Ashmori, S. Basri, P. D. D. Dominic, A. Muneer, Q. AI-Tashi *et al.,* "Blockchain-oriented software development issues: A literature review," *Software Engineering Application in Informatics*, vol. 232, pp. 48–57, 2021.
[19] M. Uddin, "Blockchain medledger: Hyperledger fabric enabled drug traceability system for counterfeit drugs in pharmaceutical industry," *International Journal of Pharmaceutics*, vol. 597, pp. 120235, 2021.
[20] B. A. Tama, B. J. Kweka, Y. Park and K. -H. Rhee, "A critical review of blockchain and its current applications," in *Proc. the 2017 Int. Conf. on Electrical Engineering and Computer Science (ICECOS)*, Indonesia, Palembang, pp. 109–113, 2017.
[21] Y. Li, J. Wei, J. Yuan, Q. Xu and C. He, "A decentralized music copyright operation management system based on blockchain technology," *Procedia Computer Science*, vol. 187, pp. 458–463, 2021.
[22] S. Ølnes, J. Ubacht and M. Janssen, "Blockchain in government: Benefits and implications of distributed ledger technology for information sharing," *Government Information Quarterly*, vol. 34, no. 3, pp. 355–364, 2017.
[23] L. Carter and J. Ubacht, "Blockchain applications in government," in *Proc. the 19th Annual Int. Conf. on Digital Government Research: Governance in the Data age*, New York, NY, USA, pp. 1–2, 2018.
[24] C. Piao, Y. Hao, J. Yan and X. Jiang, "Privacy preserving in blockchain-based government data sharing: A service-on-chain (SOC) approach," *Information Processing & Management*, vol. 58, no. 5, pp. 102651, 2021.
[25] M. Vasek and T. Moore, "There's no free lunch, even using Bitcoin: Tracking the popularity and profits of virtual currency scams," in *Proc. the Int. Conf. on Financial Cryptography and Data Security*, San Juan, Puerto Rico, pp. 44–61, 2015.
[26] W. Chen, X. Guo, Z. Chen, Z. Zheng and Y. Lu, "Phishing scam detection on ethereum: Towards financial security for blockchain ecosystem," in *Proc. the Twenty-Ninth Int. Joint Conf. on Artificial Intelligence*, Yokohama, Japan, pp. 4506–4512, 2020.
[27] J. J. Wu, Q. Yuan, D., Lin, W. You, W. L. Chen *et al.,* "Who are the phishers? Phishing scam detection on ethereum via network embedding," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 2, pp. 1156–1166, 2022.
[28] W. Wei, "An inquiry into the means of fund transfer in, and symbiosis grey industry of, telecommunication and network fraud crimes," *Policing Studies*, vol. 8, pp. 78–88, 2022.
[29] J. H. Zhong, "Control deficiencies and countermeasures of bank institutions' case-involved accounts from the perspective of anti-money laundering-an empirical analysis based on 1326 accounts involved in telecommunication fraud in shaoyang city," *Financial Accounting*, vol. 11, pp. 63–67, 2021.
[30] Y. Y. Zhang and J. Wang, "Research on the practice, problems and countermeasures in combating the 'capital chain' of telecommunication fraud," *Financial Accounting*, vol. 9, pp. 29–33, 2022.

Appendix A.

In order to distinguish between successful and unsuccessful transactions, inflow and outflow transactions , we have constructed 24 × 4 features. We use the example of the outflow transaction in the successful transaction (24):

out_degree: total outflow of transactions in the successful transaction.

out_money: total amount of outflow transactions in the successful transaction.

out_maxmoney: maximum amount of outflow transactions in the successful transaction.

out_minmoney: minimum amount of outflow transactions in the successful transaction.

out_interval_money: out_maxmoney-out_minmoney.

out_money_degree: $\frac{\text{out\_money}}{\text{out\_degree}}$.

out_begin: earliest timestamp of outflow transactions in the successful transaction.

out_stop: last timestamp of outflow transactions in the successful transaction.

out_interval: out_stop-out_begin.

out_money_interval: $\frac{\text{out\_money}}{\text{out\_interval}}$.

out_interval_degree: $\frac{\text{out\_interval}}{\text{out\_degree}}$.

out_avggas: average gas of outflow transactions in the successful transaction.

out_maxgas: maximum gas of outflow transactions in the successful transaction.

out_mingas: minimum gas of outflow transactions in the successful transaction.

out_avggasused: average gasused of outflow transactions in the successful transaction.

out_maxgasused: maximum gasused of outflow transactions in the successful transaction.

out_mingasused: minimum gasuesd of outflow transactions in the successful transaction.

out_intervalgas: out_maxgas-out_mingas.

out_intervalgasused: out_maxgasused-out_mingasused.

out_neighbour: total number of neighbours of outflow transactions in the successful transaction.

out_avgneighbour: $\frac{\text{out\_degree}}{\text{out\_neighbour}}$.

out_maxneighbour: maximum number of transactions for a single neighbour of outflow transactions in the successful transaction.

out_minneighbour: minimum number of transactions for a single neighbour of outflow transactions in the successful transaction.

out_intervalneighbour: out_maxneighbour-out_minneighbour.

In order to distinguish between successful and unsuccessful transactions for the transaction output account identity, we have constructed 5 × 2 features. We use the example of the outflow transaction in the successful transaction (5):

out_ca: total number of output transactions to ca in the successful transaction.

out_eoa: out_degree-out_ca.

out_ca_interval: out_eoa-out_ca.

out_ca_out_degree: $\frac{\text{out\_ca}}{\text{out\_degree}}$.

out_eoa_out_degree: $\frac{\text{out\_eoa}}{\text{out\_degree}}$.

In order to assess the overall situation of the nodes, we construct the following 43-dimensional features:

degree: total number of transactions.

ok_degree: total number of transactions in the successful transaction.

error_degree: total number of transactions in the unsuccessful transaction.

ok_degree_degree: $\frac{\text{ok\_degree}}{\text{degree}}$.

error_degree_degree: $\frac{\text{error\_degree}}{\text{degree}}$.

in_degree_degree: $\frac{\text{total inflow of transactions in the successful transaction}}{\text{degree}}$.

out_degree_degree: $\frac{\text{out\_degree}}{\text{degree}}$.

in_error_degree_degree: $\frac{\text{total inflow of transactions in the unsuccessful transaction}}{\text{degree}}$.

out_error_degree_degree: $\frac{\text{total outflow of transactions in the unsuccessful transaction}}{\text{degree}}$.

ok_money: total amount of transactions in the successful transaction.

ok_money_degree: $\frac{\text{ok\_money}}{\text{degree}}$.

error_money: total amount of transactions in the unsuccessful transaction.

error_money_degree: $\frac{\text{error\_money}}{\text{degree}}$.

money: total amount of transactions.

money_degree: $\frac{\text{money}}{\text{degree}}$.

ok_money_money: $\frac{\text{ok\_money}}{\text{money}}$.

error_money_money: $\frac{\text{error\_money}}{\text{money}}$.

ok_maxmoney: maximum amount of transactions in the successful transaction.

error_maxmoney: maximum amount of transactions in the unsuccessful transaction.

maxmoney: maximum amount of transactions.

ok_minmoney: minimum amount of transactions in the successful transaction.

error_minmoney: minimum amount of transactions in the unsuccessful transaction.

minmoney: minimum amount of transactions.

balance: total amount of inflow transactions in the successful transaction–total amount of outflow transactions in the successful transaction.

interval: timestamp interval in the successful transaction.

error_interval: timestamp interval in the unsuccessful transaction.

ok_money_interval: $\frac{\text{ok\_money}}{\text{interval}}$.

interval_degree: $\frac{\text{interval}}{\text{degree}}$.

error_interval_degree: $\frac{\text{error\_interval}}{\text{degree}}$.

mingas: minimum gas in the successful transaction.

maxgas: maxmum gas in the successful transaction.

avgas: average gas in the successful transaction.

intervalgas: maxgas-mingas.

mingasused: minimum gasused in the successful transaction.

maxgasused: maximum gasused in the successful transaction.

avggasused: average gasused in the successful transaction.

intervalgasused: maxgasused-mingasused.

minneighbour: minimum number of transactions for a single neighbour in the successful transaction.

maxneighbour: maximum number of transactions for a single neighbour in the successful transaction.

avgneighbour: average number of transactions for a single neighbour in the successful transaction.

intervalneighbour: maxneighbour-minneighbour.

num_neighbour: total number of neighbours in the successful transaction.

ca: whether a contract address or not.