



Article

A Collective Anomaly Detection Technique to Detect Crypto Wallet Frauds on Bitcoin Network

Mohammad Javad Shayegan ^{1,*}, Hamid Reza Sabor ¹, Mueen Uddin ²  and Chin-Ling Chen ^{3,4,5,*} 

¹ Department of Computer Engineering, University of Science and Culture, Bahar St., Shahid Qamushi St., Ashrafi Esfahani Bulvar, Tehran 1461968151, Iran; e-mueen.uddin@ubd.edu.bn

² School of Digital Science, University Brunei Darussalam, Gadong BE1410, Brunei; mueen.uddin@ubd.edu.bn

³ School of Information Engineering, Changchun Sci-Tech University, Changchun 130600, China

⁴ School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China

⁵ Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung 41349, Taiwan

* Correspondence: shayegan@usc.ac.ir (M.J.S.); clc@cyut.edu.tw (C.-L.C.)

Abstract: The popularity and remarkable attractiveness of cryptocurrencies, especially Bitcoin, absorb countless enthusiasts every day. Although Blockchain technology prevents fraudulent behavior, it cannot detect fraud on its own. There are always unimaginable ways to commit fraud, and the need to use anomaly detection methods to identify abnormal and fraudulent behaviors has become a necessity. The main purpose of this study is to use the Blockchain technology of symmetry and asymmetry in computer and engineering science to present a new method for detecting anomalies in Bitcoin with more appropriate efficiency. In this study, a collective anomaly approach was used. Instead of detecting the anomaly of individual addresses and wallets, the anomaly of users was examined. In addition to using the collective anomaly detection method, the trimmed_Kmeans algorithm was used for clustering. The results of this study show the anomalies are more visible among users who had multiple wallets. The proposed method revealed 14 users who had committed fraud, including 26 addresses in 9 cases, whereas previous works detected a maximum of 7 addresses in 5 cases of fraud. The suggested approach, in addition to reducing the processing overhead for extracting features, detect more abnormal users and anomaly behavior.

Keywords: K_means; trimmed_kmeans; Blockchain; Bitcoin; anomaly detection



Citation: Shayegan, M.J.; Sabor, H.R.; Uddin, M.; Chen, C.-L. A Collective Anomaly Detection Technique to Detect Crypto Wallet Frauds on Bitcoin Network. *Symmetry* **2022**, *14*, 328. <https://doi.org/10.3390/sym14020328>

Academic Editors: Sergei D. Odintsov and Alexander Shelupanov

Received: 29 November 2021

Accepted: 1 February 2022

Published: 5 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Blockchain was first proposed in 1991 to establish an encryption and information exchange system to address data security concerns [1]. Bitcoin, as the first electronic cryptocurrency was emerged from the Blockchain features by Satoshi Nakamoto in 2008 [2] and attracted the attention of governments around the world to use Bitcoin. Attractiveness and the amazing popularity of Bitcoin as a cryptocurrency have made Blockchain so popular. Blockchain has gained many enthusiasts in industry and academia and attracted the attention of many applications such as the Internet of Things [3]. Due to this popularity, many cybercriminals and even real-world criminals (because of the anonymity of users) became interested in using Blockchain and Bitcoin [4].

However, blockchains are not without drawbacks and limitations and is not completely immune to fraud, hack, attacks, and other malicious activities. The blockchain itself suffers from security issues. The security issues could be categorized into three levels, namely, the process level, the data level, and the infrastructure level. There exist many studies on how to incorporate different Blockchain technologies to enhance the security, transparency, and traceability of systems [5].

Bitcoin users are always at risk of being hacked, and in addition to the enormous economic losses it causes to these users, it can also cause credit crises for commercial

websites [6–9]. Due to this technology’s novelty, the developed security mechanisms for some systems do not yet exist, and there have been several hack attacks on digital currencies [6]. Although Blockchain technology prevents fraudulent behavior, it cannot detect fraud on its own, so new innovative techniques and methods are needed to track attacks [10]. The amazing attractiveness of Bitcoin, on the one hand, and the rise of cybercrime activity, on the other, have made it imperative to use anomaly detection for identifying potential scams.

One of the most important techniques for handling security issues is using anomaly detection. In data mining, anomaly detection is the identification of rare items, events, or observations that raise suspicions by differing significantly from the majority of the data. A collective anomaly refers to a group of data points that differ from the majority of the data, wherein a single data point is not treated as an anomaly [11]. Although Blockchain technology prevents fraudulent behavior due to the type of structure, technology, and use of consensus algorithms, it cannot detect fraud on its own, and there may always be unpredictable ways to steal and defraud [10]. Thus, new innovative techniques and methods are invented for handling attacks on Blockchain. The amazing attractiveness of Bitcoin, on the one hand, and the rise of cybercrime activity, on the other, have made it imperative to use anomaly detection to detect potential fraud.

This study intends to use collective anomaly detection (instead of point anomaly detection) on all one user’s wallets (instead of individual wallets) to remove features that have higher computational and operational capabilities. This approach reduces data size and helps to identify better abnormalities that have been intentionally used with multiple user wallets.

2. Previous Works

Much research has been carried out on cryptocurrency. For example, due to the attractiveness of the cryptocurrency, many studies have been carried out on its financial aspects, such as [12–15]. The main focus of the current study is on the anomaly in cryptocurrency and their architecture, namely Blockchain.

Initially, Blockchain was thought to be resistant to all kinds of attacks due to its cryptographic type and thanks to consensus algorithms, but security issues have prompted researchers to look for ways to detect anomalies in the blockchains. Several studies tackled the anomaly detection issue in Blockchain [4,7–9,16–33].

Table 1 summarizes the types of malicious attacks on blockchains and the tactics and potential strategies that can be used to confront them. As shown in Table 1, anomaly detection methods can be used to detect the most malicious attacks. For example, using anomaly detection methods, bitcoin accounts of users who have used combinational services to engage in illegal activities or money laundering can be detected and tracked.

Table 1. Summary of the types of malicious attacks on Blockchain and potential strategies confront them (extracted from [10]).

Malicious Attack	Definition	Defensive and Preventive Measures
Double Spending	An individual makes more than one payment using one body of funds.	The complexity of the mining process
Record Hacking	fraudulent transactions are inserted into the ledger.	Distributed consensus; Detection techniques
51% Attack	A single miner node with more computational resources (51%) than the rest of the network nodes dominates the verification and approval of transactions.	Detection techniques; wide adoption of the Blockchain technology
Identity Theft	The private key of an individual is stolen.	Identify and reputation blockchains
Illegal Activities	Parties transact illegal goods or commit money laundering.	Detection techniques; laws and regulations
System Hacking	The programming codes and systems that implement Blockchain are compromised.	Robust systems and advanced intrusion detection methods

According to Table 1, one of the diagnostic anomalies in the blockchains is countering the Record Hacking attack and detecting theft, hacking, fraud, which this study and the following works have considered:

Zambre et al. [9] used six features and the K-Means algorithm to identify suspicious and rogue users and found a starting point for analyzing suspicious users. Pham et al. [7] used three main social networking methods (power degree and densification laws, K-Means clustering, and local outlier factor) to diagnose anomalies. They were able to discover one of the 30 known cases of theft. In a subsequent study, Pham et al. [21] used three unsupervised learning methods, including K-Means clustering, Mahalanobis distance, and unsupervised vector machine (SVM), and were able to identify a total of 3 of the 30 known cases.

Monamo et al. [4] emphasize that anomaly detection plays an important role in data mining and considering that many remote locations have important information for further investigation, and in the Bitcoin network, diagnostics anomaly means detecting fraud, used the trimmed-Kmeans method. They successfully identified five of the addresses involved in 30 cases of theft, hacking, fraud, or loss.

Monamo et al. used the kd-trees algorithm instead of the Trimmed-Kmeans algorithm in the next study [20] and were able to discover 7 of the target addresses, which were involved in 5 out of 30 cases of theft, hacking, fraud, or loss. (In many cases of theft, hacking, or fraud, thieves participated in multiple addresses and wallets to make it difficult to detect anomalies, leaving multiple addresses and wallets in each theft.)

In a study, Signorini et al. [22] suggested using Fork instead of eliminating it to diagnose abnormalities. Chawathe [16] further analyzes the method of Monamo et al. [4] and recommends this method to detect anomalies in the blockchains. In addition to the method and algorithm used in the previous records, the subject of feature selection is also very important, which is compared in Table 2. The ✓ character shows a column method that supports the feature in that row.

According to the results obtained in previous records, several works have been conducted, and in most of these methods, improvements in anomaly detection have been achieved by changing or adding new features or changing the algorithm, but in all methods, only the anomaly detection of wallet addresses has been sought. Additionally, if the user has multiple wallets and the behavior of each of these addresses seems normal, the previous methods will be somewhat inefficient. Since abnormal users mainly use multiple wallet addresses to normalize their behavior, it can be more efficient to choose a method that can examine the user's behavior instead of the wallet address. In order to solve this problem, in addition to using the best features and algorithms in the previous records, the method of collective anomaly detection has been used to pay more attention to the anomaly detection of users with several wallet addresses.

Table 2. Comparison of feature selection in previous records.

Feature Selection	Zambre et al. [9]	Pham et al. [21]	Pham et al. [7] in the Subsequent Study	Monamo et al. [4]	Monamo et al. [20] in the Subsequent Study	Chawathe [16]
In-degree		✓	✓	✓	✓	✓
Out-degree		✓	✓	✓	✓	✓
Average amount incoming				✓	✓	✓
Average amount outgoing				✓	✓	✓
Average time interval between transactions	✓	✓	✓			

Table 2. Cont.

Feature Selection	Zambre et al. [9]	Pham et al. [21]	Pham et al. [7] in the Subsequent Study	Monamo et al. [4]	Monamo et al. [20] in the Subsequent Study	Chawathe [16]
Average time interval between our transactions	✓	✓	✓			
Clustering coefficient			✓	✓	✓	✓
Average incoming speed	✓	✓	✓			
Average outgoing speed	✓	✓	✓			
In-acceleration	✓					
Out-acceleration	✓					
Unique in-degree			✓			
Unique out-degree			✓			
Balance			✓			
Creation date			✓			
Active duration			✓			
In-degree transaction		✓				
out-degree transaction		✓				
The total value of the transaction		✓				
Triangle				✓	✓	✓
Total amount sent				✓	✓	✓
Total amount received				✓	✓	✓
Standard deviation received				✓	✓	✓
Standard deviation sent				✓	✓	✓
In-in				✓	✓	✓
In-out				✓	✓	✓
Out-in				✓	✓	✓
Out-out				✓	✓	✓

3. Research Method

As mentioned in this research, the process of anomaly detection has been carried out with a collective anomaly approach. The details of the proposed method are described below:

3.1. Dataset and Theft List

In this research, the dataset of the “ELTE Bitcoin Project” [34] has been used. This database includes the entire blockchain related to Bitcoin until 9 February 2016 and its basic version includes transactions until 28 December 2013. The basic database includes seven files: Block specification, transaction ID, Bitcoin Addresses, Block ID, Transaction output list, and Transaction input list. Each file has several features. Given that in previous works up to 7 April 2013, the Bitcoin Blockchain database had been examined, this study has also used these two datasets to date to examine the results more closely with previous works. The list of addresses that have committed theft, fraud, hacking, or loss was then extracted.

The following section provides a brief description of the features in Table 2:

- In-degree: Number of transactions received by a given user.
- Out-degree: Number of transactions sent by a given user.
- Unique in-degree: Number of unique users a given user has received transactions.
- Unique out-degree: Number of unique users a given user has sent transactions.
- Average in-transaction: Average number of bitcoins received per incoming transaction.
- Average out-transaction: Average number of bitcoins sent per outgoing transaction.
- Average time interval between in-transactions.
- Average time interval between out-transactions.
- Number of public keys owned by a given user.
- Balance: Net number of bitcoins retained by the user.
- Clustering coefficient: the measure of connectivity amongst neighbors of a given user.
- Creation date: timestamp of the first transaction associated with a given user.
- Active duration: time difference between first and most recent transactions associated with a given user.
- Balance: Net number of bitcoins for a given transaction considering all in- and outgoing edges from that transaction.
- Clustering coefficient: the measure of connectivity amongst neighbors of a given transaction.
- Currency features: total amount sent, the total amount received, average amount sent, the average amount received, standard deviation received, standard deviation.
- Creation date: timestamp of the first edge associated with a given transaction.
- Active duration: time difference between first and most recent edges associated with a given transaction.
- Network/graph features: in-degree, out-degree, clustering coefficient, number of triangles.
- Average neighborhood (source–target) whereby concerning each query node: source refers to the origin on incoming transaction and target is the destination. The four features identified: in–in, in–out, out–out, out–in.
- Average amount incoming: The average amount of bitcoins received to the address of the user’s wallet.
- Average Amount outgoing: The average amount of bitcoins sent to the user wallet address.
- Total amount sent: The total amount of bitcoins sent to the user’s wallet address.
- Total amount received: The total amount of bitcoins received to the address of the user’s wallets.
- Standard deviation received: The standard deviation of the number of bitcoins received to the address of the user’s wallets.
- Standard deviation sent: The standard deviation of the number of bitcoins sent to the user’s wallet address.
- Average neighborhood (In-in): The average neighborhood of inputs to inputs of all outputs.
- Average neighborhood (In-out): The average neighborhood of inputs to outputs of all outputs.
- Average neighborhood (Out-in): The average neighborhood of outputs to inputs of all outputs.
- Average neighborhood (Out-out): The average neighborhood of outputs to outputs of all outputs.

3.2. Preprocessing

Since the best results in previous works are related to Monamo et al. [20], and they also used the 14 features listed in Table 2; in this study, investigations were performed on these features, and data preprocessing was conducted in the following three general steps:

- Data wiping: Records that have no input or output are removed. Consequently, the number of records is reduced from 13,086,527 to 10,800,406.
- Data aggregation and data size reduction: to detect collective anomalies in this research, using the Contraction feature, all the addresses, and wallets of a user are aggregated to extract the appropriate features according to this aggregation of data, and as a result, the number of records reached 5,305,678 records.
- Considering that there is a computational relationship between the two in-degree and out-degree features, and according to the principle of data aggregation, two in-degree and out-degree features can be eliminated compared with the method of Monamo et al. [20].
- Because in many thefts, hacking, or fraud cases, the criminals work with multiple addresses and wallets to make it difficult to diagnose the anomaly; therefore, in the previous works that used the method of point anomaly detection, two important features of clustering coefficient and triangle were used to extract better results by realizing the multiplicity of connections between these addresses and wallets. On the other hand, according to the new approach of this research in diagnosing collective anomalies of users (with multiple addresses and possible wallets), instead of identifying point anomalies of addresses and wallets, two clustering coefficients and triangle features require high computational and operational power which can be removed for extraction.
- Data conversion: The min–max linear method was used to normalize the data.

3.3. Feature Extraction

Table 3 shows the employed features of the proposed approach in the research along with a brief description of the features.

Table 3. The features of the proposed approach.

Feature	Definition
Average amount incoming	The average amount of bitcoins received to the address of the user's wallet
Average Amount outgoing	The average amount of bitcoins sent to the user wallet address
total amount sent	The total amount of bitcoins sent to the user's wallet address
total amount received	The total amount of bitcoins received to the address of the user's wallets
standard deviation received	The standard deviation of the number of bitcoins received to the address of the user's wallets
standard deviation sent	The standard deviation of the number of bitcoins sent to the user's wallet address
Average neighborhood (In-in)	The average neighborhood of inputs to inputs of all outputs
Average neighborhood (In-out)	The average neighborhood of inputs to outputs of all outputs
Average neighborhood (Out-in)	The average neighborhood of outputs to inputs of all outputs
Average neighborhood (Out-out)	The average neighborhood of outputs to outputs of all outputs

3.4. Trimmed K-Means Algorithm

Clustering is one of the famous techniques for anomaly detection because clustering potentially throws outlier data into a separate cluster. Among the clustering algorithms, K-means is one of the most popular algorithms. Although some authors, such as [4], believe that K-means is not a technique for outlier detection, it lays the basis to evaluate methods given that outliers will be found furthest from the centroids of clusters they are associated with. Moreover, K-means inherits that lack of robustness from the mean. Instead of K-means, some researchers suggested an extended version of this algorithms is called trimmed K-means. The trimmed K-means is based on partial trimming that is more robust than classical K-means clustering in [35]. The general approach of trimmed K-means is as follows:

The value of α at the input is specified to determine the percentage of outlier data, which is a number between 0 and 1. The K is the number of clusters in the input. A penalty

function is denoted by Φ . For each set A that $P(A) \geq 1 - \alpha$ and any k -set $M = m_1, m_2, \dots, m_k$ in a vector space with d dimensions, the variation of M given A :

$$V_{\Phi}^A(M) = \frac{1}{P(A)} \int_A \Phi(\inf_{i=1, \dots, k} \|X - m_i\|) dP$$

To obtain changes in the k cluster, there is the following relation to minimize M :

$$V_{k, \Phi}^A = \inf_{M \subset R^D, |M|=k} V_{\Phi}^A(M)$$

To obtain a cluster 0 by α percent of the dataset, there is the following relation to minimize A :

$$V_{k, \Phi, \alpha} = V_{k, \Phi, \alpha}(X) = V_{k, \Phi, \alpha}(P_X) = (\inf_{A \in \beta^d, P(A) \geq 1-\alpha} V_{k, \Phi}^A)$$

The main purpose of the algorithm is to obtain a set of outlier data called A_0 , and to obtain k sets that fit inside each cluster, i.e., $M = M_1^0, M_2^0, \dots, M_K^0$, providing the following condition:

$$V_{\Phi}^{A_0}(M_0) = V_{k, \Phi, \alpha}$$

Briefly, in trimmed K-means, by observing the maximum $O(1 - \alpha)$ number of samples, the centers of the clusters can be determined. In this way, by selecting a subset of data, the centers of the clusters can be determined with reasonable accuracy. One of the most important features of this algorithm is to place α percent of the outlier, which is very far from other clusters' centers, in the 0 cluster. This feature is particularly important in the case of the considered problem and anomaly detection.

Due to a large number of records and data dimensions and also the reduction in clustering time, Monamo et al. [4] applied the clustering operation to one million records, but in the proposed method, due to data aggregation and size reduction, the experiment was conducted on all the records to extract more reliable results.

4. Findings

In this section, the results of the experiment are presented and compared with previous works. The proposed approach was run on a VPS Server DL380 G9 with 16 CPU core and 16GB RAM. We used MATLAB for implementation. The MATLAB FSDA toolbox [36] was used for developing the algorithms.

4.1. Experimental Results

As shown in Table 4, the proposed method uses the collective anomaly detection method for the first time compared with previous records. It succeeds in detecting anomalies of users who intend to show their behavior, usually by having multiple wallet addresses, and the proposed method was successful in detecting 14 users with 26 addresses involved in 9 cases of theft, fraud, hacking, or loss.

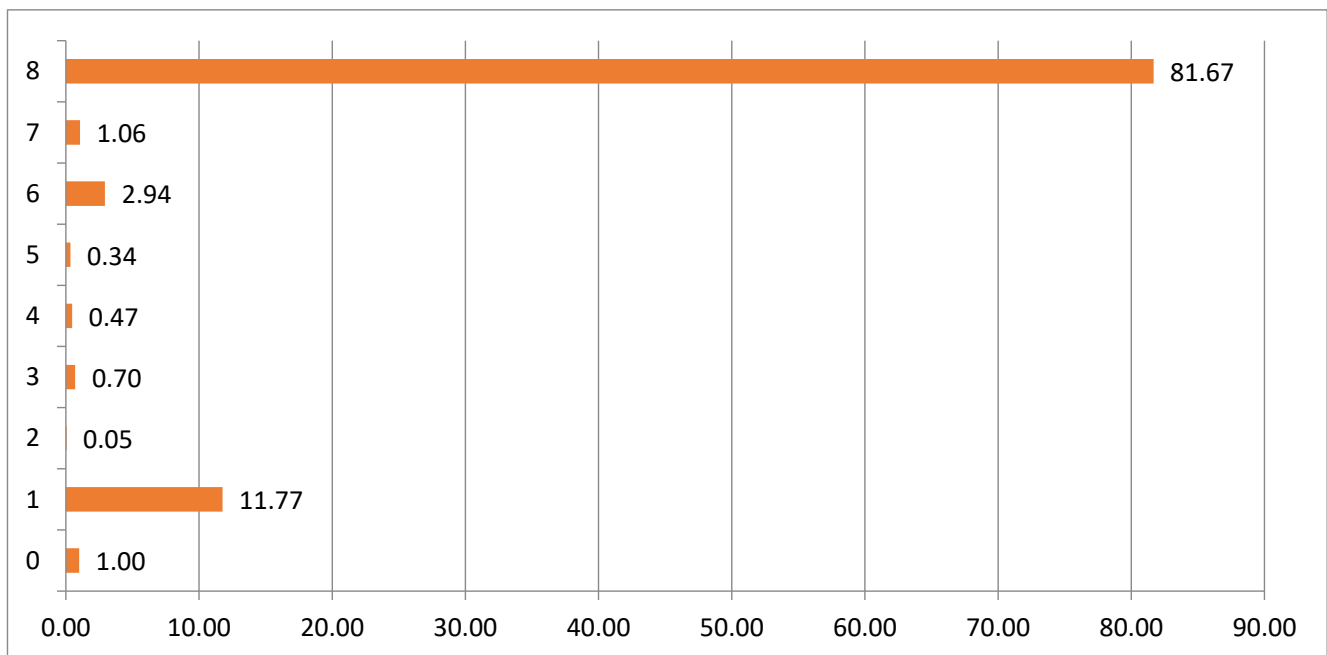
As shown in Figure 1, the detected anomalous addresses are all in the 0 cluster, which is the same as the outliers and makes up exactly one percent of the total data.

4.2. Comparison with Previous Works

The following section presents comparisons of the current study with the previous works. The comparisons are based on features, employed algorithms, and the performance of the studies.

Table 4. Results of anomaly diagnosis with the proposed method.

Row	User Number	Address Number of Wallet	Theft Number	Theft Name
1	882066	882066	1	Stone Man Loss
2	3216635	3216635	3	Stefan thomas loss
3	913570	5034989	4	Allinvain Theft
4	149	5463950	6	Mass MyBitcoin Thefts
5	64	5125978	11	October 2011 Mt. Gox Loss
6	135	924292	14	Linode Hacks
7	135	1095327	14	Linode Hacks
8	135	2000790	14	Linode Hacks
9	135	2021669	14	Linode Hacks
10	135	2720178	14	Linode Hacks
11	135	4941747	14	Linode Hacks
12	135	5679585	14	Linode Hacks
13	731	827543	14	Linode Hacks
14	9538	3283795	14	Linode Hacks
15	9538	5295593	14	Linode Hacks
16	9538	5911894	14	Linode Hacks
17	1363830	2305801	14	Linode Hacks
18	1363830	3707950	14	Linode Hacks
19	1698477	1698477	17	May 2012 Bitcoinica Hack
20	1914	818018	23	Bitfloor Theft
21	1914	1740332	23	Bitfloor Theft
22	1914	4524766	23	Bitfloor Theft
23	1914	5517289	23	Bitfloor Theft
24	833694	833694	23	Bitfloor Theft
25	4212450	4212450	23	Bitfloor Theft
26	7083219	11225439	24	Cdecker Theft

**Figure 1.** Dispersion rate of addresses after clustering in the proposed method.

4.2.1. Comparison of Features

As shown in Table 5, the proposed method is placed in the middle of the table in terms of the number of extracted features. At the same time, the proposed method does not use the clustering coefficient feature, the extraction of which has a high time complexity; therefore, the proposed method has acceptable performance in feature extraction in terms of computational and processing power.

Table 5. Comparison of the number of features, time, and computational and operational power.

Research name	Number of Features	Approximate Computational and Operational Time and Power
Zambre [9]	6	Moderate
Pham [7]	9	Moderate
Pham et al. [21], in the subsequent study	12	High
Monamo [4]	14	High
Monamo [20], in the subsequent study	14	High
Chawathe [16]	14	High
The proposed method	10	Moderate

4.2.2. Comparison of the Used Algorithms

As shown in Table 6, the proposed method was able to detect anomalies using only one algorithm and had a proper performance in selecting the algorithm.

Table 6. Comparison of used algorithms.

Study	The Proposed Algorithm
Zambre [9]	K-Means
Pham [7]	Power Degree and Densification Laws,
Pham [21]	K-Means Clustering, Local Outlier Factor
Monamo [4]	K-Means, Mahalanobis, SVM
Monamo [20]	trimmed_kmeans
Chawathe [16]	kd_tree
The proposed method	trimmed_kmeans

4.2.3. Comparison of Success of the Suggested Approach

As shown in Table 7 and Figure 2, the proposed method identified 26 of the anomalous addresses that were present in the nine detected anomalies, and in this respect, performed better than the previous works.

In terms of the number of features, the lowest number of features is related to Pham [21]. The proposed method is in the middle of the comparison table. Because the proposed method uses the diagnosis of collective anomalies, a reduction in the number of records has been created, and in general, it has been successful in reducing the dimensions (number of records and features).

In terms of time and operational and computational power, the proposed method performed better than previous records that managed to detect anomalies.

In terms of the number of algorithms used to detect anomalies and suspicious transactions, the proposed method using the Trimmed_Kmeans algorithm has performed fine.

The most important part of comparing the proposed method with others is the success in performance and results. In this regard, the proposed method has been able to achieve the best performance compared with other methods and was able to detect 14 users with 26 addresses (wallets) who committed 9 cases of theft, fraud, hacking, or loss, and compared with Monamo's latest method [20], which was able to find 7 addresses (wallets) that committed 5 thefts, scams, hacks, or losses, has a much better performance.

Table 7. Types of thefts, hacks, scams, and losses detected by anomaly detection methods.

Theft Number	Name of Theft, Hack, Fraud, Loss	Pham [7]	Pham [21] in the Subsequent Study	Monamo [4]	Monamo [20], in the Subsequent Study	The Proposed Method
1	Stone Man Loss				1	1
3	Stefan thomas loss					1
4	Allinvain Theft				1	1
5	June 2011 Mt. Gox Incident			1	1	
6	Mass MyBitcoin Thefts					1
11	October 2011 Mt. Gox Loss					1
14	Linode Hacks			3	3	13
17	May 2012 Bitcoinica Hack					1
23	Bitfloor Theft					6
24	Cdecker Theft					1
25	2012 50BTC Theft			1	1	
	Sum	1	3	5	7	26

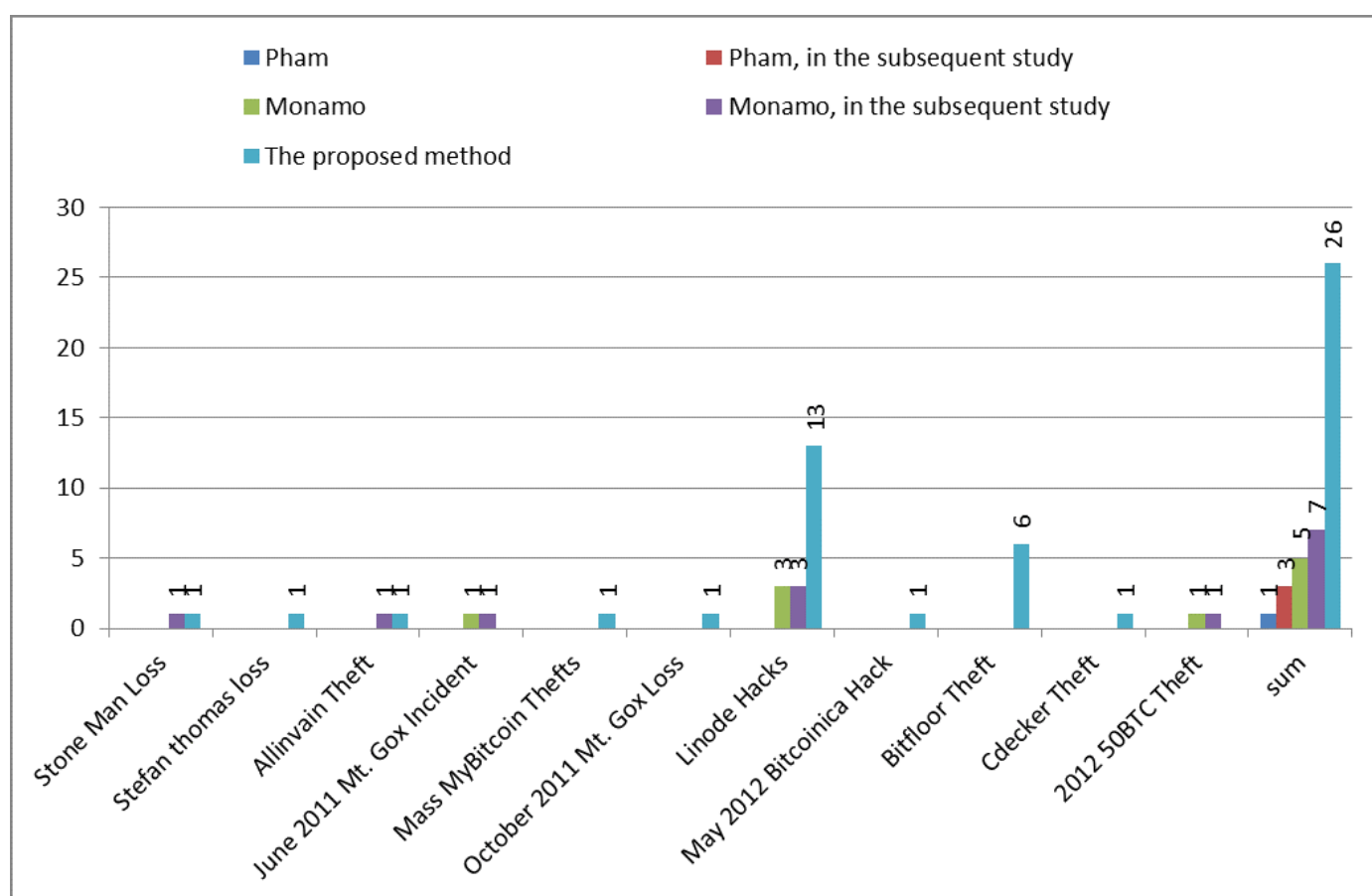
**Figure 2.** Thefts, hacks, scams, and losses detected by various Bitcoin anomaly detection methods.

Figure 3 shows how many frauds have been detected in each method, and how many wallets were involved in the frauds. It should be mentioned the number of detected users has been found only in the proposed method due to the new approach in the detection of anomalies that is user-centered.

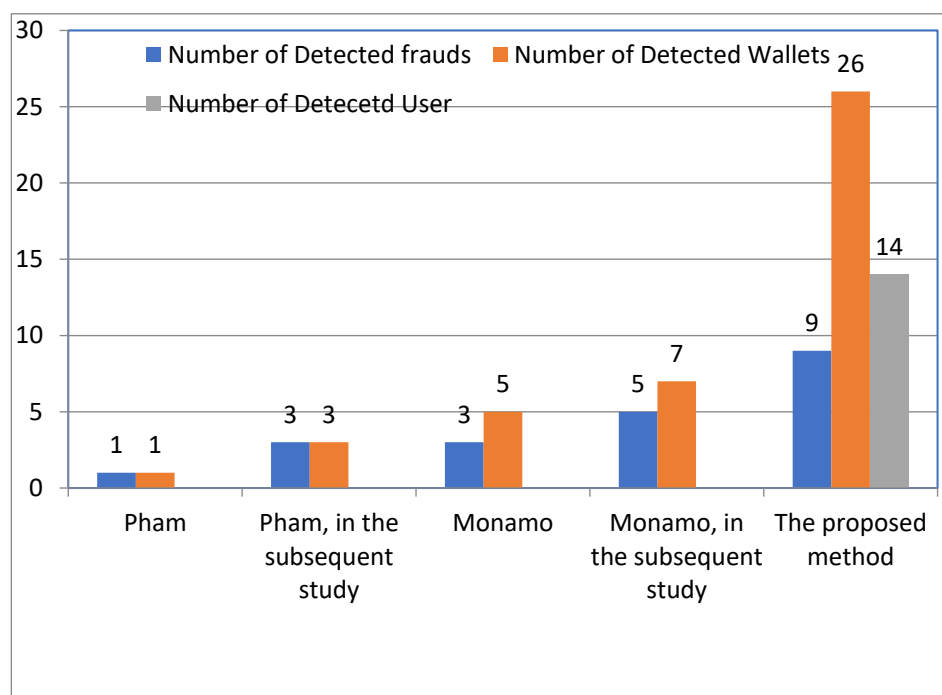


Figure 3. Comparison of the number of detected anomalies in different approaches.

The most important novelty of this research compared with previous methods is the use of collective anomaly detection (user) instead of individual anomalies (wallet address), i.e., instead of seeking anomaly detection in digital wallet addresses, it seeks to detect anomalies among the behavior of users who mainly use several digital wallet addresses. The advantages of the suggested approach are:

1. Using collective anomaly detection (users' behavior with multiple digital wallets) instead of individual anomaly detection (digital wallet behavior).
2. Data aggregation and reduction in data dimensions.
 - a. Remove the two properties, in-degree and out-degree, that had a high correlation with other attributes.
 - b. Elimination of clustering coefficient and triangle with high time overhead due to the use of collective anomaly detection. Employing the collective strategy has been eliminated for these two factors.
 - c. The reduced number of records from 10,800,406 to 5,305,678 records due to aggregation of digital wallet addresses per user.

5. Conclusions and Suggestion

According to the results, it was found that people who intend to commit fraud and malicious activities in the Bitcoin network use several addresses and so-called digital wallets to normalize their activities as normal users. In a way, these users' activity with multiple addresses makes them look almost like normal users. To diagnose this type of anomaly, such as an in-disguise anomaly, one must find a small deviation in these users' behavior. In the previous works, anomaly detection was carried out by extracting new features that rely on the connection between a user's digital wallets. However, in the proposed method, using collective anomaly detection, the user's digital wallets are aggregated, and instead of detecting the anomaly of the wallet address, the anomaly of users who own one or more digital wallets was examined.

On the other hand, due to the significant expansion of this network, it becomes very difficult to extract features that depend on high power or computing time, and in practice, it seems very difficult to detect anomalies in this network with these methods. Therefore, in order to integrate and reduce the problem-solving dimensions of anomaly detection in

Blockchain and Bitcoin networks, four features, two of which had high processing and computing power, were removed. The proposed method also uses the Trimmed_KMeans algorithm for clustering, which has a more robust method for solving anomaly detection problems than similar algorithms such as the KMeans algorithm. In the end, the proposed method was able to identify 14 users who had 26 known anomalous addresses. Thus, in comparison with the previous methods, in addition to reducing the dimensions of the problem from 10,800,406 records to 5,305,678 and also from 14 features to 10, the processing power and computational time of extracting each feature was also reduced. In addition, in the most important part of the evaluation and performance result, the number of detected thefts increased from five to nine compared with the previous best methods, and the number of addresses of the perpetrators of theft was increased from 7 to 26. Additionally, in this method, for the first time, 14 users who committed these cases were identified.

As future works, it is suggested to do new work in two parts in general. In one step, features and algorithms should be selected that require low computational and operational power to extract and execute. In another step, features and algorithms having the best diagnosis of the anomaly should be found.

Author Contributions: All authors declare that they contribute to all parts of this research and the extracted paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Ministry of Science and Technology, Taiwan, R.O.C., under Contract MOST 110-2218-E-305-001-MBK and Contract MOST 110-2410-H-324-004-MY2.

Data Availability Statement: The datasets generated during the current study are available from the authors on reasonable request.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Fischer, A.M. Public Key/Signature Cryptosystem with Enhanced Digital Signature Certification. U.S. Patent 4,868,877, 19 September 1989.
2. Nakamoto, S.; Bitcoin, A. A Peer-to-Peer Electronic Cash System. 2019. Available online: www.bitcoin.org (accessed on 28 November 2021).
3. Wan, S.; Li, M.; Liu, G.; Wang, C. Recent advances in consensus protocols for blockchain: A survey. *Wirel. Netw.* **2020**, *26*, 5579–5593. [\[CrossRef\]](#)
4. Monamo, P.; Marivate, V.; Twala, B. Unsupervised learning for robust Bitcoin fraud detection. In Proceedings of the 2016 Information Security for South Africa (ISSA), Johannesburg, South Africa, 16–18 August 2016; pp. 129–134.
5. Leng, J.; Zhou, M.; Zhao, J.L.; Huang, Y.; Bian, Y. Blockchain Security: A Survey of Techniques and Research Directions. *IEEE Trans. Serv. Comput.* **2020**, *1*. [\[CrossRef\]](#)
6. Kshetri, N. Blockchain's roles in strengthening cybersecurity and protecting privacy. *Telecommun. Policy* **2017**, *41*, 1027–1038. [\[CrossRef\]](#)
7. Pham, T.; Lee, S. Anomaly detection in the bitcoin system—a network perspective. *arXiv* **2016**, arXiv:1611.03942.
8. Prado-Romero, M.A.; Doerr, C.; Gago-Alonso, A. Discovering bitcoin mixing using anomaly detection. In *Iberoamerican Congress on Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 534–541.
9. Zambre, D.; Shah, A. Analysis of Bitcoin network dataset for fraud. *Unpubl. Rep.* **2013**, *27*, 2013.
10. Xu, J.J. Are blockchains immune to all malicious attacks? *Financ. Innov.* **2016**, *2*, 1–9. [\[CrossRef\]](#)
11. Li, Z.; Xiang, Z.; Gong, W.; Wang, H. Unified model for collective and point anomaly detection using stacked temporal convolution networks. *Appl. Intell.* **2021**, 1–14. [\[CrossRef\]](#)
12. Bariviera, A.F.; Merediz-Solà, I. Where do we stand in cryptocurrencies economic research? A survey based on hybrid analysis. *J. Econ. Surv.* **2021**, *35*, 377–407. [\[CrossRef\]](#)
13. Corbet, S.; Lucey, B.; Urquhart, A.; Yarovaya, L. Cryptocurrencies as a financial asset: A systematic analysis. *Int. Rev. Financ. Anal.* **2019**, *62*, 182–199. [\[CrossRef\]](#)
14. Guesmi, K.; Saadi, S.; Abid, I.; Ftiti, Z. Portfolio diversification with virtual currency: Evidence from bitcoin. *Int. Rev. Financ. Anal.* **2019**, *63*, 431–437. [\[CrossRef\]](#)
15. Vidal-Tomás, D.; Ibáñez, A.M.; Farinós, J.E. Weak efficiency of the cryptocurrency market: A market portfolio approach. *Applied Econ. Lett.* **2019**, *26*, 1627–1633. [\[CrossRef\]](#)
16. Chawathe, S.S. Clustering blockchain data. In *Clustering Methods for Big Data Analytics*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 43–72.

17. Hirshman, J.; Huang, Y.; Macke, S. *Unsupervised Approaches to Detecting Anomalous Behavior in the Bitcoin Transaction Network*, 3rd ed.; Technical Report; Stanford University: Stanford, CA, USA, 2013.
18. Huang, B.; Liu, Z.; Chen, J.; Liu, A.; Liu, Q.; He, Q. Behavior pattern clustering in blockchain networks. *Multimed. Tools Appl.* **2017**, *76*, 20099–20110. [\[CrossRef\]](#)
19. Meiklejohn, S.; Pomarole, M.; Jordan, G.; Levchenko, K.; McCoy, D.; Voelker, M.G.; Savage, S. A fistful of bitcoins: Characterizing payments among men with no names. In Proceedings of the 2013 Conference on Internet Measurement Conference, Barcelona, Spain, 23–25 October 2013; pp. 127–140.
20. Monamo, P.M.; Marivate, V.; Twala, B. A multifaceted approach to Bitcoin fraud detection: Global and local outliers. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; pp. 188–194.
21. Pham, T.; Lee, S. Anomaly detection in bitcoin network using unsupervised learning methods. *arXiv* **2016**, arXiv:1611.03941.
22. Signorini, M.; Pontecorvi, M.; Kanoun, W.; di Pietro, R. BAD: Blockchain Anomaly Detection. *arXiv* **2018**, arXiv:1807.03833.
23. Zhang, Q.; Wan, S.; Wang, B.; Gao, D.W.; Ma, H. Anomaly detection based on random matrix theory for industrial power systems. *J. Syst. Archit.* **2019**, *95*, 67–74. [\[CrossRef\]](#)
24. Zhang, R.; Zhang, G.; Liu, L.; Wang, C.; Wan, S. Anomaly detection in bitcoin information networks with multi-constrained meta path. *J. Syst. Archit.* **2020**, *110*, 101829. [\[CrossRef\]](#)
25. Bhowmik, M.; Chandana, T.S.S.; Rudra, B. Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain. In Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 8–10 April 2021; pp. 539–541.
26. Dhieb, N.; Ghazzai, H.; Besbes, H.; Massoud, Y. A secure ai-driven architecture for automated insurance systems: Fraud detection and risk measurement. *IEEE Access* **2020**, *8*, 58546–58558. [\[CrossRef\]](#)
27. Kamišalić, A.; Kramberger, R.; Fister, I. Synergy of Blockchain Technology and Data Mining Techniques for Anomaly Detection. *Appl. Sci.* **2021**, *11*, 7987. [\[CrossRef\]](#)
28. Liu, L.; Tsai, W.-T.; Bhuiyan, M.Z.A.; Peng, H.; Liu, M. Blockchain-enabled fraud discovery through abnormal smart contract detection on Ethereum. *Future Gener. Comput. Syst.* **2022**, *128*, 158–166. [\[CrossRef\]](#)
29. Lorenz, J.; Silva, M.I.; Aparício, D.; Ascensão, J.T.; Bizarro, P. Machine learning methods to detect money laundering in the Bitcoin blockchain in the presence of label scarcity. *arXiv* **2020**, arXiv:2005.14635.
30. Martin, K.; Rahouti, M.; Ayyash, M.; Alsmadi, I. Anomaly detection in blockchain using network representation and machine learning. *Secur. Priv.* **2021**, e192. [\[CrossRef\]](#)
31. Podgorelec, B.; Turkanović, M.; Karakatič, S. A machine learning-based method for automated blockchain transaction signing including personalized anomaly detection. *Sensors* **2020**, *20*, 147. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Pourhabibi, T.; Ong, K.-L.; Kam, B.H.; Boo, Y.L. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decis. Support Syst.* **2020**, *133*, 113303. [\[CrossRef\]](#)
33. Wang, T.; Wu, X.; He, T. Trustable and Automated Machine Learning Running with Blockchain and Its Applications. *arXiv* **2019**, arXiv:1908.05725.
34. Kondor, D.; Pósfai, M.; Csabai, I.; Vattay, G. Do the rich get richer? An empirical analysis of the Bitcoin transaction network. *PLoS ONE* **2014**, *9*, e86197.
35. Cuesta-Albertos, J.A.; Gordaliza, A.; Matrán, C. Trimmed k-means: An attempt to robustify quantizers. *Ann. Stat.* **1997**, *25*, 553–576. [\[CrossRef\]](#)
36. FSDA Matlab Code. EU SCIENCE HUB. Available online: <https://ec.europa.eu/jrc/en/scientific-tool/fsda-matlab-code> (accessed on 7 August 2021).