Malicious Bitcoin Transaction Tracing Using Incidence Relation Clustering

Baokun Zheng[1,2], Liehuang Zhu[1], Meng Shen[1](✉), Xiaojiang Du[3], Jing Yang[1], Feng Gao[1], Yandong Li[1], Chuan Zhang[1], Sheng Liu[4], and Shu Yin[4]

[1] Beijing Institute of Technology, Beijing, China
zhengbk168@163.com, {liehuangz,shenmeng,jingy,leeyandong}@bit.edu.cn,
gaofengbit@foxmail.com, chuanzdlut@163.com
[2] China University of Political Science and Law, Beijing, China
[3] Temple University, Philadelphia, USA
dxj@ieee.org
[4] Union Mobile Financial Technology Co., Ltd., Beijing, China
{liusheng,yinshu}@umfintech.com

Abstract. Since the generation of Bitcoin, it has gained attention of all sectors of the society. Law breakers committed crimes by utilizing the anonymous characteristics of Bitcoin. Recently, how to track malicious Bitcoin transactions has been proposed and studied. To address the challenge, existing solutions have limitations in accuracy, comprehensiveness, and efficiency. In this paper, we study Bitcoin blackmail virus WannaCry event incurred in May 2017. The three Bitcoin addresses disclosed in this blackmail event are only restricted to receivers accepting Bitcoin sent by victims, and no further transaction has been found yet. Therefore, we acquire and verify experimental data by example of similar Bitcoin blackmail virus CryptoLocker occurred in 2013. We focus on how to track malicious Bitcoin transactions, and adopt a new heuristic clustering method to acquire incidence relation between addresses of Bitcoin and improved Louvain clustering algorithm to further acquire incidence relation between users. In addition, through a lot of experiments, we compare the performance of our algorithm with another related work. The new heuristic clustering method can improve comprehensiveness and accuracy of the results. The improved Louvain clustering algorithm can increase working efficiency. Specifically, we propose a method acquiring internal relationship between Bitcoin addresses and users, so as to make Bitcoin transaction deanonymisation possible, and realize a better utilization of Bitcoin in the future.

Keywords: Bitcoin · Blockchain · Incidence relation · Cluster

1 Introduction

On May 12, 2017, Bitcoin blackmail virus WannaCry was burst globally. Criminals blackmailed Bitcoin [1] equaling to USD 300 from users infected with the

J. Hu et al. (Eds.): MONAMI 2017, LNICST 235, pp. 313–323, 2018.
https://doi.org/10.1007/978-3-319-90775-8_25

virus. For a short while, many users in the world suffered from serious loss. Bitcoin can be sent by anyone to any other person everywhere. Bitcoin uses a public key based wallet address as a pseudonym on the blockchain, where transactions between different users are realized through this pseudonym. Bitcoin accounts are anonymous and cannot be reviewed. In order to implement anonymous transactions, Bitcoin system allows users to generate multiple wallets addresses freely. Users can use different wallet addresses to reduce the transaction characteristics of individual wallet addresses. Because of the anonymity of Bitcoin account, Bitcoin may be used for some illegal behavior and the black market, such as the purchase of guns and drugs. Thus, obtaining trading rules by analyzing the user transaction records, and even speculating the identity information of users, is particularly important in the prevention of crime.

[2–9] studied the relationship between the Bitcoin addresses based on the heuristic method. However, the comprehensiveness of the heuristic methods is not fully considered in these papers. They did not study the output addresses of coinbase transaction, and the judgement on change address is insufficient. In addition, few papers have studied the relation between users.

Under such background, this paper aims to better known traceability of Bitcoin movement and explore better use in the future. Most importantly, this paper does not aim to carry out deanonymisation for all Bitcoin users because it is impossible according to abstract user protocol design. Instead, this paper aims to recognize anonymous users according to specific behaviors of Bitcoin users in Bitcoin network.

In this paper, the methodology is based on the availability of Bitcoin blockchain, using digital signature secret key disclosed on every transaction, and decoding a graphic data structure by Bitcoin activities. To summarize, the ***contributions*** of our work include:

(1) We propose a new heuristic method with three rules to acquire incidence relation between addresses of Bitcoin. In this new method, multi-input transactions, coinbase transactions and change address are studied. It improves the accuracy and comprehensiveness of the relationship between Bitcoin addresses. We verified the comprehensiveness and accuracy of the actual transaction through the Bitcoin addresses we controlled, and the results reach 100%. By using this method, we find 2118 CryptoLocker blackmail addresses.

(2) We use Louvain [10] clustering algorithm to analyze relation between users. Louvain clustering algorithm that is a community division method based hierarchical clustering can divide transaction addresses closely related to a community so as to find out incidence relation between Bitcoin users. We also improve Louvain clustering algorithm in this paper. In the graph of Bitcoin transactions, we preprocess the leaf nodes and carry out the optimization module of coarse-grained inverse operation in order to increase the community modularity and working efficiency.

(3) Performance evaluation via extensive experiments also demonstrates that our methods can efficiently trace Bitcoin transaction.

The remainder of this paper is organized as follows: Sect. 2 introduces the method model, definitions, and preliminaries of our work; Sect. 3 gives a concrete description of our measurement methodology; Sect. 4 carries out performance analysis; Sect. 5 introduces relevant work; Sect. 6 concludes this paper.

2 Model, Definitions and Preliminaries

2.1 Model

We show the system model of transaction tracking in Fig. 1. There are three processes in the model: Data acquisition, Data analysis, and Data presentation. Data acquisition contains Bitcoin transaction data and Bitcoin addresses with disclosed identity in network. Data analysis contains acquisition of transaction relation between Bitcoin addresses and confirmation of relationship between Bitcoin users. Data presentation presents relationship between Bitcoin addresses and users in a visualized way.

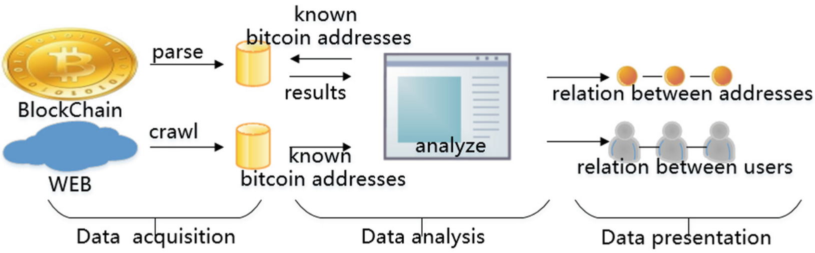


Fig. 1. System model of transaction tracking.

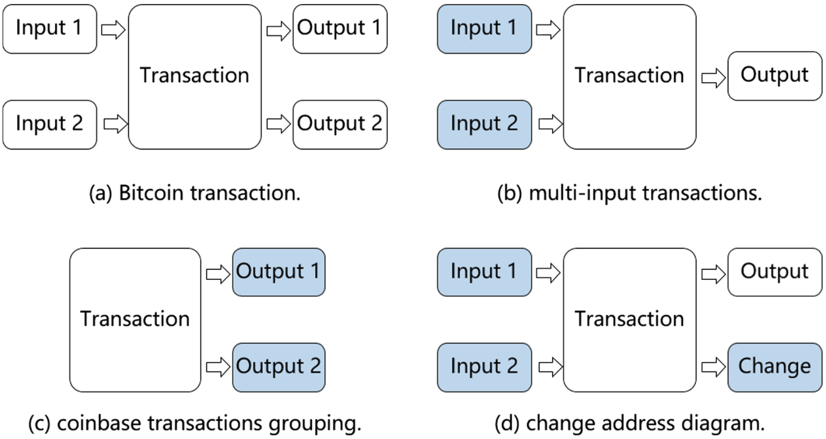
2.2 Definitions

Definition 1 (Address attribution). The user set is represented as $U = \{u_1, u_2, \dots u_n\}$, the Bitcoin address set is represented as $A = \{a_1, a_2, \dots a_n\}$, and the transaction set is represented as $T = \{t_1, t_2, \dots, t_n\}$. The input transaction is represented as Input(t), and the output of the transaction is represented as Output(t).

Bitcoin transaction consists of a set of input addresses, a set of output addresses and change address.

Definition 2 (Change address). If a public key ***pk*** meets the following conditions, the ***pk*** is the one-time change address of transaction ***t***:

- The ***pk*** is only as output of one transaction ***t***.
- Transaction ***t*** is not a coinbase transaction.
- For ***pk'*** $\in$ ***output(t)***, no ***pk'***$\in$ ***inputs(t)***, i.e. transaction ***t*** is not a transaction of "self-change".
- No ***pk'***$\in$ ***output(t)***, and ***pk'***$\neq$***pk***, but ***pk*** is used as transaction output more than once.



(a) Bitcoin transaction.

(b) multi-input transactions.

(c) coinbase transactions grouping.

(d) change address diagram.

Fig. 2. Bitcoin transaction.

Definition 3 (Transaction Matrix). When presenting data, we need to convert the user's transaction data into a matrix. Give an atlas $\boldsymbol{G = (V, E)}$. $\boldsymbol{V}$ represents cluster of vertexes of atlas $\boldsymbol{G}$ which is transferred from account addresses in Bitcoin transaction network. $\boldsymbol{E}$ is the cluster of edges of atlas $\boldsymbol{G}$ and is transferred from transaction relation between account addresses in Bitcoin network.

2.3 Preliminaries

In this section, we formalize: (i) Bitcoin transaction process revealing incidence relation between Bitcoin addresses, and (ii) Louvain algorithm that shows principle of clustering Bitcoin addresses and incidence relation between users.

Bitcoin Transaction. Bitcoin transaction comprises a group of input, output and change address. The input addresses belong to the payer, output addresses belong to the receiver, and change address (optional) is used to store the remaining Bitcoin after transaction, belonging to the payer. The transaction protocol of Bitcoin regulates that the input of a new transaction must be explicit value outputted by previous transaction. Transactions can be divided to single input and multiple outputs, multiple inputs and single output, and multiple inputs and multiple outputs, as shown in Fig. 2(a). In the figure, the input of a new transaction may refer to outputs of multiple transactions previously [2,3,9].

Louvain Algorithm. Louvain [10] algorithm is based on modularity increment ΔQ, and mainly divided to two stages. Firstly, every node is initiated as a community. All nodes in network are traversed ceaselessly, and taken out from original community. The modularity increment generated by the node's joining in each community is calculated. If the modularity increment is larger than zero, the community with maximum modularity increment shall be selected, and combined with the node. Aforesaid process is repeated until community is not integrated in network [11]. Secondly, a new network will be constructed according to the first layer of community divided, and the weight between new nodes is the weight between original communities. The process of aforesaid stage will be repeated until no community can be combined.

3 Measurement Methodology

3.1 Data Acquisition

Bitcoin Transaction Data. All transaction data used in this paper is confirmed Bitcoin transactions. We collected them from the blockchain maintained in Bitcoin system, starting from block 1 to block 464283 corresponding to the block creation time from the first one on Jan. 3, 2009 to May 1, 2017 by Bitcoin client with total capacity of 106.87 GB. During this period, a total number of 236242063 Bitcoin transactions have been successfully released and globally confirmed. After acquiring historical transaction, the improved Bitcoin-DatabaseGenerator [12] tool is used to parse the data to acquire data including 284821377 distinct Bitcoin addresses. All the results from this paper are based upon the transactions and addresses from this data set.

Tracing Data. This paper aims to study the Bitcoin blackmail event in May 2017. But study shows the three Bitcoin addresses disclosed in the blackmail event are only restricted to receivers accepting Bitcoin sent by victims, and no further transaction has been found yet. Therefore, this paper acquires and verifies experimental data by example of similar Bitcoin blackmail virus CryptoLocker occurred in 2013 to verify research designs. Program Scrapy web spider and get disclosed Bitcoin addresses in relevant forum to trace flow of Bitcoin. By Scrapy, 5 CryptoLocker blackmail addresses are acquired.

3.2 Data Analysis

Confirmation of Bitcoin Addresses' Incidence Relation. In Bitcoin transaction network, the addresses are connected by transaction activities. Thus, it can be confirmed that two addresses connected are of certain incidence relation; and the source and flow direction of Bitcoin can be known according to characteristics of Bitcoin transaction protocol. We propose a new heuristic method with three rules to acquire incidence relation between addresses of Bitcoin. The method can confirm which Bitcoin addresses belong to the same user, and it can be concluded and described as:

- ***multi-input transactions grouping***
  If two or more addresses are inputs of the same transaction, they are controlled by the same user; for instance, for any transaction $\boldsymbol{t}$, all $\boldsymbol{pk} \in \boldsymbol{inputs(t)}$ are controlled by the same user, as shown in Fig. 2(b).
- ***coinbase transactions grouping***
  If two or more addresses are outputs of the same coinbase transaction, they are controlled by the same user; for instance, for any coinbase transaction, all $\boldsymbol{pk} \in \boldsymbol{output(t)}$ are controlled by the same user, as shown in Fig. 2(c).
- ***change address guessing***
  The one-time change address and transaction input addresses are controlled by the same user; for instance, for any transaction $\boldsymbol{t}$, the controller of $\boldsymbol{input(t)}$ controls the one-time change address $\boldsymbol{pk} \in \boldsymbol{output(t)}$, as shown in Fig. 2(d).

Confirmation of Bitcoin Users' Incidence Relation. During transaction, every user can participate in transaction by multiple Bitcoin addresses. As previously mentioned, different Bitcoin addresses of the same user are confirmed. However, for the Bitcoin blackmail event just occurred in May 2017, maybe criminal gang comprises many people and everyone has multiple Bitcoin address. After the blackmail, criminals gathered the Bitcoin blackmailed to an account of a higher-level member. In this section, we set up a data set of all blackmail addresses which we found to determine the relationship between users. The strategy of this paper is to adopt Louvain algorithm to confirm Bitcoin users' incidence relation and improve it. We improve time complexity and optimal modular based on the enhanced Louvain algorithm proposed by Gach and Hao [13], and it can be concluded and described as:

- ***Node pretreatment***
  In Bitcoin transaction network, account address participating in transaction is node of the network; transactions between accounts are edges connecting nodes. For address $\boldsymbol{i}$ and address $\boldsymbol{j}$, if $\boldsymbol{i}$ is the only connection of $\boldsymbol{j}$, address $\boldsymbol{j}$ will surely be divided to the same community with address $\boldsymbol{i}$, which can be proved:
  If as assumed before, node $\boldsymbol{i}$ and node $\boldsymbol{j}$ belong to the same community, then:
  $Q_{i\to C(j)} = \sum(A_{ij}) - a_j^2$
  Increment of corresponding modularity:
  $\triangle Q_{i\to C(j)} = \frac{\sum_{i.j}}{2m^2}(2m - k_j)$
  If node $\boldsymbol{i}$ and node $\boldsymbol{j}$ are not in the same community, $Q_{i\to C(j)} \leq 0$, then $2m < k_j$. Thus, if node $\boldsymbol{i}$ is the only connection of node $\boldsymbol{j}$, node $\boldsymbol{j}$ will surely be divided to the same community with node $\boldsymbol{i}$. If node $\boldsymbol{j}$ is classified before community division, the modularity $\boldsymbol{Q}$ computation of certain nodes can be reduced so as to improve efficiency of community division.
- ***optimized modularity***
  In the enhanced Louvain algorithm proposed by Gach and Hao [13]. During coarseness inverse operation optimization, the attribution of nodes in the community connecting with outside will be confirmed again by K-medoids

algorithm [14]. Take the node and the node most closely connecting to node in the community as mass points to calculate community attribution of the node. Since number of V_i and V_j is limited, K-medoids algorithm implementation efficiency is high, which can save time to recalculate several $\varDelta Q$ of every V_i, so as to improve efficiency.

3.3 Data Presentation

To better understand the relation of Bitcoin addresses and users, we use Gephi [15], an open source software for network visualization and exploration, to visualize outgoing transactions from Bitcoin transaction data.

Presentation of Addresses Transaction Relation. Take a CryptoLocker blackmail address to carry out test. We use "multi-input transactions grouping", "coinbase transactions grouping", and "change address guessing" to cluster addresses, and the heuristic method iterate tenth. The addresses incidence relation graph is acquired. The addresses belong to the same user. Take these addresses as vertexes, and transactions between addresses as edges; save after converting to graph, and output visually.

Presentation of Users Incidence Relation. As previously mentioned, we set up a data set of all blackmail addresses which we found. Take these addresses as vertexes, and transaction between addresses as edges. By improved Louvain algorithm, we mark the addresses belonging to the same user as the same color, and we distinguish 17 distinct sub-communities in the CryptoLocker blackmail addresses network. We see that the ransom balances from all addresses within a community are transferred to a single aggregate address at the center.

4 Performance Analysis

The scheme is implemented in Python language. The database of Bitcoin blockchain data storage is SQL Server. The experimental machine with 2.40 GHz, Intel i5-2430M CPU and 8 GB RAM.

4.1 Comprehensiveness and Accuracy

We have carried out 45 transactions on the two Bitcoin addresses which we have controlled. The actual transaction addresses involving our control are 126 and 158 respectively, and the experimental results are in good agreement with the actual data, as shown in Table 1. The results indicate that the heuristic method of this paper has a good comprehensiveness and accuracy.

We use the CryptoLocker blackmail addresses which are acquired By Scrapy. Meanwhile, according to burst time of the virus and amount of Bitcoin transacted, 2118 victim addresses, 795 transactions and 1128.40 Bitcoin value are inquired from historical Bitcoin transaction data by our heuristic method.

Table 1. Experimental results of test addresses

Test address sample	Number of associated addresses	Comprehensive rate	Accuracy rate
1***P	126	100%	100%
1***y	158	100%	100%

4.2 Efficiency

In the Bitcoin transaction network, the address represents the vertex, and the transaction between the addresses represents the edge. As mentioned earlier, there are 2118 vertices and 795 edges. In order to verify the efficiency of the algorithm, the Bitcoin transaction data is processed by Louvain algorithm and improved Louvain algorithm respectively, and the data volume is gradually increased. Figure 3 shows a comparison of accumulated average runtime between Louvain algorithm and improved Louvain algorithm with different data volume. The results show that improved algorithm reduces runtime by about 4.533% compared with the original algorithm.

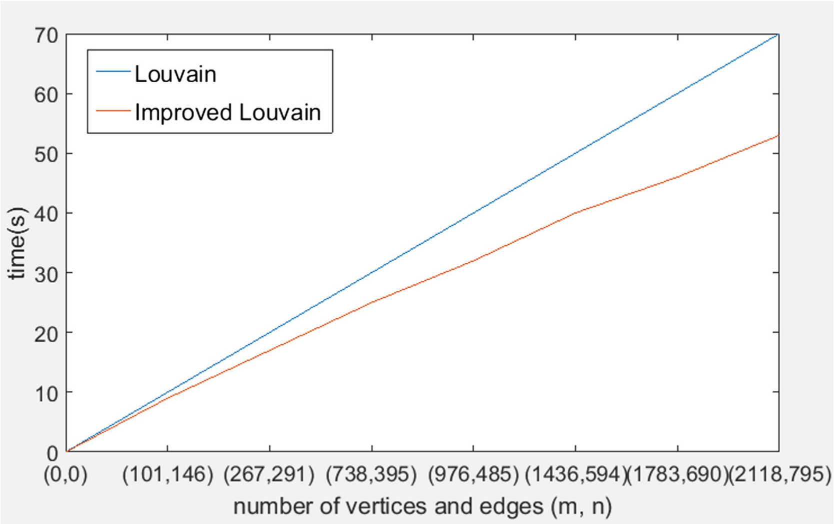


Fig. 3. Runtime comparison between Louvain and improved Louvain.

5 Related Work

In recent years, security and privacy issues have been a hot topic of research [16–25, 27, 28], and Bitcoin privacy issues are a key attention. The related work can be classified into two categories: clustering analysis based on incidence relationship of Bitcoin address and Louvain community algorithm clustering analysis.

5.1 Clustering Analysis Based on Incidence Relationship of Bitcoin Addresses

Large number of researches show inherent flaws of Bitcoin system in privacy. Reid and Harrigan [2] found incidence relationship between transaction addresses by studying Bitcoin transaction, generated transaction network and user network, analyzed quantity, amount and related address of transaction incurred by addresses that Wiki Leaks disclosed so as to find out flow direction of Bitcoin. Meiklejohn et al. [3] heuristically studied cluster of multi-input transaction address and change address. Ron and Shamir [4] generated user map pursuant to incidence relationship of transactions, carried out in-depth study on the largest Bitcoin transaction in history, and concluded based on the data that Bitcoin system has large amount of hoarding behavior, and most capital is not circulating. Androulaki et al. [5] carried out Bitcoin privacy test by actual Bitcoin system and simulating Bitcoin system in university; deeply studied change address; 40% users were participated in the test, and user data could effectively realize deanonymisation of Bitcoin by clustering technology with accuracy of 80%. Zhao [6] studied clustering of multi-input transaction, coinbase transaction and change address, and found out flow direction of Bitcoin. Spagnuolo et al. [7] used heuristic method to realize clustering of multi-input transaction and change address. Monaco [8] identified user by measuring biological characteristics of user identity according to time sequence of transaction sample during a period of time. Liao et al. [9] studied flow of blackmail software ransom, and traced blackmailing addresses by classifying addresses receiving ransom by clustering technology.

5.2 Louvain Community Algorithm

Bitcoin transaction data is vast, and relationship is complicated. The community division method based hierarchical clustering can divide transaction addresses closely related to a community so as to find out incidence relation between Bitcoin addresses. Blondel et al. [10] proposed Louvain algorithm, which divides community by modularity calculation, and can rapidly process big data. Gach and Hao [13] proposed an enhanced Louvain algorithm, and adopted multi-level method to maximize module. De Meo et al. [26] optimizes modularization ideas of Louvain algorithm. The optimal program is to realize maximum network module by computing route from central point so as to improve operating efficiency.

6 Conclusion

This paper clusters incidence relation between Bitcoin addresses by a new heuristic method, and further confirms incidence relation between users by improved Louvain algorithm. However, the heuristic method mentioned in this paper may generate certain error for judgment on change address. Louvain algorithm needs to be further improved for efficiency implementation. Different iteration frequencies of the two methods may lead in different quantities. But the larger the iteration frequency is, and the lower efficiency will be. In the future, studies can be carried out around those problems.

Acknowledgements. This work was supported in part by the National Science Foundation of China under Grant 61602039, in part by the Beijing Natural Science Foundation under Grant 4164098, in part by CCF-Venustech Open Research Fund, in part by BIT-UMF research and development fund, and in part by Education and Teaching Reform Project of China University of Political Science and Law under Grant 1000/10717130.

References

1. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system. Consulted (2008)
2. Reid, F., Harrigan, M.: An analysis of anonymity in the bitcoin system, pp. 1318–1326 (2011)
3. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., Mccoy, D., Voelker, G.M., Savage, S.: A fistful of Bitcoins: characterizing payments among men with no names. In: Conference on Internet Measurement Conference, pp. 127–140. ACM (2013)
4. Ron, D., Shamir, A.: Quantitative analysis of the full Bitcoin transaction graph. In: Sadeghi, A.-R. (ed.) FC 2013. LNCS, vol. 7859, pp. 6–24. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39884-1_2
5. Androulaki, E., Karame, G.O., Roeschlin, M., Scherer, T., Capkun, S.: Evaluating user privacy in Bitcoin. In: Sadeghi, A.-R. (ed.) FC 2013. LNCS, vol. 7859, pp. 34–51. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39884-1_4
6. Zhao, C.: Graph-based forensic investigation of Bitcoin transactions (2014)
7. Spagnuolo, M., Maggi, F., Zanero, S.: BitIodine: extracting intelligence from the bitcoin network. In: Christin, N., Safavi-Naini, R. (eds.) FC 2014. LNCS, vol. 8437, pp. 457–468. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-45472-5_29
8. Monaco, J.V.: Identifying Bitcoin users by transaction behavior. In: SPIE DSS (2015)
9. Liao, K., Zhao, Z., Doupe, A., Ahn, G.J.: Behind closed doors: measurement and analysis of CryptoLocker ransoms in Bitcoin. In: Electronic Crime Research, pp. 1–13. IEEE (2016)
10. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech. Theor. Exp. **30**, 155–168 (2008)
11. (U) Bitcoin Virtual Currency: Unique Features Present Distinct Challenges for Deterring Illicit Activity (2011)
12. https://github.com/ladimolnar/BitcoinDatabaseGenerator/releases
13. Gach, O., Hao, J.-K.: Improving the Louvain algorithm for community detection with modularity maximization. In: Legrand, P., Corsini, M.-M., Hao, J.-K., Monmarché, N., Lutton, E., Schoenauer, M. (eds.) EA 2013. LNCS, vol. 8752, pp. 145–156. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11683-9_12
14. Park, H.S., Jun, C.H.: A simple and fast algorithm for K-medoids clustering. Expert Syst. Appl. **36**(2), 3336–3341 (2009)
15. Gephi. https://gephi.org/
16. Shen, M., Ma, B., Zhu, L., Mijumbi, R., Du, X., Hu, J.: Cloud-based approximate constrained shortest distance queries over encrypted graphs with privacy protection. IEEE Trans. Inf. Forensics Secur. **13**(4), 940–953 (2018)

17. Du, X., Shayman, M., Rozenblit, M.: Implementation and performance analysis of SNMP on a TLS/TCP base. In: IEEE/IFIP International Symposium on Integrated Network Management Proceedings IEEE, pp. 453–466 (2001)
18. Du, X., Wu, D.: Adaptive cell relay routing protocol for mobile ad hoc networks. IEEE Trans. Veh. Technol. **55**(1), 278–285 (2006)
19. Zhang, M., Nygard, K.E., Guizani, S.: Self-healing sensor networks with distributed decision making. Int. J. Sens. Netw. **2**(5/6), 289–298 (2007)
20. Du, X., et al.: An effective key management scheme for heterogeneous sensor networks. Ad Hoc Netw. **5**(1), 24–34 (2007)
21. Du, X., Chen, H.H.: Security in wireless sensor networks. Wirel. Commun. IEEE **15**(4), 60–66 (2008)
22. Xiao, Y., Chen, H.H., Du, X., et al.: Stream-based cipher feedback mode in wireless error channel. IEEE Trans. Wirel. Commun. **8**(2), 622–626 (2009)
23. Du, X., Guizani, M., Xiao, Y., Chen, H.H.: A routing-driven elliptic curve cryptography based key management scheme for heterogeneous sensor networks. IEEE Trans. Wirel. Commun. **8**(3), 1223–1229 (2009)
24. Yao, X., Han, X., Du, X., Zhou, X.: A lightweight multicast authentication mechanism for small scale IoT applications. IEEE Sens. J. **13**(10), 3693–3701 (2013)
25. Liang, S., Du, X.: Permission-combination-based scheme for Android mobile malware detection. In: IEEE International Conference on Communications, pp. 2301–2306. IEEE (2014)
26. De Meo, P., Ferrara, E., Fiumara, G., Provetti, A.: Generalized Louvain method for community detection in large networks. In: International Conference on Intelligent Systems Design and Applications, pp. 88–93. IEEE (2012)
27. Fahad, A., Alshatri, N., Tari, Z., et al.: A survey of clustering algorithms for big data: taxonomy and empirical analysis. IEEE Trans. Emerg. Top. Comput. **2**(3), 267–279 (2014)
28. Almalawi, A.M., Fahad, A., Tari, Z., Cheema, M.A., Khalil, I.: kNNVWC: an efficient k-nearest neighbors approach based on various-widths clustering. IEEE Trans. Knowl. Data Eng. **28**(1), 68–81 (2016)