"Summary of Understanding Ethereum via Graph Analysis[23]"

Ting Chen, AYuxiao Zhu, Zihao Li, Jiachi Chen, Xiaoqi Li, Xiapu Luo, and Xiaodong Lin

(Blockchain Data Analytics, University of Manitoba, Fall 2020)

1. Introduction:

Ethereum is one of the most important cryptocurrencies and the most significant cryptocurrency that supports smart contracts in terms of market capitalization. A smart contract is a program that is automatically executed based on how it is defined before by the users.

In this paper, the authors intend to conduct an analysis of Etheruem based on graphs. Their methodology consists of constructing three main graphs, which are MFG(money flow graph), smart contract creation graph(CCG), and smart contract invocation graph(CIG), and they intend to extract useful information from the Etheruem by using these main tools. Their analysis is based on some metrics such as degree distribution, clusters, degree correlation, node activity, strongly/weakly connected components, Etc.

2. Contributions:

The main contributions of this paper could be summarized as follows:

- This paper was the first paper to construct systematic analysis on Etheruem while using the graph analysis as the main methodology. The authors introduce a novel approach to collect the entire transaction data and construct the MFG(money flow graph), smart contract creation graph(CCG), and smart contract invocation graph(CIG) for analysis.
- From this study, authors received some new observations regarding Ethereum, which they deem very valuable.
- The authors also introduce novel approaches for handling some security issues in Ethereum, such as attack forensics and anomaly detection. Furthermore, they verify how their approaches could be useful by going through some real cases.

3. Background and Data Collection:

Ethereum is the most prominent cryptocurrency that supports smart contracts. One of the features that ethereum provides is that the users could develop many applications based on smart contracts, which will run in Ethereum virtual machine. Paying a transaction fee is mandatory in Ethereum in order to avoid malicious tasks. In Ethereum, the basic unit is account[9]. Every transaction in Ethereum has several fields that consist of a recipient field, which holds the address of the recipient of the transaction, and if a transaction is used for contract creation, then the amount of this field would be zero. In Ethereum, transactions could fail due to several reasons, and considering this fact, the authors only consider successful transactions when they are inspecting smart contract creation.

The authors took that the complete transactions that have taken place from the launch of Ethereum on July 30th, 2015 to June 10th, 2017(namely 28,502,131 external transactions and 19,759,821 internal transactions).

4. Graph construction:

In this section, three graphs that the authors proposed will be defined.

4.1. **MGF.**

Definition 1. *MGF=(V, W, w), such that V is a set of nodes, and E is a set of edges. And w would be a function that maps the edges to their weights. $V = V_n \cup V_{sc}$, V_n is the set of EOAs(externally owned accounts) and the V_{sc} would be the set of smart of contracts. In this definition E is the set of ordered pairs of notes, $E = \{(v_i, v_j) \mid v_i, v_j \in E\}$. In this graph, the order of an edge represents the direction of the flow of the money. $w : E \rightarrow$. And the w maps each edge to a weight that is the whole amount of ether that is moved via that edge. Because of the graph structure, it could be stated that MGF is a directed and weighted graph.*

4.2. **CCG.**

Definition 2. *CCG=(V, W), such that V is a set of nodes and E is a set of edges. $E = \{(v_i, v_j) \mid v_i, v_j \in E\}$ in a way that every edge (v_i, v_j) corresponds to the creation of the smart contract v_j by the account v_i.*

4.3. **CIG.**

Definition 3. *CIG=(V, W), such that V is a set of nodes, and E is a set of edges. And w would be a function that maps the edges to their weights. In this graph E refers to an ordered pair of nodes, $E = \{(v_i, v_j) \mid v_i \in v, v_j \in v_{sc}\}$. Here an edge corresponds to an account v_i invoking the account v_j. $w : E \rightarrow$ maps each edge to a weight that is the amount invocations via an edge. Based on the graph's structure, it could be stated that CIG is a directed and weighted graph.*

5. Main Findings:

While constructing the three graphs that were mentioned earlier authors reached some insights:

- **Finding1:** They observed that in Ethereum, users prefer to transfer money rather than using the smart contract. They mentioned that this might be due to the fact that the smart contract is a relatively new concept, and the users might not be introduced properly to it considering the fact that we did not see smart contracts in Bitcoin where users might be more familiar with.
- **Finding2:** They found out that smart contracts are not prevalent. The reason for this also could be the fact there exists only a small number of applications supported by smart contracts and the fact that even among those small number of applications, only a few of them are related to the financial domain.
- **Finding3:** And they saw that all the users do not regularly use Ethereum. This might be because of the fact that a large number of users just intend to try Ethereum.
- **Finding4:** They observed that a few developers developed a large number of smart contracts. Furthermore, after the authors conducted an evaluation of all smart contracts, they saw that only a smart percentage of them are unique.
- **Finding5:** There understood that financial applications are dominant in Ethereum, even considering the fact in Etheruem different sorts of other applications are also allowed to operate.

6. Applications and Future research

6.1. **Applications:**

Attack Forensics: In attack forensics, when we are dealing with a malicious smart contract, our goal is to reach out to every account which is controlled by anyone who is attacking the network. In order to do that, the authors perform a correlation on CCG and CIG to get their hands on the list of every smart contract that is initiated by an attacker and, in addition to that, every account that would initiate such contracts later on. You could see their results in figure 8, where the malicious contract is the node BD37(the notation x:y here means that the contract y is created by node x)

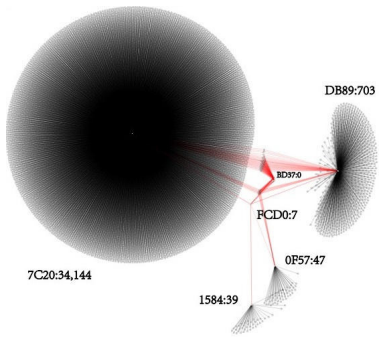

Fig. 8. Attack forensics of *BD37*. We randomly select 10,000 contracts created by *7C20* for the ease of illustration. No contracts created by *7C20* are invoked.

Anomaly Detection: In addition to attack forensics, the authors also introduce a new algorithm for identifying the sorts of contracts created in an abnormal way that their aim is the exhaustion of resources. The inputs of this algorithm are an account x, MFG, CCG, CIG, and three thresholds, and if the algorithm seeks an abnormal contract creation by after receiving those inputs, it will return true. You could see this algorithm in the figure below.

Algorithm 1 Detection of abnormal contract creation

Inputs: *x*, the detected account
MFG, money flow graph
CCG/CIG, contract creation/invocation graphs
T1,T2,T3, thresholds
Outputs: True/False, *x* is abnormal/benign

```
sc_set = created_sc(CCG, x);
if size(sc_set) < T1 return False;
for each node y in sc_set
    caller_set = inedge(CIG, y);
    for each edge z in caller_set
        num += z.weight;
    sender_set = inedge(MFG, y);
    for each edge s in sender_set
        value += s.weight;
if num > T2 × size(sc_set) || value > T3 × size(sc_set)
    return False;
else return True;
```

6.2. **Future Directions:** The authors state that in the future, they intend to conduct a more comprehensive study on Ethereum, and they want to assess how it is evolving based on some additional metrics, and they also intend to come up with more applications for Ethereum, such as detection of Dos attacks.

References

[1] Kehrli, Jerome. "Blockchain 2.0-from bitcoin transactions to smart contract applications." Niceideas, November. Available at: https://www. niceideas. ch/roller2/badtrash/entry/blockchain-2-0-frombitcoin (Accessed: 5 January 2018) (2016).

[2] Li, Xiaoqi, et al. "A survey on the security of blockchain systems." Future Generation Computer Systems 107 (2020): 841-853.

[3] Chen, Ting, et al. "An adaptive gas cost mechanism for ethereum to defend against under-priced dos attacks." International Conference on Information Security Practice and Experience. Springer, Cham, 2017.

[4] Chen, Ting, et al. "Under-optimized smart contracts devour your money." 2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER). IEEE, 2017.

[5] Reid, Fergal, and Martin Harrigan. "An analysis of anonymity in the bitcoin system." Security and privacy in social networks. Springer, New York, NY, 2013. 197-223.

[6] Meiklejohn, Sarah, et al. "A fistful of bitcoins: characterizing payments among men with no names." Proceedings of the 2013 conference on Internet measurement conference. 2013.

[7] Zhao, Chen, and Yong Guan. "A graph-based investigation of bitcoin transactions." IFIP International Conference on Digital Forensics. Springer, Cham, 2015.

[8] Maesa, Damiano Di Francesco, Andrea Marino, and Laura Ricci. "An analysis of the bitcoin users graph: inferring unusual behaviours." International Workshop on Complex Networks and their Applications. Springer, Cham, 2016.

[9] Buterin, Vitalik. "What is Ethereum?." Ethereum Official webpage. Available: http://www. ethdocs. org/en/latest/introduction/what-is-ethereum. html. Accessed 14 (2018).

[10] Wood, Gavin. "Ethereum: A secure decentralised generalised transaction ledger." Ethereum project yellow paper 151.2014 (2014): 1-32.

[11] Just, Winfried, Hannah Callender, and M. Drew LaMar. "Clustering coefficients." Department of Mathematics, Ohio University, Athens. Dosegljivo 22.3 (2015): 2017.

[12] Lee Rodgers, Joseph, and W. Alan Nicewander. "Thirteen ways to look at the correlation coefficient." The American Statistician 42.1 (1988): 59-66.

[13] Langville, Amy N., and Carl D. Meyer. "Deeper inside pagerank." Internet Mathematics 1.3 (2004): 335-380.

[14] Phanny, I. T. H. "Guideline for Interpreting Correlation Coefficient." Data & Analytics 5 (2014).

[15] Antonopoulos, Andreas M. Mastering Bitcoin: unlocking digital cryptocurrencies. " O'Reilly Media, Inc.", 2014.

[16] Bojja Venkatakrishnan, Shaileshh, Giulia Fanti, and Pramod Viswanath. "Dandelion: Redesigning the bitcoin network for anonymity." Proceedings of the ACM on Measurement and Analysis of Computing Systems 1.1 (2017): 1-34.

[17] Miller, Andrew, et al. "Discovering bitcoin's public topology and influential nodes." et al (2015).

[18] Decker, Christian, and Roger Wattenhofer. "Information propagation in the bitcoin network." IEEE P2P 2013 Proceedings. IEEE, 2013.

[19] Neudecker, Till, Philipp Andelfinger, and Hannes Hartenstein. "Timing analysis for inferring the topology of the bitcoin peer-to-peer network." 2016 Intl IEEE Conferences on Ubiquitous Intelligence Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld). IEEE, 2016.

[20] Biryukov, Alex, Dmitry Khovratovich, and Ivan Pustogarov. "Deanonymisation of clients in Bitcoin P2P network." Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. 2014.

[21] Donet, Joan Antoni Donet, Cristina Pérez-Sola, and Jordi Herrera-Joancomartí. "The bitcoin P2P network." International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg, 2014.

[22] Ranshous, Stephen, et al. "Exchange pattern mining in the bitcoin transaction directed hypergraph." International Conference on Financial Cryptography and Data Security. Springer, Cham, 2017.

[23] T. Chen et al., "Understanding Ethereum via Graph Analysis," IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, Honolulu, HI, 2018, pp. 1484-1492, doi: 10.1109/INFOCOM.2018.8486401.