Abstract—

Index Terms—

Analyzing the Ethereum Blockchain

Christin Sauerbier
Humboldt University Berlin
Berlin, Germany
christin.sauerbier@outlook.com

Roman Proskalovich
Humboldt University Berlin
Berlin, Germany
romanarion@gmail.com

Thomas Siskos
Humboldt University Berlin
Berlin, Germany
thomas.siskos@hu-berlin.de

I. Introduction

II. Literature Review

We aspire to adopt the approaches towards the study of blockchain described in [Lischke and Fabian, 2016]. The main difference of our research is that it targets ethereum instead of the bitcoin. For this reason we pay special attention to the papers focused on ethereum blockchain. Thus, [Payette et al., 2017] research ethereum address space; [Chan and Olmsted, 2017] and [Chen et al., 2018] perform ethereum graph analysis; [Li et al., 2017] develop a query layer for ethereum blockchain; [Anoaica and Levard, 2018] perform quantitative description of internal activity on the ethereum blockchain; [Somin et al., 2018] study the dynamics of the social signals of ethereum network, provide insights about the ecosystem and the forces acting within it and demonstrate that the network displays strong power-law properties.

We also go through the generic information about ethereum: including its white [Buterin et al., 2014]) and yellow ([Wood, 2014] papers), comparative studies of crypto-currencies (see for instance [Maesa, 2018], [Rudlang, 2017], [Sapuric et al., 2017], [Anderson et al., 2016]). Given the role that smart-contracts play in ethereum blockchain we also consider research on this topic. For instance: [Grishchenko et al., 2018] do semantic analysis of ethereum smart-contracts; [Tikhomirov et al., 2018] come up with tools for their static analysis; [Bartoletti et al., 2017] research practical applications of smart contracts and their design patterns; [Bartoletti et al., 2017] suggest ways to identify Ponzi schemes.

To complete the picture we also consider research on network analysis of bitcoin blockchain that were suggested in [Lischke and Fabian, 2016] and can provide some additional inspiration: [Reid and Harrigan, 2013], [Baumann et al., 2014], [Drainville, 2012], [Ober et al., 2013], [Meiklejohn et al., 2013], [Spagnuolo et al., 2014], [Androulaki et al., 2013], [Kaminsky, 2011], [Ortega, 2013].

III. Ethereum Theory

IV. Ethereum and the Blockchain

V. Methodology

A. Data Preparation

B. Graph Analysis

C. Descriptive Analytics

VI. Findings

A. Data Preparation

B. Graph Analysis

C. Descriptive Analytics

Descriptive analysis of the data allows getting insights that can be useful at the time of the network analysis. By itself alone it can point at important information with regard to the network structure, interactions between the nodes, underlying business processes, design problems, potential privacy and security issues.

For instance, while the median value of a transaction is equal to 0,9 ether, the value of the biggest one is 1 million ether (138 million U.S. dollars as of 15.03.2019). Figure 1 shows the distribution of value sent in individual transactions. There is a significant number of outliers indicating at certain concentration of ether holdings.

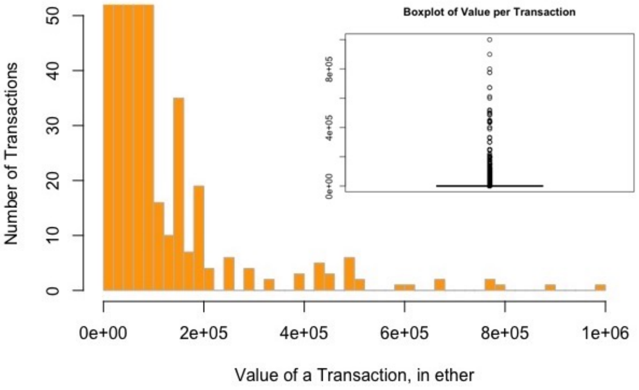


Fig. 1: Distribution of Value sent per Transaction.
**The number of transactions on the Y axis is limited to 50, so that the less frequent values remain visible.*

The amount of fees paid in transactions gives a hint at the underlying variation in businesses activity and Ethereum design issues. Cryptocurrencies are often promoted as a fast medium for micro-payments. However, fast transactions with low fees are not always the case. Limitations in the block size and the number of blocks in a period of time may result in an overflow of transactions pending to be included in a block. To speed up transactions people pay higher fees. At the same time, when the exchange rate of a cryptocurrecny rises, fees tend to become lower.

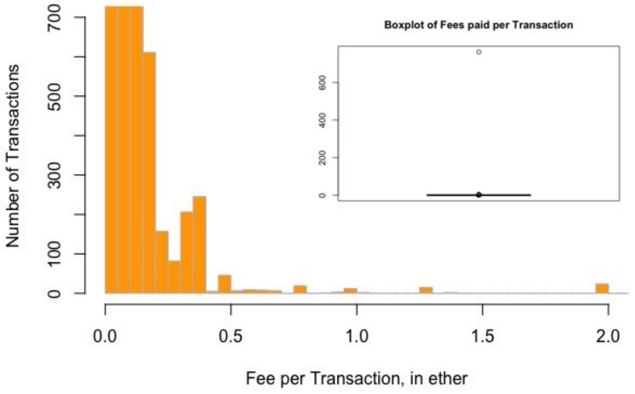


Fig. 2: Distribution of Fees paid per Transaction.
**The number of transactions on the Y axis is limited to 700, so that the less frequent fees remain visible.*

Interestingly, the is a big outlier - once there was a fee of more than 600 ether (80 000 U.S. dollars as of 15.03.2019). Apparently this is due to a mistake made by a payer. Such a mistake shows one of the disadvantages of the blockchain-based currencies where all transactions are irreversible.

Analysis of the distribution of gas used in transactions (see Figure 3) points at differences in their types. As gas measures how much "work" an action or a set of actions takes to perform, this potentially can be used to make a preliminary judgement about the prevalence of the smart contracts or about distribution of their complexity.

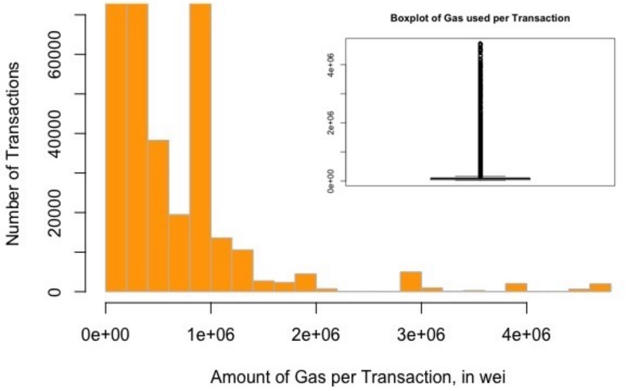


Fig. 3: Distribution of Gas used per Transaction.
**The number of transactions on the Y axis is limited to 70 000, so that the less frequent amount of gas remain visible.*

There are obviously two major clusters that probably separate regular payment transactions and smart contracts plus several groups of more complex contracts that however represent minority in the total volume of transactions. Both median and the 3rd quartile amount of gas is equal to 90 000 wei. In case the above hypothesis is correct, regular payment transactions will cover at least 75 % of the total number and are all represented on the Figure 3 by the 1st bar (cut in order to make other values visible).

Analysis of the transactions number over time makes the concern about unequal distribution of wealth in the network visible. As illustrated on the Figure 4 the number of transactions on the most active days more than doubles the respective number on the days with the lowest activity. Such significant variation indicates at the presence of major nodes.

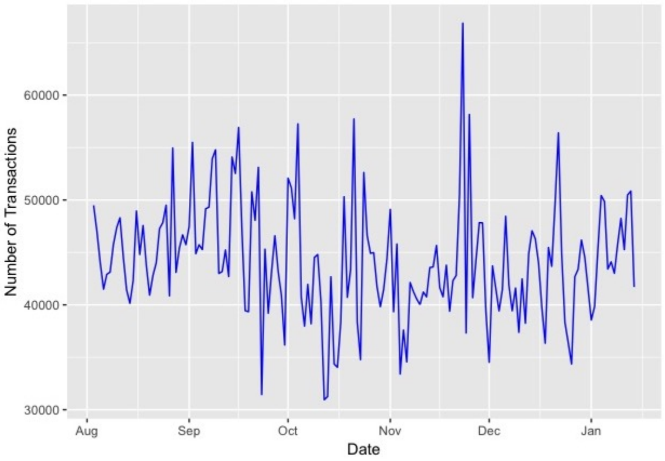


Fig. 4: Number of Transactions over Time.

This becomes even more visible if we consider the monetary volume of transactions over time (see Figure 5). Increases in volume of up to a hundred times from one day to another are an example of activity that would lead to a drastic changes in prices if these transactions are made not on chain but on currency exchanges.

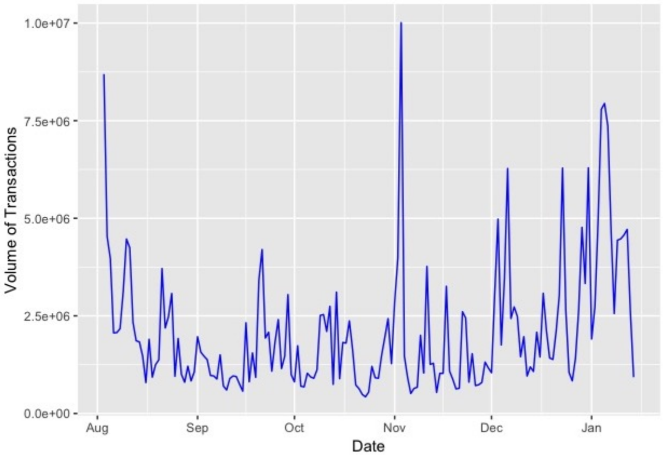
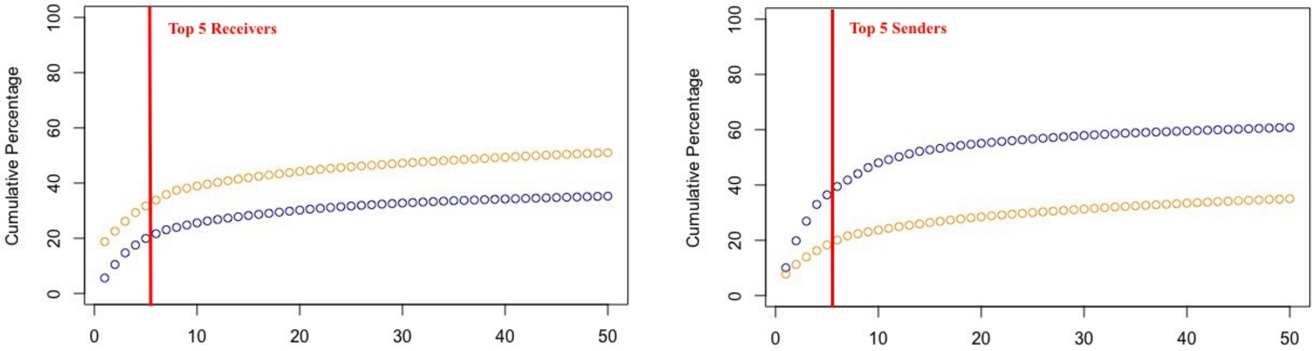


Fig. 5: Volume of Transactions over Time, in ether.



(a) Top 50 Receivers.

(b) Top 50 Senders.

Fig. 6: Cumulative Percentage of Transactions (blue) and Volume (orange) covered by Major Receivers (a) and Senders (b).

In order to verify the hypothesis about the presence of major nodes we analyse transactions number and volume more in depth. This is done by associating transactions with individual accounts. We treat receivers and senders of transactions separately.

As shown on the Figure 6a top 5 major accounts have received 19,9 % of all transactions (1,45 out of 7,3 million transactions). Top 50 accounts cover slightly more than one third of the total transactions' number. When it comes to the monetary value, the situation is even more centralized. Top 5 accounts have received 31,8 % of the total transactions' value (112 out of 352 million ether received in all transactions). Major 50 Receivers cover more than a half of the volume transacted within the network.

Figure 6b shows the data for major senders of transactions. Here, in terms of the transactions' number the situation is even more centralized. Top 5 accounts have made 36,4 % of all transactions. Top 50 senders cover 60,1 % of the entire transactions' number. At the same time in terms of volume the picture is different. Top 5 and Top 50 senders cover only 18,3 % and 35,0 % of the total transactions' value respectively.

Interestingly, there are significant discrepancies in the degree of centralization if the data about receivers and spenders is compared to each other. This is the case for centralization both in terms of transactions' number and volume. Analysis of these discrepancies may shade light on some of the roles that major nodes play in the network.

Discrepancies in the Transactions' Volume between Receivers and Senders. There is a significant difference between the amount of ether received and spent by the major accounts. While Top 50 Receivers have accumulated 172 million ether, Top 50 Senders have spent only 123 million. Besides, top receivers and senders should not necessarily be the same entities. Consequently there is a number of major nodes that spend less than their receive or don't spend at all. Figure 7 shows the net balances of the accounts within the studied period of time. While absolute majority of accounts have balance around 0, there are few that have accumulated a significant amount of wealth.

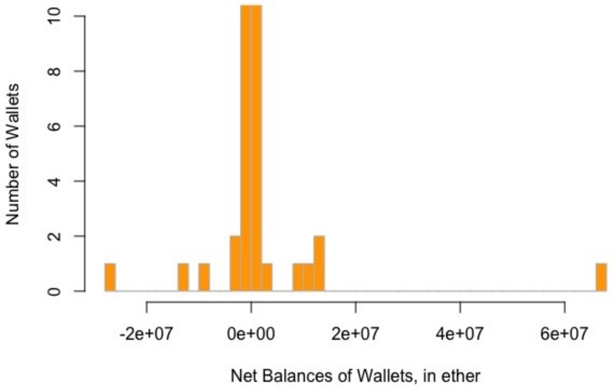


Fig. 7: Distribution of Balances within Studied Period.
**The number of wallets on the Y axis is limited to 10, so that the less frequent wallets remain visible.*

Those could be mining pools that receive new coins and hold them anticipating appreciation of ether. However the amount of Ethereum accumulated by such nodes seems to be too big to be explained this way. Despite a new block in Ethereum is mined every 10-20 seconds and remunerates miners with 5 ether, during the analyzed period of time this results in only approximately 5 million new ether.

Probably, these nodes are currency exchanges that accumulate wealth of their participants. They keep track of exchange transactions internally without recording them on-chain. The wealth is recorded as spent only when a participant makes a payment transaction.

If the nodes are actually currency exchanges, there is a deviation from the Ethereum narrative of ”no trusted third party involvement”. Given the level of centralization associated with the nodes, this could imply a major source of vulnerability. An attack on the exchange records would lead to losses for a significant number of the exchange participants.

In order to verify how relevant this risk actually is, we conducted further investigation of the accounts that received the biggest volumes of ether. Using data of the Ethereum scanners such as Etherscan.io, Etherchain.org, Blockchair.com,

TABLE I: Top 5 Major Receivers (Volume)

Wallet	Amount Received, mill. ether	Type	Affiliation
0xAA1A6e3e6EF20068f7F8d8C835d2D22fd5116444	66,0	Smart-Contract	ReplaySafeSplit
0xBFC39b6F805a9E40E77291afF27aeE3C96915BDD	13,2	Smart-Contract	Poloniex
0x209c4784AB1E8183Cf58cA33cb740efbF3FC18EF	12,6	Smart-Contract	Poloniex 2
0x7727E5113D1d161373623e5f49FD568B4F543a9E	11,1	Smart-Contract	Bitfinex 2
0xFa52274DD61E1643d2205169732f29114BC240b3	8,5	Smart-Contract	Kraken 4

TABLE II: Top 5 Major Senders (Number of TXs)

Wallet	Number of TXs send	Type	Affiliation
0xEA674fdDe714fd979de3EdF0F56AA9716B898ec8	733818	Address	Ethermine
0x2a65Aca4D5fC5B5C859090a6c34d164135398226	715017	Address	DwarfPool 1
0xD34DA389374CAAD1A048FBDC4569AAE33fD5a375	516932	Address	GenesisMining
0x52bc44d5378309EE2abF1539BF71dE1b7d7bE3b5	443544	Address	Nanopool
0x61C808D82A3Ac53231750daDc13c777b59310bD9	251566	Address	F2Pool 1

Bloxy.info, etc.[1] we found out affiliation of the Top 5 Receivers (in terms of volume).

Four out of five Top Receivers actually belong to the currency exchanges (see Table I). However, all these accounts in reality appeared to be not regular addresses but smart contracts. The remaining Top Receiver (the first one in the list) is also a smart contract. None of them is supposed to accumulate wealth and by design should transfer it further. This puts under question the validity or completeness of the transactions data we have collected.

Deeper research allowed restoring the faith in the dataset. Appeared that money received by identified major accounts was actually spent. Bit in another network.

The point is that on the 20th of July 2016 (shortly before the analyzed period) Ethereum experienced a hard fork. As a consequence of the DAO incident - a hack of a complicated smart-contract that resulted in a loss of about 12.7 million ether (worth around 150 million U.S. dollars at the time) - the community was split in two parts. Most of the network participants decided to make a fork chain where the stolen coins would be returned to their owners. However, some of the community decided to continue the old chain (Ethereum Classic or ETC), arguing that ”the code is the law” and the blockchain data should be irreversible[2].

This lead to a number of issues related to the interaction between chains. One of the most relevant of them was ”replay attack”. The mechanics of it is as follows: if there is a valid transaction on one chain, it can also be offered - replayed - on the other chain to duplicate the received amount. For instance, a user of the currency exchange can deposit and withdraw her money from this exchange on the old chain, and then use the 2nd transaction to withdraw money also on the new (forked) chain[3]. This concern was solved through inter-mediation of the smart contracts. Thus, the post by Ethereum founder Vitalik Buterin states: ”users who are interested in taking any actions with their ETC, including creating and participating in applications, converting to another asset, etc. are advised to use the splitter contract at address 0xAA1A6e3e6EF20068f7F8d8C835d2D22fd5116444 to move their ETC to a separate newly created account so as to avoid replay attacks; we also encourage the ETC community to consider adopting a secondary hard fork to change transaction formats to make further replay attacks impossible[4].

The address mentioned in the post is the biggest Receiver in our dataset[5]. It has transmitted all the received money to its senders, but on the other chain. The other 4 accounts, which are intermediate exchange wallets were apparently performing similar function of splitting wealth between different chains.

Thus, our concern about possible attack is confirmed only partially. None of the Top 5 Receivers were accumulating wealth and it all was automatically transmitted to other accounts. At the same time while exploring the ”ReplaySafeSplit” smart-contract we have found that a number of users lost their money due to the mistakes made by them during the split[6]. The smart-contract itself appeared to be susceptible to a number of bugs[7] and was later substituted by another one. The DAO incident described in the context of our Receivers exploration also exemplifies how the risk of the attack on an account or contract that holds funds from different users can and has actually played out.

Discrepancies in the Transactions' Number between Receivers and Senders. As has been described by the Figure 6 the number of transactions received by major Receivers is significantly smaller than the number of transactions made by major Senders. This discrepancy can probably be partially explained by the activity of the mining pools. There, one

[1] For details see:
https://etherscan.io/,
https://www.etherchain.org/,
https://blockchair.com/Ethereum/,
https://bloxy.info/.

[2] See for instance https://www.cryptocompare.com/coins/guides/the-dao-the-hack-the-soft-fork-and-the-hard-fork/.

[3] See for instance https://vessenes.com/hard-fork-and-replay-concerns/.

[4] For details see https://blog.Ethereum.org/2016/07/26/onward_from_the_hard_fork/.

[5] The original address in the post was substituted as the original smart-contract was changed.

[6] See for instance https://medium.com/@chevdor/safer-version-of-the-replaysafesplit-smart-contract-a29c347e8a7.

[7] For details see https://etherscan.io/address/0xaa1a6e3e6ef20068f7f8d8c835d2d22fd5116444#code.

account receives a mining reward and then sends it in small fractions to the many participants of the pool.

Alternative explanation can be related to various approaches of the network participants to gain more privacy. One of them is the so called ”peeling-chain”. The technique consists in dividing and sending the wealth of an account to multiple addresses again and again. The goal is to create an impression that several users are doing transactions instead of one.

The process is illustrated on the Figure 8. Here, the received amount of 50 coins is splitted in several iterations to multiple accounts in an attempt to complicate tracking of a certain persons wealth.

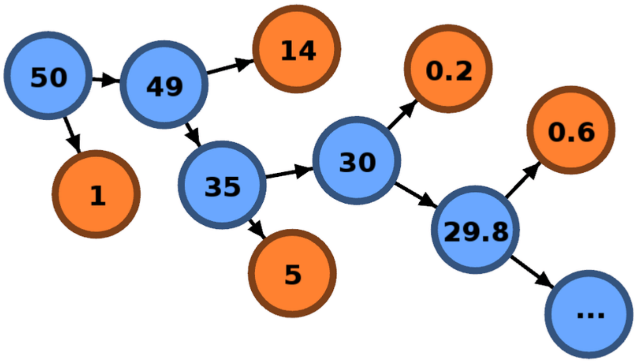


Fig. 8: How the peeling-chain works.
From [de Balthasar and Hernandez-Castro, 2017]

Further research of the Top 5 transaction Senders allowed verifying the hypotheses about the role of different nodes and their activities. Using data of the Ethereum scanners as before we disclosed identities of the major accounts. All of them appeared to be mining pools (see Table II. Among Top 5 Receivers of transactions there are no mining pools (the accounts are a digital token YoCoin, 2 Poloniex exchange wallets, currently closed by the U.S. government cryptocurrency trading platform BTC-e and already mentioned ReplaySafeSplit smart-contract). Thus the hypothesis about the activity of mining pools is confirmed. ”Peeling-chain” practices are neither confirmed nor disproved.

VII. Conclusion

References

[Anderson et al., 2016] Anderson, L., Holz, R., Ponomarev, A., Rimba, P., and Weber, I. (2016). New kids on the block: an analysis of modern blockchains. *arXiv preprint arXiv:1606.06530*.

[Androulaki et al., 2013] Androulaki, E., Karame, G. O., Roeschlin, M., Scherer, T., and Capkun, S. (2013). Evaluating user privacy in bitcoin. In *International Conference on Financial Cryptography and Data Security*, pages 34–51. Springer.

[Anoaica and Levard, 2018] Anoaica, A. and Levard, H. (2018). Quantitative description of internal activity on the ethereum public blockchain. In *New Technologies, Mobility and Security (NTMS), 2018 9th IFIP International Conference on*, pages 1–5. IEEE.

[Bartoletti et al., 2017] Bartoletti, M., Carta, S., Cimoli, T., and Saia, R. (2017). Dissecting ponzi schemes on ethereum: identification, analysis, and impact. *arXiv preprint arXiv:1703.03779*.

[Baumann et al., 2014] Baumann, A., Fabian, B., and Lischke, M. (2014). Exploring the bitcoin network. volume 1.

[Buterin et al., 2014] Buterin, V. et al. (2014). A next-generation smart contract and decentralized application platform. *white paper*.

[Chan and Olmsted, 2017] Chan, W. and Olmsted, A. (2017). Ethereum transaction graph analysis. In *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 498–500. IEEE.

[Chen et al., 2018] Chen, T., Zhu, Y., Li, Z., Chen, J., Li, X., Luo, X., Lin, X., and Zhange, X. (2018). Understanding ethereum via graph analysis. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pages 1484–1492. IEEE.

[de Balthasar and Hernandez-Castro, 2017] de Balthasar, T. and Hernandez-Castro, J. (2017). An analysis of bitcoin laundry services.

[Drainville, 2012] Drainville, D. (2012). An analysis of the bitcoin electronic cash system. *Waterloo, Canada: University of Waterloo*.

[Grishchenko et al., 2018] Grishchenko, I., Maffei, M., and Schneidewind, C. (2018). A semantic framework for the security analysis of ethereum smart contracts. In *International Conference on Principles of Security and Trust*, pages 243–269. Springer.

[Kaminsky, 2011] Kaminsky, D. (2011). Black ops of tcp/ip. *Black Hat USA*, page 44.

[Li et al., 2017] Li, Y., Zheng, K., Yan, Y., Liu, Q., and Zhou, X. (2017). Etherql: a query layer for blockchain system. In *International Conference on Database Systems for Advanced Applications*, pages 556–567. Springer.

[Lischke and Fabian, 2016] Lischke, M. and Fabian, B. (2016). Analyzing the bitcoin network: The first four years. *Future Internet*, 8(1):7.

[Maesa, 2018] Maesa, D. D. F. (2018). Blockchain applications: Bitcoin and beyond.

[Meiklejohn et al., 2013] Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G. M., and Savage, S. (2013). A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 127–140. ACM.

[Ober et al., 2013] Ober, M., Katzenbeisser, S., and Hamacher, K. (2013). Structure and anonymity of the bitcoin transaction graph. *Future internet*, 5(2):237–250.

[Ortega, 2013] Ortega, M. S. (2013). *The bitcoin transaction graphanonymity*. PhD thesis, Masters thesis, Universitat Oberta de Catalunya.

[Payette et al., 2017] Payette, J., Schwager, S., and Murphy, J. (2017). Characterizing the ethereum address space.

[Reid and Harrigan, 2013] Reid, F. and Harrigan, M. (2013). An analysis of anonymity in the bitcoin system. In *Security and privacy in social networks*, pages 197–223. Springer.

[Rudlang, 2017] Rudlang, M. (2017). Comparative analysis of bitcoin and ethereum. Master's thesis, NTNU.

[Sapuric et al., 2017] Sapuric, S., Kokkinaki, A., and Georgiou, I. (2017). In which distributed ledger do we trust? a comparative analysis of cryptocurrencies. *MCIS 2017 Proceedings*.

[Somin et al., 2018] Somin, S., Gordon, G., and Altshuler, Y. (2018). Social signals in the ethereum trading network. *arXiv preprint arXiv:1805.12097*.

[Spagnuolo et al., 2014] Spagnuolo, M., Maggi, F., and Zanero, S. (2014). Bitiodine: Extracting intelligence from the bitcoin network. In *International Conference on Financial Cryptography and Data Security*, pages 457–468. Springer.

[Tikhomirov et al., 2018] Tikhomirov, S., Voskresenskaya, E., Ivanitskiy, I., Takhaviev, R., Marchenko, E., and Alexandrov, Y. (2018). Smartcheck: Static analysis of ethereum smart contracts.

[Wood, 2014] Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, 151:1–32.