

DOI: 10.37943/AITU.2020.36.57.005

A. Seitenov

MSc, Senior lecturer of ICT Educational Program
altynbek.seitenov@astanait.edu.kz, orcid.org/0000-0001-5777-4363
Astana IT University, Kazakhstan

G. Smagulova

Master of technics and technology, Senior lecturer of ICT Educational Program
gulnur.smagulova@astanait.edu.kz, orcid.org/0000-0002-2878-5597
Astana IT University, Kazakhstan

DISTRIBUTION OF ETHEREUM BLOCKCHAIN ADDRESSES

Abstract: In the contemporary world, Ethereum is a very reliable financial saving among cryptocurrencies. It is also well known as a blockchain platform for creating and launching its own cryptocurrency. The applications run on Ethereum executed by a platform-specific cryptographic token, ether. During 2014, Ethereum had launched a pre-sale for ether, which had received an overwhelming response. Ether is used broadly for two purposes: it is traded as a digital currency exchange like other cryptocurrencies, and it is used inside Ethereum to run applications and even to monetize work. It should be noticed that the smart contract has brought a significant share of the success to Ethereum. The smart contract is a computer programme that independently performs assigned tasks between network participants without the participation of a third trusted party. Smart contracts and their intranet transactions have facilitated the rapid expansion of the Ethereum network. Smart contracts are widely represented on the market, either as electronic transaction payments or as applications for the implementation of logistics supplies, gambling, and other sectors. These transactions are irreversible and fully tracked online. Whereas the electronic records are available in a public distributed ledger and include data about user addresses, whereas the real names are hidden. The article explains the usage of decentralized accounts and their electronic transactions in the Ethereum network. The results are presented through different application sectors. Additionally, a new method for extracting blockchain records through node cluster via IPFS path is implemented in the research.

Keywords: classification of smart contracts, exporting blockchain transactions, Ethereum, blockchain technology, blockchain data mining, Ethereum data analysis, Erc20 token, Ethereum categories, solidity language, Infura nodes.

Introduction

According to CoinMarketCap, there are more than 2000 cryptocurrencies are traded on various digital exchanges [1]. In reality, Bitcoin has got more than half the capitalization of the crypto market (ibis). It can be exchanged for a wide range of goods and services in the financial market [2].

Ethereum takes second place in the market capitalization of blockchain technology [1]. Although Bitcoin and Ethereum are represented as cryptocurrencies while Ethereum possesses a significant opportunity in comparison with Bitcoin. One of them is a smart contract that works through an Ethereum development platform with a wide variety of applications [3].

Applications can be deployed in an exchange system, social and government sector, as well as a part of the supply chain and food industry. Dharma is an outstanding example of the use of a smart contract allowing to organize crypto-asset loans to private individuals [4]. Another example, IDEX is a decentralized exchange that provides funds for trading various tokens in real-time [5].

Since the launch of the first Ethereum block, its blockchain network has been steadily expanding. This fact is required notable attention to digital technologies based on smart contracts. The interest in studying the understanding of Ethereum system behaviour is concentrated in this report. Measurement results from electronic transaction analysis could help to better understand the development trend of the blockchain industry for the near future [6-8].

Blockchain

Blockchain is an innovative data structure that works as a digital ledger that contains information about transactions. It distributes a set of transactions among network members who work together to maintain a network [9]. Data is synchronized through the network and cannot be modified [10-11]. Digital blocks ensure that the received transaction is immutable and announced in the form it was created [12].

A digital signature using a cryptographic encryption algorithm signs each entry. The cryptographic encryption is a combination of the private key and the corresponding public key [12]. A transaction is performed between two or more participants who have a unique digital address that identifies it [11]. A newly arrived transaction is recorded in the global chain only after the verification process. Confirmation of the transaction is achieved by the agreement of the majority of nodes and is updated by miners. This process allows maintaining the blockchain network by transaction approval and registration [13].

In most cases, the blockchain platforms have the same design with the same functions. They are expanded by linking the generated block broadcast through a network of nodes. Each node has an updated real-time copy of a linearly constructed chain [14].

Ethereum environment

This work is related to the Ethereum Blockchain, launched in 2013 [15-16]. According to the creators of the platform, Ethereum is a universal IT solution for blockchain technologies, where software protocols are executed by using a smart contract. This self-executing computer application provides the opportunity to create cryptographic coins [12]. All programme's launches are carried out on the Ethereum Virtual Machine (EVM), and Ether is used to execute the operation process [17]. Ether (ETH in a short way) is a token unit for exchanging goods or representing various assets [18].

A set of rules called ERCs, have been issued to regulate the riskless and proper development of blockchain applications [19]. For example, the ERC20 token standard provides basic functionality for interchangeable tokens [8]. The developed standards provide programs with the ability to manage and carry out efficiently and safely transactions within the network [20]. In addition to the ERC20, there are other standards such as ERC621, ERC223, ERC667 and ERC777 [8].

Related approaches

In particular, there are various approaches to data mining from the blockchain for different platforms. The most familiar way to get data from Ethereum blockchain is to use the Geth client implemented in Go programming language [18]. The Geth client synchronizes data from the first block created on the network to the current time, as shown in Figure 1 [21]. An

alternative way to Geth command line is to handle the graphical interface implemented in Mist program [17].

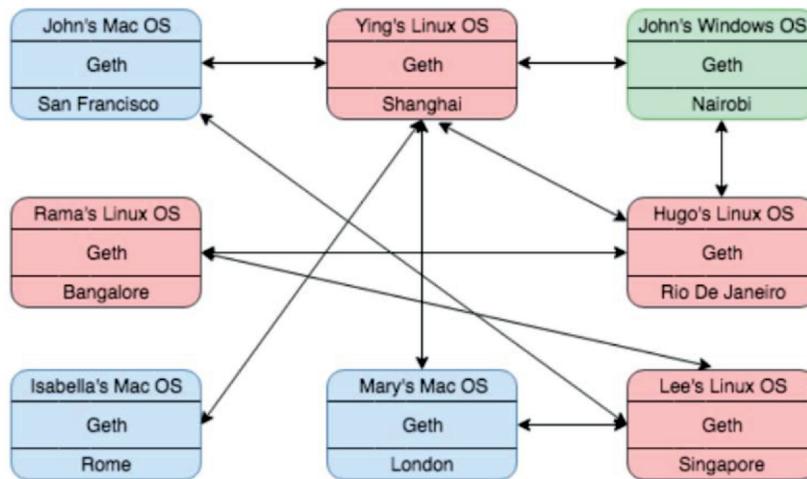


Fig. 1. Connection to Ethereum via Geth

The next way to access data is to manage an Ethereum explorer such as Etherscan. The explorer was used by Bartoletti and Pompianou for an empirical analysis of smart contracts. Smart contracts were analysed manually through Etherscan and classified into different categories [7]. In addition, there is a paid service that allows sending requests to receive responses from a database in the Google BigQuery system [22]. These methods cover all Ethereum database. [23-24]. The research considers a method that differs from the noticed programs. It enables to analyse electronic records without importing all data in the user computer. The method is described in the next chapter.

Data selection and collection

The section provides details on the data received approach and a technical process for exporting records from the network.

To export data from the Ethereum blockchain, access has been organized to one of the clusters of nodes. Infura, the selected cluster, allows the programmer to connect the user interface with external smart contracts (Figure 2) [25]. The method admits solving problems in the mode of full node synchronization via web service as Amazon Web Service (AWS), which requires a decent level of requirements in processing power and memory. Infura infrastructure provides stable and reliable access to Ethereum via IPFS path [26].



Fig. 2. Exporting raw data from Ethereum

We exercise 7 days period to study and analyse Ethereum transactions. The collection of digital recordings is retrieved from June 21, 2019, to June 27, 2019. 40 thousand blocks (blocks of 8,000,000 – 8,040,000) are extracted to satisfy the data requirements. The total number of records in the block is 5,083,421 rows. Additionally, 968,475 unique recipient addresses are identified. It should be noted that the weekly average volume is almost 4,550

thousand transactional transactions in 2019 [24].

During the selected period, network activity was higher than compared with the average 7-day productivity. Moreover, the average daily price of the selected ether exceeded \$309, while the average cost reached \$186 [24]. This fact can serve as serious evidence of forecasting high activity in the blockchain and interest among investors and developers.

As mentioned earlier, Ethereum blocks are loaded with a specific range. Electronic transactions (a unique transaction hash) are extracted from each block in Hexbytes format [26]. Then the records are converted to a string type for suitable manipulation and analysis. The receipt of the record contains a sender and recipient address, transaction fee committed for the operation. Next, we classify the data by transaction addresses and accumulate by the value of the transaction, as well as how many times the recipient/sender addresses participated in a given time interval. Finally, the results are sorted in ascending order to determine the most popular addresses according to various factors in this period.

Record results

It should be noted that the total amount of exported raw data from the network is 1.35 GB. The extracted raw data from the Ethereum blockchain is divided into several files for further convenient manipulation. The first type of file contains a transaction list of the corresponding blocks. The other includes receipts of each peer transaction.

The extracted transactions from the blockchain are grouped by address. Further, the addresses are sorted by the number of participants and accumulated by transaction value. The article analyses the top 50 popular addresses for each result. The classification is based on manually scan each address through Etherscan and related forums while the source code of each smart contract has been investigated.

The first result displays the top 50 popular addresses for their participation in Ethereum transactions over a research period (See goo.gl/FXaSQ). Observed addresses participates in 1,606,543 transactions (31.6% of total). 41 accounts (out of 50) are smart contracts. Only 1 address is associated with the N/A group. Hence, More Gold Coin (MGC) token is the most attractive contract. It participates in 339,812 transactions (21.2%) and is shown in Table 1. MGC is a storage payment service platform with a digital asset wallet [27].

Table 1: Top 10 addresses by the total number of executions

Nº	Address	Total value (in Ether)	Number of transactions	Type	Name	Category
1	0x174BfA6600Bf90...	0,00	339 812	Contract	More Gold Coin (MGC)	Token ERC-20
2	0xdAC17F958D2ee5...	0,00	103 551	Contract	Tether USD (USDT)	Token ERC-20
3	0x2a0c0DBEcC7E4...	23237,81	71 265	Contract	IDEX	Exchange
4	0xd3EBdAea9AEac...	0,00	65 315	Contract	SCF	Token ERC-20
5	0x21aB6c9fAC80C5...	0,00	61 080	Contract	Baer Chain	Token ERC-20
6	0x686B30a8082634...	0,00	54 813	Contract	Healthy Happy Harmony	Token ERC-20
7	0xF4a2eFf88a408ff4...	327443,78	53 304	Address	N/A	N/A
8	0x7090a6e22c83846...	10,27	51 401	Contract	Zuck Bucks	Token ERC-20
9	0x773301B28178f56...	3,77	50 891	Contract	Block.Network.DAO	Token ERC-20
10	0x8E766F57F7d16C...	0,00	41 797	Contract	Maximine Coin	Token ERC-20

The histogram illustrates the taxonomy of the 50 best addresses that determine membership in a particular group (Figure 3). The first result shows that standard ERC-20 tokens is the

largest group in the diagram (28 contracts) and conducts more than 1.09 million transactions. 5 out of 28 contracts are marked as a stable coin, and they involve in 180,882 records. There are Tether USD (USDT), USD Coin (USDC), TrueUSD (TUSD), Dai Stablecoin (DAI) and Paxos Standard. 6 addresses are connected with exchanges of exchanges like IDEX, Binance, Make, Ox, EtherDelta and Upbit.

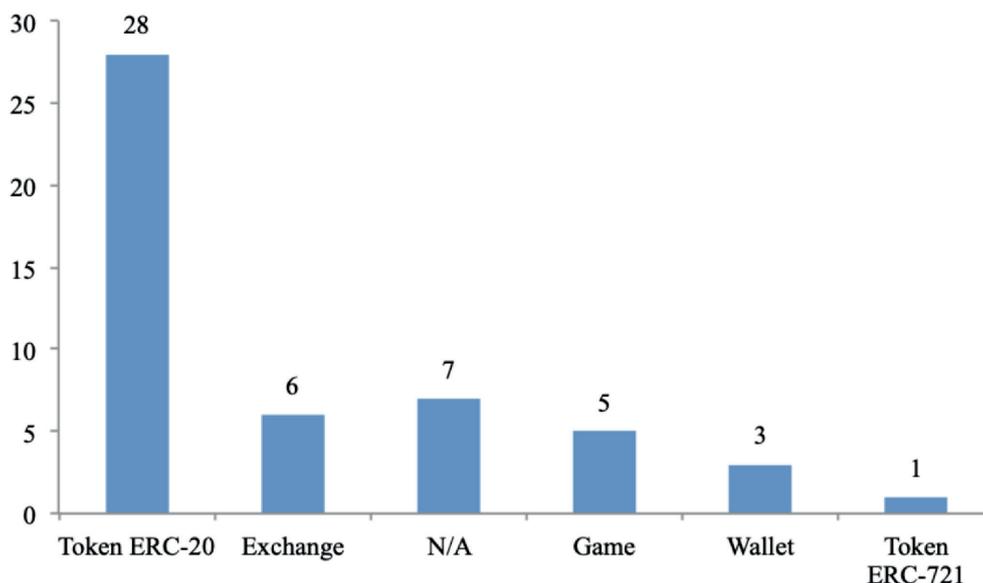


Fig. 3. Address categories by transaction number

It should be mentioned that the gaming category contains five smart contracts relating to the game (2 contracts) and the gambling industry. There are three smart contracts marked as gambling: Dice2Win, Arina Land Tycoon, and FCK.com. The next category is a wallet and includes three addresses: Cloud Token, Cryptonator, and BitBank. ERC-721 Token is one of the interesting groups among others. The address belonged to CryptoKitties. It is noted that the ERC-721 token is a unique asset that is wholly owned by the holder and cannot be played and picked up [28]. The ERC-721 standard includes basic functions for tracking non-replaceable tokens [29].

The second result shows the 50 most popular transactional addresses for the same week (See goo.gl/FRsDE). The total amount of transactions at 50 leading addresses exceeds 7 683 thousand Ether. It is 41.9% share while over this period the total turnover of coins amounted to 18 329 thousand ether. Binance is the leader of the rank with a turnover of 796 843 Ether (10.4% of the total number of addresses of the top 50 addresses). According to CoinMarketCap, Binance is a leader among cryptocurrency exchanges [1].

The diagram below shows the various categories related to the top 50 addresses by the total transmitted value (Figure 4). As indicated in the table of upper addresses, we see that there are 5 different categories: Exchange, Wallet, Game, N/A and ERC-20 token. In this case, the result shows that Binance and Bitfinex are awarded as the most popular among exchanges, there are 6 and 4 addresses, respectively. There are CoolTlon Token based on ERC-20 standard, Last Winner related to Game category as well as Kyber Network is associated with the innovative wallet. Kyber Network is a decentralized platform that acts as a wallet and payment system [30].

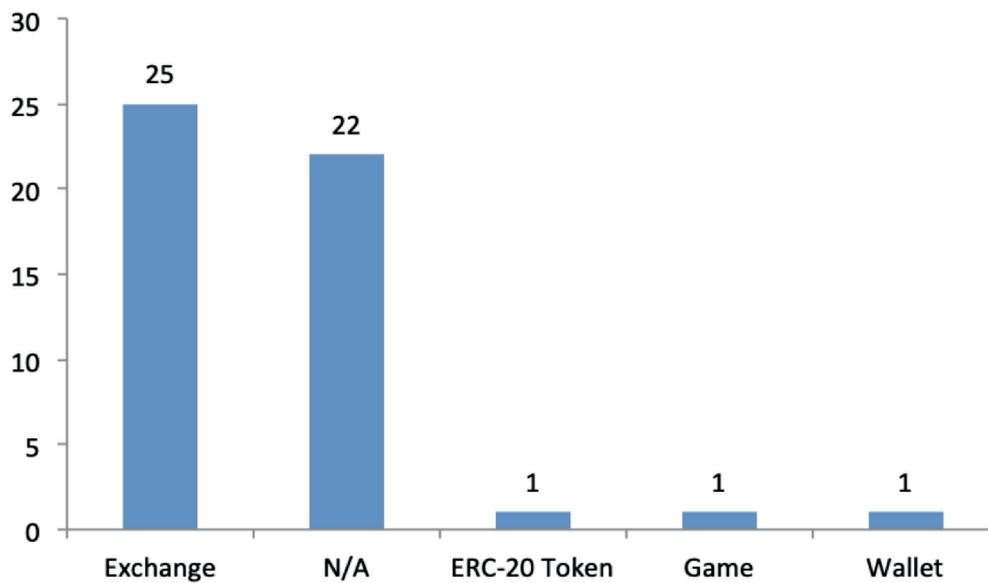


Fig. 4. Address categories by accumulated transaction value

From evaluating both taxonomy results for transactions, we can find that they contain slightly different categories. The first result consists of 6 categories (ERC-20 token, ERC-721 token, N/A, Exchange, Game and Wallet), and in the second experiment, there are 5 categories (exchange, N/A, ERC-20 token, Wallet and Game). Moreover, if the addresses of the ERC-20 tokens are the most popular in the first experiment, the second test shows that only one view (CoolTion) is in the top 50 list.

At the same time, records belonging to the Exchange group are often observed in both experiments. In particular, there are 13% of the main addresses and 57% of the transmitted value from the leading addresses. However, in both experiments, only Binance scores are compared. However, the N / A category is one of the dominant groups. It is not possible to pinpoint that anonymous account for a special group, even if they are very active in Ethereum.

On the other hand, the result of the first experiment shows that the distribution of addresses belonging to ERC-20 tokens dominates in categories by transaction number. In particular, 28 of the top 50 addresses are classified as ERC-20 tokens, and they participate in 1,091,428 payments of 1,606,543 entries from the top 50 addresses. Just as in the research paper of Victor and Luders, it is also implied that more than half of the transactions are related to ERC-20 tokens [8].

Result observation

Finally, from an analysis of the project's experiments and the results of the research of Bartoletti and Pompianu, as well as differences in the taxonomy of Ethereum transactions have found similarities [7]. The financial sector was the most common group in both outcomes (Financial is 46% and Token ERC-20 is 56%). Particularly, the project results have shown that the total number of transactions with the ERC-20 token, ERC-721 and the Exchange groups occupied a significant dominant share (70%) The compassion outcome results in Figure 5 [7].

On the contrary, the researchers did not mention exactly the exchange category distribution in their work. Moreover, it has been identified that it was the Exchange group that dominated both project experiments: the classification of Ethereum addresses by transaction number and total transaction cost. The game section could also be mentioned as an observable category with indicators of 10% and 19.4% from various research works. The next category is a wallet

and stands at 7% in our analyze and thus the difference between the Bartoletti and Pompianu analyzes is almost 4%. In unclassified category, the difference in two outcomes is just over five percent (Bartoletti and Pompianu – 19% and research results – 14%).

Besides, the article notes that one of the most commonly used categories was the financial and Token ERC-721, and this group was not defined in our experiments for a certain period [7]. However, we investigated tokens based on ERC standards using Etherscan [23].

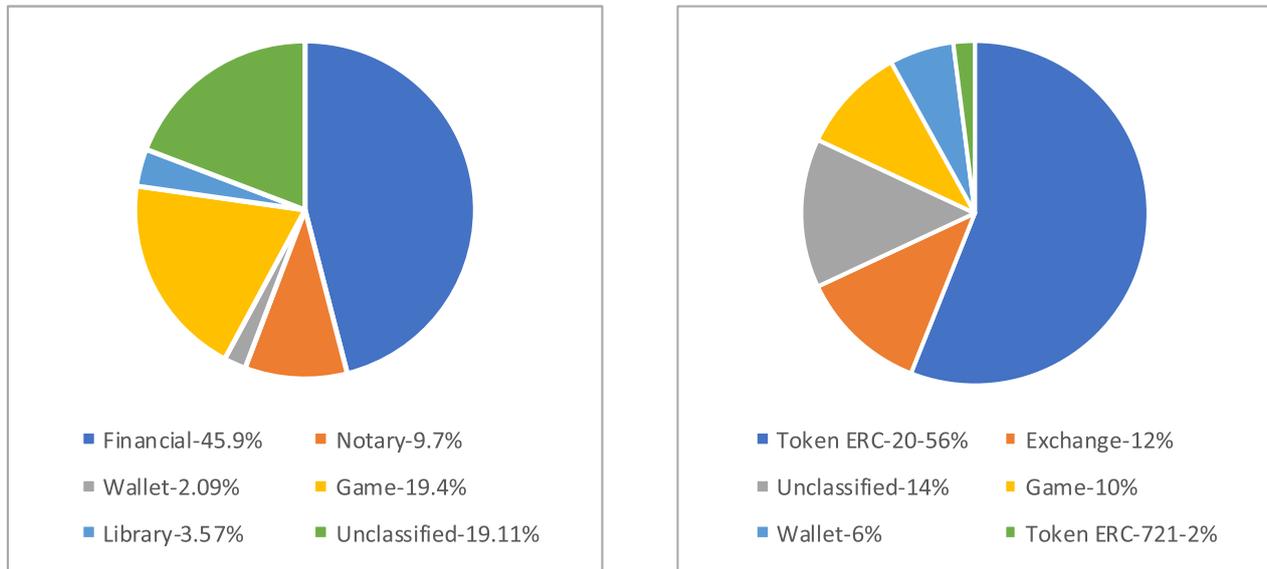


Fig. 5. Observation of two analyses.

Note that in some cases there are not enough elements to distribute smart contracts. This happens e.g. when the contract does not link to the project webpage, and there are neither comments in online forums nor in the contract sources.

Conclusion

This research paper has demonstrated the observational analysis of the electronic transactions in Ethereum blockchain. We gave a result overview for the activity of the network participants and the propagation of previous transactions. Therefore, this led to a study of the current situation in the Ethereum blockchain and the implementation of an analysis process to evaluate the transactions in various approaches. The first evaluation method illustrates the result, evaluated on the same principle. The second approach showed that the estimate depends on the transmitted Ether value. Raw data analysis results are compared with previous studies. After a review of existing publications identified with the taxonomy of smart contracts, we have found a gap in the classification of accounts and electronic payments over the past period. The previous work was related to the smart contract identification template and influenced the ERC-20 token on the network [7-8]. Meanwhile, the main goals and contributions were the development and analysis of the implementation process for the classification of the addresses of Ethereum.

The foundation for further research is able to be a complete classification of the blockchain addresses instead of the top 50 popular addresses on the Ethereum network. The following analysis model can be updated by evaluating the addresses sorted by the transaction fee. To sum up, potential proposals can be included in the continuation of this research work, and it can also be the basis for launching an innovative project with a relevant topic.

References

1. CoinMarketCap (n.d.). [Online] Available: <https://coinmarketcap.com>
2. Barber, S., Boyen, X., Shi, E., & Uzun, E. (2012, February). Bitter to better—how to make bitcoin a better currency. In *International conference on financial cryptography and data security* (pp. 399-414). Springer, Berlin, Heidelberg.
3. Salman, A., & Razaq, M.G.A. (2018). Bitcoin and the world of digital currencies. *Financial Management from an Emerging Market Perspective*, 271-281.
4. Dharma.io (n.d.). [Online]. Available: <https://www.dharma.io>
5. IDEX (n.d.). [Online]. Available: <https://idex.market/eth/idex>
6. Bartoletti, M., Carta, S., Cimoli, T., & Saia, R. (2017). Dissecting Ponzi schemes on Ethereum: identification, analysis, and impact. *Future Generation Computer Systems*, 102, 259-277. *arXiv preprint arXiv:1703.03779*.
7. Bartoletti, M., & Pompianu, L. (2017). An empirical analysis of smart contracts: platforms, applications, and design patterns. In *International conference on financial cryptography and data security* (pp. 494-509). Springer, Cham. *arXiv.org*, 10323, p.633.
8. Victor, F., & Luders, B. K. (2019). *Measuring Ethereum-based ERC20 Token Networks*. In press.
9. International Finance Corporation (2019). *Blockchain: Opportunities for Private Enterprises in Emerging Markets*, Second Edition.
10. Natarajan, H., Krause, S.K., & Gradstein, H.L. (2017). *Distributed Ledger Technology and blockchain*. 1. FinTech Note.
11. McCorry, P. (2018). *Applications of the Blockchain using Cryptography* (Doctoral dissertation, Newcastle University).
12. Crosby, M., Pattanayak, P., Verma, S., & Kalyanaraman, V. (2016). Blockchain technology: Beyond bitcoin. *Applied Innovation*, 2 (6-10), p.71.
13. Drescher, D. (2017). *Blockchain basics*. Berkeley, CA: Apress.
14. Nofer, M., Gomber, P., Hinz, O., & Schiereck, D. (2017). Blockchain. *Business & Information Systems Engineering*, 59 (3), pp. 183–187.
15. Buterin, V. (2013). A next-generation smart contract and decentralized application platform. <https://github.com/ethereum/wiki/wiki/White-Paper>
16. Buterin, V. (2014). A next-generation smart contract and decentralized application platform. *White Paper*, 3, p.37.
17. Kehrl, J. (2016). Blockchain 2.0—from bitcoin transactions to smart contract applications. *Niceideas*. [Online]. Available: <https://www.niceideas.ch/roller2/badtrash/entry/blockchain-2-0-frombitcoin>.
18. Dannen, C. (2017). *Introducing Ethereum and Solidity*, (Vol. 318). Berkeley: Apress.
19. Vogelsteller, F., & Buterin, V. (2018). *ERC-20 token standard, 2015*. [Online]. Available: <https://github.com/ethereum/EIPs/blob/master/EIPS/eip-20.md>. GitHub Site.
20. Buterin, V. (2018). Standardized contract APIs.
21. Zastrin.com. (2019). Geth/Parity Clients. [Online]. Available: <https://www.zastrin.com/courses/ethereum-primer/lessons/3-1>.
22. Day, A., & Medvedev, E. (2019). Ethereum in BigQuery: a public dataset for smart contract analytics. *Google Cloud Blog*. [Online]. Available: <https://cloud.google.com/blog/products/data-analytics/ethereum-bigquery-public-dataset-smart-contract-analytics>
23. *Ethplorer.io* (n.d.) [Online]. Available: <https://ethplorer.io>
24. *Etherscan.io*. (2019) [Online]. Available: <https://etherscan.io>
25. *Infura.io*. (n.d.) [Online]. Available: <https://infura.io>
26. Merriam, P. & Carver, J. (2018). *Web3.py – Web3.py 4.9.2 documentation*. [Online] Available: <https://web3py.readthedocs.io>.
27. *MGCToken.io*. (n.d.) [Online]. Available: <https://mgctoken.io>
28. *CryptoKitties.co*. (n.d.) [Online]. Available: <https://cryptokitties.co>
29. Entriken, W., Shirley, D., Evans, J., & Sachs, N. (2018). Non-Fungible Token Standard, document ERC-721.
30. *Kyber Network* (n.d.). [Online] Available: <https://kyber.network>