EXPLORING THE ETHEREUM BLOCKCHAIN: ANALYTICS

*

Abstract. Ethereum is a community-built technology behind the cryptocurrency ether (ETH), a largest decentralised blockchain platform for smart contracts with over \$385 billion in market capitalization (as of 31st August, 2021).It provides open access to digital money, easy to use banking services regardless of your background or location. Smart contracts are the applications that allow users to do transactions with each other without having the need for a trusted central authority. They are immutable making them attractive for various decentralised applications (or dApps) like e-governance, healthcare management and data provenance. Ethereum, and cryptocurrencies in general, have been in the center of criticism regarding their volatility and the speculation in their prices. In this study, an in-depth big data analysis on the Ethereum blockchain is performed, revealing interesting trends. The main focus is on understanding what the number of transactions in the Ethereum network represents and what the volume of ether in circulation between the different addresses is. For that purpose, the Google BigQuery Ethereum database is used in order to analyse all the available data and answer questions.

Keywords: Ethereum, Blockchain, Ether, ETH, Smart Contracts, Cryptocurrency, BigQuery

1 INTRODUCTION

Satoshi Nakamoto introduced Bitcoin in 2009, a digital currency for peer to peer transactions which opened the door for the whole cryptocurrency world.. Bitcoin is a cryptocurrency that operates only as a digital form of money, whereas Ethereum has a multifunctional character allowing users to create smart contracts that can interact and be executed. As a result, Ethereum offers the vision of a decentralized future, where third-parties are eliminated and users have direct access to the markets. Understanding how Ethereum works helps us identify what the role of blockchain will be in tomorrow's world, and when it makes sense to use it. A signif-

icant portion of this work is inspired from Athina Voulgari's Master Thesis at ETH Zürich, Management Technology and Economics Department.

2 BACKGROUND

The elementary idea that led to the birth of blockchain was first proposed in 1990 by S. Haber and W.S. Stornetta. They introduced a novel, computationally feasible set of procedures to timestamp digital data, which would make it impractical to alter the timestamp after its creation, making it immutable [8]. Virtually anything of value can be tracked and traded on a blockchain network, thereby reducing risk and cutting costs for all involved.One of the most important implications of this approach was that there would be no need for a third-party service to keep record of the timestamps. Since then, blockchain technology has revolutionised many fields including healthcare [3] [6], transportation [18] [15], digital forensics [10], and cybersecurity [1] [19] due to its reliability, immutability, and transparency. These characteristics are the direct results of the blockchain structure: data is divided into a collection of blocks that are all linked together by means of cryptography. This structure prevents tampering with any arbitrary block without changing all others; hence, achieving immutability. The decentralized data handling also prevents the two parties in a transaction from manipulating the data stored in the network. In essence, the blockchain establishes a general agreement that verifies the details of atransaction without the need to trust the parties involved [21].

3 DISTRIBUTED LEDGER TECHNOLOGY (DLT)

Distributed Ledger Technology (DLT),refers to a novel approach to recording and sharing data across multiple data stores (ledgers), which store the exact same data and are maintained and controlled by a distributed network of computer servers, which are called nodes [14]. The innovative feature is that the ledger is not maintained by a central authority.
Each node, independently, builds and records the updates in the ledger. After that, these updates must be accepted by most of the network. In order to overcome the challenge of ensuring that all the nodes agree upon the validity of the ledger updates, there have been different mechanisms that define the way of agreeing on this; this agreement is called consensus. Once the consensus has been reached, the distributed ledger updates itself and the latest, agreed-upon version of the ledger, is saved in each node separately [16].
There are different consensus mechanisms to validate the agreement given the set-up of the distributed ledger. Consensus algorithms define the steps and the order of the steps that need to be done in order to produce an output [12]. They constitute a fundamental component of distributed networks and are crucial for their functionality.

4 BLOCKCHAIN TECHNOLOGY

Blockchain was first mainly mentioned in the Bitcoin white paper [13]from Satoshi Nakamoto, published in 2008, as the underlying technology of Bitcoin. It was the first fully functional Distributed Ledger Technology (DLT) and the only one people have known for many years. Blockchain is nothing more,a DLT with a specific set of features that distinguish blockchain from all existing DLT technologies [7]. It is a database, shared by means of blocks, which are connected with each other in the form of a chain.

Blockchain offers the main functionalities of DLTs. The three main properties that determine the Blockchain technology are decentralization, transparency, and immutability [9]. The term of transparency comes in contradiction to the concept of privacy that is claimed to be a feature of blockchain. A person's identity is represented by a public address which is cryptographed. It is not the name of the sender that is referred to in a transaction, but the hash representation. As a result, the physical identity of a user is secure, while the digital identity is monitored and recorded, so that transparency is secure as well. Furthermore, blockchain ensures that no recorded transaction can be altered or removed. Even the slightest change in a block will completely change the hash address of the block. This immutability feature makes blockchain fully reliable. Finally, in a decentralized system, the information is not stored in a single entity. On the contrary, a basic component of a blockchain is its peer-to-peer (P2P) network. Each user can use the network and provide resources at the same time. A node can be any active electronic device that is connected to the Internet and provides an IP address. Although all nodes are equal, they can serve the network in different ways. A special reference needs to be made to these nodes that are called "full nodes" or miners. The purpose of a full node is to copy the latest blockchain data to a single device, while the device is connected to the network. As a result, this information is stored in all full nodes of the network and can only be altered if all these nodes are destroyed, which makes the system less vulnerable.

Blockchains have the potential to build a new generation of transactional applications that establish trust, accountability, and transparency at their core while streamlining business processes and legal constraints. There are three main types of blockchains- Public Blockchain, Private Blockchain and Federated or Consortium Blockchain.

5 ETHEREUM BLOCKCHAIN

The Ethereum blockchain is a type of distributed ledger technology (DLT), a general term used for databases that store and share information in a decentralized network of independent nodes.

Ethereum is an open source, blockchain-based platform, which eliminates the need for trusted intermediaries. According to Dr. Gavin Wood, co-founder of Ethereum organization, Ethereum, taken as a whole, can be viewed as a transaction-based

state machine: we begin with a genesis state and incrementally execute transactions to morph it into some final state. It is this final state which we accept as the canonical "version" of the world of Ethereum [22]. It allows everyone to set their own rules in transactions and state transition functions. This is done by involving smart contracts, a set of cryptographic rules that are executed under the conditions that the creator of the contract has set up.

Essentially, blockchains help in maintaining a tamper-proof log that is used to establish ownership of actions. Therefore, it has wide applications and can help establish the trust of the users in the system.

Ethereum allows developers to program their own smart contracts, or "autonomous agents", as the Ethereum whitepaper [4] calls them. The solidity language is "Turing-complete", which means that it supports a broader set of computational instructions. In a nutshell, smart contracts can:

- Function as "multi-signature" accounts, so that funds are spent only when a required percentage of people agree.
- Manage agreements between users, e.g. if one purchases an insurance product from another.
- Provide utility to other contracts (similar to how a software library works).
- Store information about an application, such as domain registration information or membership records.

One trade-off of Ethereum is related to its consensus mechanism (Proof of Work). It is the most popular algorithm for verifying the correctness of a transaction happening in the network. On one hand, it requires a lot of computational power in order to verify any transaction and at the end to build a block and validate it. On the other hand, the fact that it requires so much computational power makes it difficult to replicate and change the whole chain and as a result, reduces the risk of a 51% attack. At the same time, each individual solution is easy for the community to verify, which makes it easy to check all transactions for trustworthiness. The fact that it doesn't rely on a single third-party entity makes it a "trustless" and transparent network.

Even though most terms used in the context of Ethereum are not exclusive to this particular blockchain, there are a few terms that are. This section provides a brief introduction to the technical terms related to the Ethereum blockchain that are used in this paper. A comprehensive explanation of terms and concepts can be found in the Ethereum whitepaper [4].

6 THE ETHEREUM MODEL

6.1 Accounts

Accounts play an important role in Ethereum. There are two kinds of accounts: Externally Owned Accounts (EOAs) and Contract Accounts. All the accounts in blockchain are characterised by four core elements:

- The nonce, a counter that is used to ensure that each transaction can only be processed once representing the number of transactions successfully sent from this account (if it is an EOA), or the number of contracts created by it (if it is a contract account).
- The account's current ether balance (the number of wei owned by an account).
- The account's code (only for contract accounts).
- The account's storage (permanent data stored, only for contract accounts).

6.2 Smart Contracts

The concept of smart contracts was introduced by N. Szabo in 1994 in an unpublished manuscript and then formally in 1997 [20]. Smart contracts are essentially blockchain-based applications, concisely, self executing programs which contain the terms and agreements between the parties involved.

A smart contract is represented by its 20-byte address, similar to an EOA address, such as '0xC2C747E0F7004F9E8817Db2ca4997657a7746928', and it can be developed in high-level languages,such as Solidity. When the contract is deployed on the network, the solidity code is compiled as bytecode. The feature that sets smart contracts apart is that they automatically verify whether the terms of the agreement have been fulfilled or not. Additionally, to ensure reliability, fault tolerance, and transparency, the codes (smart contracts) are replicated on many nodes in the blockchain.

There is an infinite amount of contracts that can be created in the network. However, there are some "templates" that are commonly used, as they provide some main functionalities for contracts that want to achieve similar behaviours. These are token standards that define a common list of rules and are called ERC (Ethereum Request for Comments). Some of these are described below:

ERC20: This standard is a simple interface for creating tokens and can be reused from any application. It is the most commonly used standard in Ethereum. It consists of 6 main functions and 2 events. These functions are:

- Total Supply: defines the initial total supply of tokens
- Balance: monitors the balance of the contract every time
- Transfer to: sends tokens to specific accounts (wallets)
- Transfer from: enables token holders to exchange tokens after the initial distribution of the tokens
- Approve: approves other accounts to withdraw tokens from the account calling the function
- Allowance: after the approval of withdrawing tokens, it is used to see the amount of tokens that is withdrawn from the account

ERC20 is used from wallets in decentralised exchange platforms. Moreover, it is the main standard used for ICO contracts [5].

ERC721: This is a standard developed for non-fungible tokens. These tokens are completely unique. This feature of the standard makes it suitable for representation of assets that cannot be duplicated. A good example is the ownership over assets, such as music, art pieces, or digital collectibles.

6.3 Transactions

Ethereum works on an account-based model, as our known banking system. The state of Ethereum changes every time a transaction of value or information is executed between two accounts. So, the state consists of these account addresses. There are three kinds of transactions:

- Ether transfer: An external account can transfer Ether to another external account or a contract. This kind of transaction is very similar to a Bitcoin transfer.
- Contract creation: An external account can create a contract by transferring ether to a zero recipient's account. Then a new contract is created. The transaction should contain a code defining what the new contract will do. The execution of the transaction will create the bytecode of the new contract.
- Contract call: After a contract is deployed, an external account can call a contract when an account intends to implement one or more functions of a contract. The input data contains all the instructions related to the execution of the contract.

6.4 Ether

Ether (ETH) is the currency of Ethereum. It has intrinsic value as any currency. Ether derives its value from different factors. It is used to pay transaction fees in Ethereum. Furthermore, it can be accepted as payment currency from some retailers, such as the largest online Swiss retailer Digitec Galaxus AG. It can also be lent or borrowed. Finally, it is used as a medium of exchange to purchase Ethereum-based tokens. One example of the latter is an Initial Coin Offering (ICO) [17].
Ether serves all the three conditions in order to be considered as money within an economy:

- It works as a store of value, where investors purchase it and hold it for investment purposes given its predictable supply growth and congenital utility.
- It serves as a medium of exchange within the Ethereum network
- It is used as a unit of account by various parties (including companies that have raised Ether via ICOs) [17].

The smallest denomination of Ether's metric system is Wei, where 1 Ether $= 1e^{18}$ Wei.

6.5 Gas

On the Ethereum blockchain, the cost of performing transactions or processing smart contracts is measured by gas. It can be considered as the fuel for operating Ethereum. The price of gas itself is not constant, but is reported by miners based on the complexity and computational resources required for the execution of each block. Every time a transaction or a contract is executed, users spend tokens which are translated to gas to run the computations. A user has to pay for the computational power spent, regardless of whether the transaction was successful or not. Gas is therefore an implicit incentive for developers to produce contracts that are of low-cost execution and for miners to validate transactions. Like this, the community avoids a waste of resources and eliminates the motivation for denial-of-service attacks.

6.6 Ethereum Virtual Machine

Ethereum Virtual Machine (EVM) is the computation engine of the Ethereum network and handles smart contracts deployment and execution. EVM is required in order to update the Ethereum state. An ether transfer from an account to another does not require an update of the state and as a result, EVM is not involved. However, it is used for any other action in the network. Specifically, EVM is responsible for all of the following executions in the network:

- EVM confirms the validity of a transaction. This means that it checks the correct number of values, the validity of the signature and the match of the none with the nonce of the particular transaction account. If any of these elements is not correct, an error will return back.

- EVM checks if the gas is enough in order to execute the transaction. In case gas is not enough, the transaction fails. However, the transaction fee in such a case would not be refunded and would be paid to the miner.

- EVM transfers the value of Ether to the recipient's address.

- EVM calculates the total gas used and the transaction fee, in order to initialize the gas payment to the miner.

6.7 Blocks

Blocks are a fundamental component of Ethereum. All the information is stored in blocks. Each block consists of a block header and a set of transactions mined in the block. The chain is created by hashing some of the information of the previous block and including it in the new one. The hash IDs and timestamp associated with each block are the virtual glue that connects the blocks and forms the blockchain. Any alteration of the ledger or an individual block will make the code unworkable and, as a result, would be detectable.

The Ethereum blockchain began its life with its own genesis block. After this zero-block initiation, all other activities, such as mining, transactions, and contracts change the state every time.

6.8 Mining

Mining is the process of validating new transactions and creating new blocks. As a result, miner nodes will collect transactions from the transaction pool, will execute them in the EVM, solve the nonce and create the new block that fits to the chain. However, Ethereum is on the verge of moving to a Proof of Stake (PoS) consensus mechanism and the whole way of validating and receiving rewards would change in proof of stake consensus and would reform the way mining is done in the Ethereum blockchain.

7 DATA ACCUMULATION

The Ethereum database size is currently around 280 GB, according to blockchair. There is no central administrative system where the data can be downloaded from. One feature of Ethereum is that all data is publicly available. The nodes, which share data among each other, store a copy of the data, while the network executes the Ethereum protocol, which defines the rules of interaction of nodes with each other and/or smart contracts over the network.

In August 2018, Google Big Query managed to download all the blocks' data in an SQL database, in real-time. The first phase relies on Ethereum ETL - an open-source tool developed in order to export the Ethereum blockchain into CSV or JSON files. It connects to an Ethereum node via its JSON RPC interface. The exported files are then moved to Google Cloud Storage, loaded into BigQuery, and finally verified. Then, the user can query the data in the BigQuery console or via API [11]. Since all the data is available there and is updated every day, this source of data will be used as the main one for the analysis of the Ethereum data.

8 DATA EXTRACTION AND USAGE

We extract and use all relevant data from the crypto_ethereum dataset [2] under the Google Cloud bigquery-public-data repository. Google BigQuery is a data warehouse that handles large-scale data and makes it easy to access via the SQL interface. Ethereum blockchain data is available on it and is updated daily. The entire blockchain data is stored in seven different tables: 'blocks', 'contracts', 'logs', 'token_transfers', 'tokens', 'traces', and 'transactions'.

9 STRUCTURE OF THE DATA

We depict a relational diagram of the database in Figure 1. There are some fields that overlap in the tables, such as the block number and the block timestamp. The tables are structured this way in order to have a relational structure and be easy to analyze. Although the tables consist of all the data fields displayed, only some of them will be used in our analysis.
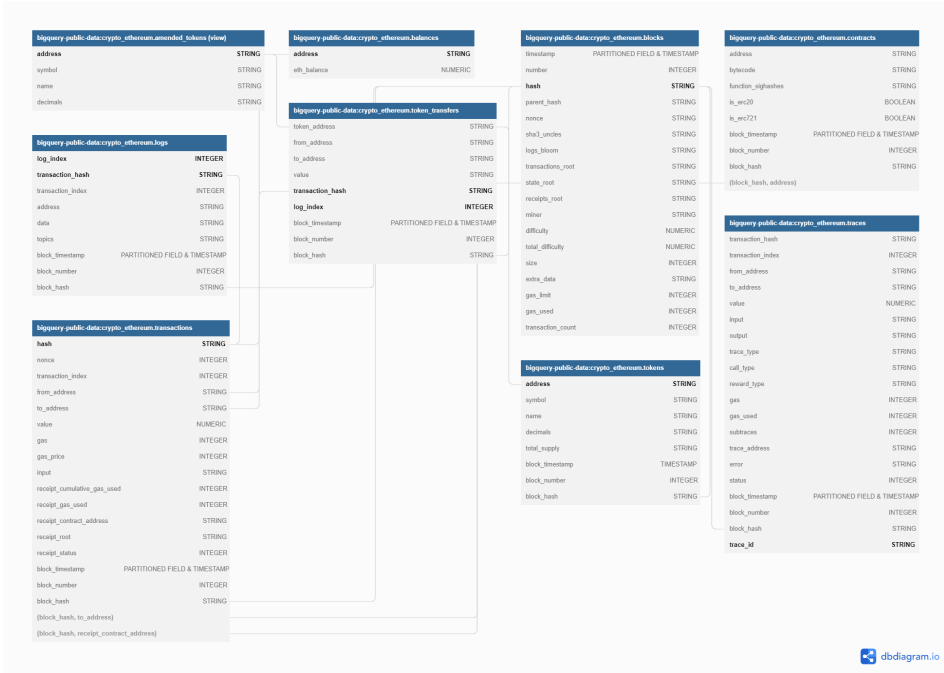

Fig. 1. Dataset relational diagram

10 DATA ANALYTICS

In the previous sections, distributed ledger technology, and blockchain technology were analyzed as concepts and the focus was mainly on the Ethereum structure. The Ethereum blockchain contains a lot of data, not only for the pure currency transactions but also for the smart contracts that operate in the network and change the state of Ethereum.

In the succeeding sections, the focus will be initially given to a comparison of how Ethereum is being used in general over time and to its "trends". Then, the number of transactions and the total value of Ether will be used as a common metric in order to compare different external accounts, different contracts, and different tokens.

Additionally, the top external accounts, contracts, and tokens will be identified and compared with each other, in an attempt to define the reason why these accounts present a large number of transactions or value of Ether. At last, a closer look will be taken at the contracts analysis, and more particularly, their lifecycle and how they are used.

Etherscan.io, a Block Explorer, Search, API, and Analytics Platform for Ethereum, will be used for identifying specific details for different addresses.

10.1 Generic Analytics

In this section we will analyze the demography of the Ethereum blockchain and provide a snapshot of it as of 31st August 2021.

10.1.1 Transactions

The Ethereum network has processed 1,266,449,457 transactions as of 31st August 2021. Fig. 2 shows a visualization of the number of transactions occurring over time in the Ethereum network. The peak was recorded on 9th May 2021 when 1,716,600 transactions went through in a single day!

The Ethereum blockchain has attracted a lot of attention in 2021 primarily because of NFTs being created on it. The number of transactions recorded on 31st August, 2021 is 1,206,314 which is more than double the total average number of daily transactions since the genesis of the blockchain, which is roughly 571,244.

It is also interesting to see what the total value of ether is in the transactions. There may be a lot of transactions, but of small ether value, or the other way around, meaning only a few transactions where a large amount of ether is transferred. The total value of ether transferred between accounts is 8,708,084,863 ETH. Fig. 3 shows a visualization of the value of ether transferred per day.

The peak was recorded on 11th August 2017 with approximately 46,000,000 ETH being transferred. The average value of ether transferred per day sits around 3,927,868 ETH per day.

10.1.2 Gas Used

The more complex a transaction is, the more gas is consumed to process the transaction. Fig. 4 shows a visualization of the daily gas consumption in the Ethereum network. The continuous increase in gas being used to process transactions is proving to be a problem for the Ethereum network and therefore developers are looking at other alternatives like Solana and Cardano. The average gas used is 36,438,799,019 per day. The highest amount of gas was used on 7th August 2021 with 101,523,049,266 gas being used in a single day! The total gas used since the genesis of the Ethereum blockchain is 81,112,766,616,753.

In January 2018, the Ethereum network started operating, not only as a simple ether transfer network, but also as a contract execution network, and since the latter
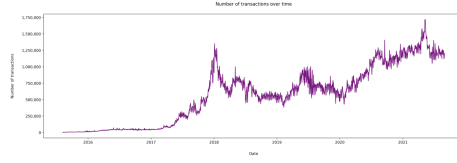
Fig. 2. Number of transactions over time in the Ethereum network. The maximum number of transactions recorded on a single day was on 9th May, 2021 with 1,716,600 transactions.
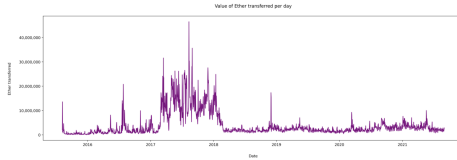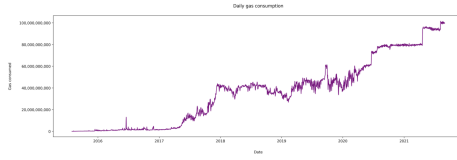

Fig. 3. Value of Ether transferred per day
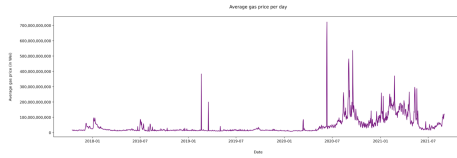

Fig. 4. Daily gas consumption

Fig. 5. Daily average gas price (in Wei)

type of transaction costs more in terms of gas, we have seen a continuous increase in gas being used to process transactions.

10.1.3 Gas Price

The gas price is set by the initiator of a transaction. The higher the price is, the more probable for the transaction to be prioritized higher than others. Fig. 5 shows the average gas price per day in the Ethereum network. This average takes into consideration only the gas prices of the successful transactions. Gas prices of transactions that were not executed are not, since this would violate the real level of gas evaluation.

On June 11, 2020, the peak in the daily average gas price, just greater than

12

722,000,000,000 was recorded. The sum of gas prices since the genesis is 61,575,600,617,544█ and the average gas price of the Ethereum blockchain is 43,485,593,656.

10.1.4 Active addresses growth

Users constitute a major part of any network's success. The Ethereum network contains not only individuals but also contracts. Fig. 6 shows a visualization of the active address growth in the Ethereum network. The addresses are extracted based on their first transaction where the transaction value is greater than 0. As of 31st August 2021, there are 95,429,623 addresses in the Ethereum network which have done a transaction with a value larger than zero. In almost a year, compared to 2020, the number of distinct addresses has doubled in 2021.

However, the same graph, considering also addresses with a zero balance, shows a significant difference (Fig. 7). In Fig. 8, a sharp increase in the number of addresses by 20 Million can be observed. On 22nd September 2016, almost 20,000,000 addresses were suddenly created. This was the result of a DDoS attack. A hacker took advantage of an underpriced Opcode and created around 20,000,000 useless Ethereum accounts. This led to the creation of millions of transaction traces, making data analysis difficult. At that specific time, the network was slowing down, making it more time-consuming for miners and nodes to process blocks. The Ethereum network underwent a hard fork after that at block number 2,463,000.

So, it can be stated that the total number of addresses is approximately 95M, given the 115M transactions minus the 20M useless addresses.
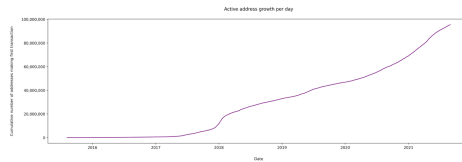
Fig. 6. Cumulative active address growth per day (having made a transaction with value greater than 0)
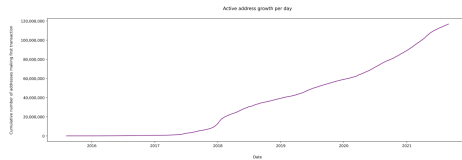
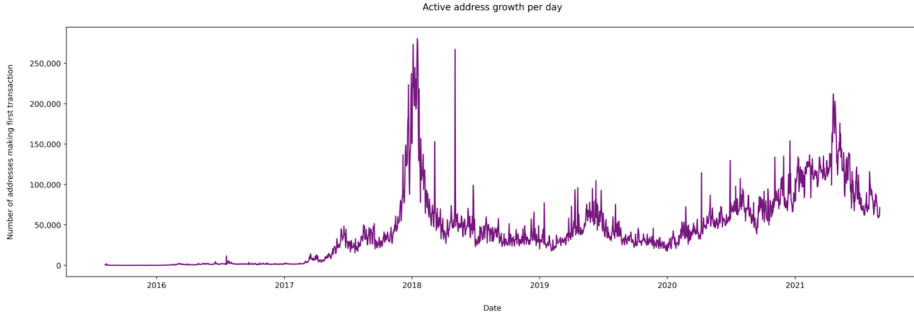Fig. 7. Cumulative active address growth (also considering addresses that haven't made any transaction)

Fig. 8. Active address growth per day (having made a transaction with value greater than 0

10.1.5 Smart Contracts

There are 46,376,065 addresses that represent smart contracts on the Ethereum blockchain. The growth of smart contract addresses is displayed in Fig. 9 and Fig. 10. Also, the contracts that are created directly from external users will be separated from the ones that are created indirectly from a contract generating another contract. This can be seen in Fig. 11 and Fig. 12. The number of smart contract addresses generated directly by users are 4,265,952. This means that around 42 Million smart contracts have been created indirectly from a contract generating another one.

The contract creation does not follow the pattern of transactions having high fluctuations. This means that over time users have understood the major functionality of Ethereum and its utility, i.e, Smart Contracts.
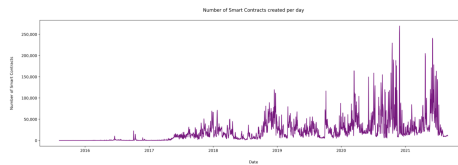

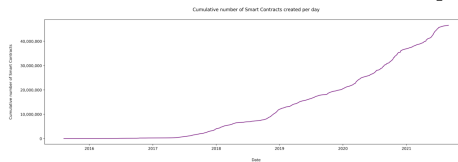Fig. 9. Number of Smart Contracts created per day

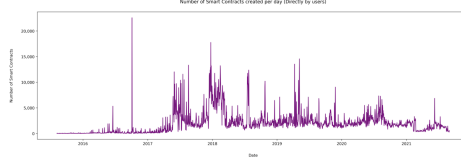Fig. 10. Cumulative number of Smart Contracts created per day

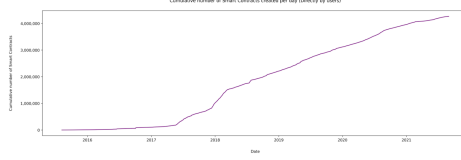Fig. 11. Number of Smart Contracts created per day (Directly by users)

Fig. 12. Cumulative number of Smart Contracts created per day (Directly by users)

10.1.6 Classification of Addresses

In Figs. 6 to 8, the Ethereum address growth for external user addresses is shown and in Fig. 10, the Ethereum address growth for smart contracts is shown. It is visible that the allocation of the total addresses has changed since the inception of Ethereum. Initially, most of the addresses were primarily external users and few contracts, but through time, contracts gradually gained more ground and are continuing to do so.

10.1.7 ETH Price

A visualization representing ETH price in USD, as reported on [CoinMarketCap], is shown in Fig. 13. The ETH price has followed the same pattern as the number of transactions. The highest price of $4168.7 was recorded on 11th May 2021. Ethereum's price as of 31st August 2021 is $3433 with a total market cap of around $403,531,078,688.

10.2 Focused Analytics

10.2.1 Analytics of transactions

Transactions are one of the major sources for answering many questions about Ethereum. They display all the activity that is happening from external accounts to other external accounts or contracts. A transaction between two external accounts can only transfer value. On the other hand, a transaction from an external account to a contract account activates the function of the contract performing different actions such as creating new contracts, writing to internal storage, transferring tokens, performing other actions, etc.
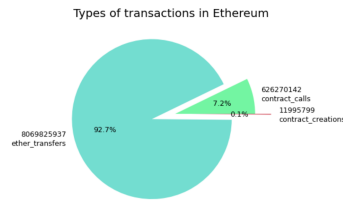
Fig. 13. ETH Price over time in USD

Fig. 14. Types of transactions in Ethereum

Figure 14 shows that 92.7% of the transactions are Ether transfers. However, almost 7.2% of the transactions are calling a contract in order to execute a function. This indicates that a significant fraction of the activity in Ethereum is because of Ether transfers which are followed by contract calls, and just 0.1% of all transactions are contract creation transactions which is very low. It is also important to mention that these numbers only refer to contracts created by external users.

The nature of transactions has changed a lot over the years. From the initialization of the Ethereum network until now, there is a gradual increase for all types of transactions but Ether transactions are in favor. This is because Ethereum launched after the success story of Bitcoin and the generic evolution of digital currencies. Users were considering Ethereum just another cryptocurrency for peer-to-peer transactions or a speculative investment opportunity.

10.2.2 Analytics of volume

A total of 117 Million Ether has been supplied in the network, while all the Ethereum transactions have circulated more than 8.7 Billion Ether, as of 31st August 2021. The daily aggregated ether volume of these transactions, both for simple ether transfers, as well as for contract calls, can be found in Fig. 16. Most of the ether value is circulated in simple ether transfers, from account to account. For almost a year, from 2017 to 2018, the average volume of ether transfers is around 10M ether per day. The number of contract calls has risen significantly between 2020 and 2021. The primary reason for this is a surge in the NFT market.

Address	Balance (ETH)	Name
0x00000000219ab540356cbb839cbe05303d7705fa	7,280,770	Ethereum 2.0 deposit contract
0xc02aaa39b223fe8d0a0e5c4f27ead9083c756cc2	7,032,141	Wrapped Ether
0xbe0eb53f46cd790cd13851d5eff43d12404d33e8	2,296,896	Binance 7
0x73bceb1cd57c711feac4224d062b0f6ff338501e	1,949,046	-
0x4ddc2d193948926d02f9b1fe9e1daa0718270ed5	1,747,254	Compound Ether (cETH)
0x53d284357ec70ce289d6d64134dfac8e511c8a3d	1,378,734	Kraken 6
0x9bf4001d307dfd62b26a2f1307ee0c0307632d59	1,040,000	-
0x61edcdf5bb737adffe5043706e7c5bb1f1a56eea	989,498	Gemini 3
0xc61b9bb3a7a0767e3179713f3a5c7a9aedce193c	700,010	Proxy
0x229b5c097f9b35009ca1321ad2034d4b3d5070f6	657,334	Huobi 18

Table 1. Top 10 addresses (by ether balance)

Address	Number of txns sent	Balance(ETH)	Name
0xea674fdde714fd979de3edf0f56aa9716b898ec8	37,053,732	38,526.8	Ethermine
0x52bc44d5378309ee2abf1539bf71de1b7d7be3b5	17,588,058	11,164.2	Nanopool
0x829bd824b016326a401d083b33d092293333a830	10,795,620	19,566.9	F2Pool Old
0xfbb1b73c4f0bda4f67dca266ce6ef42f520fbb98	10,572,809	14,156.8	Bittrex
0x3f5ce5fbfe3e9af3971dd833d26ba9b5c936f0be	7,735,908	3,079.2	Binance
0x5a0b54d5dc17e0aadc383d2db43b0a0d3e029c4c	6,061,515	34,644.3	Spark Pool
0xa7a7899d944fe658c4b0a1803bab2f490bd3849e	5,818,086	0	IDEX 2
0xd551234ae421e3bcba99a0da6d736074f22192ff	4,973,831	0.26	Binance 2
0x0681d8db095565fe8a346fa0277bffde9c0edbbf	4,943,611	2.46	Binance 4
0x564286362092d8e7936f0549571a803b203aaced	4,861,789	0.015	Binance 3

Table 2. Top 10 addresses (by number of transactions sent)

Address	Number of txns received	Balance(ETH)	Name
0xdac17f958d2ee523a2206206994597c13d831ec7	110,324,410	0	Tether: USDT
0x7a250d5630b4cf539739df2c5dacb4c659f2488d	46,110,132	0	Uniswap V2: Router 2
0xa0b86991c6218b36c1d19d4a2e9eb0ce3606eb48	13,381,871	0	USD Coin
0xa090e606e30bd747d4e6245a1517ebe430f0057e	12,161,990	2.1	Coinbase
0x8d12a197cb00d4747a1fe03395095ce2a5cc6819	11,538,109	17141.5	EtherDelta 2
0x2a0c0dbecc7e4d658f48e01e3fa353f44050c208	9,785,344	18076.9	IDEX
0x3f5ce5fbfe3e9af3971dd833d26ba9b5c936f0be	9,282,994	3079.2	Binance
0x174bfa6600bf90c885c7c01c7031389ed1461ab9	9,088,327	0	-
0x06012c8cf97bead5deae237070f9587f8e7a266d	4,875,001	6.9	CryptoKitties: Core
0x514910771af9ca656af840dff3e8264ecf986ca	4,764,625	0	ChainLink: LINK Token

Table 3. Top 10 addresses (by number of transactions received)

Type of contract	Number of contracts	Number of unique creators	Number of unique bytecodes
User-created contracts	4,209,516	210,172	424,875
Contract-created contracts	42,507,077	49,037	35,756

Table 4. Different types of contracts

Address	Transaction Count	Name
0x06012c8cf97bead5deae237070f9587f8e7a266d	4,879,451	CryptoKitties: Core
0x06a6a7af298129e3a2ab396c9c06f91d3c54aba8	646,132	0xUniverse: PLANET Token
0xd73be539d6b2076bab83ca6ba62dfe189abc6bbe	442,760	Blockchain Cuties: Old BC Token
0x1a94fce7ef36bc90959e206ba569a12afbc91ca1	180,531	Crypton (CTN)
0xf5b0a3efb8e8e4c201e2a935f110eaaf3ffecb8d	147,500	Axie Infinity: AXIE Token
0x7fdcd2a1e52f10c28cb7732f46393e297ecadda1	119,472	HyperDragons Token
0x8c9b261faef3b3c2e64ab5e58e04615f8c788099	108,041	LucidSight-MLB-NFT
0x2a46f2ffd99e19a89476e2f62270e0a35bbf0756	95,986	MakersTokenV2 (MKT2)
0xf915bbfbb6c097dc327e64eec55e9ef4d110d627	78,666	Servant (SVT)
0xd2f81cd7a20d60c0d558496c7169a20968389b40	60,852	Etherbots (ETHBOT)

Table 5. Top 10 Ethereum collectibles (ERC721 contracts) by popularity, based on number of transactions

Address	Transaction Count	Name
0xdac17f958d2ee523a2206206994597c13d831ec7	111,998,435	Tether: USDT
0x174bfa6600bf90c885c7c01c7031389ed1461ab9	9,088,327	More Gold Coin (MGC)
0x514910771af9ca656af840dff83e8264ecf986ca	4,865,359	ChainLink Token (LINK)
0x86fa049857e0209aa7d9e616f7eb3b3b78ecfdb0	2,970,081	EOS: Old Token
0xc02aaa39b223fe8d0a0e5c4f27ead9083c756cc2	2,899,686	Wrapped Ether (WETH)
0x0d8775f648430679a709e98d2b0cb6250d2887ef	2,725,166	Basic Attention: BAT Token
0xd26114cd6ee289accf82350c8d8487fedb8a0c07	2,509,541	OMG Network (OMG)
0x8fdcc30eda7e94f1c12ce0280df6cd531e8365c5	2,229,775	CpcToken (CPCT)
0xf230b790e05390fc8295f4d3f60332c93bed42e2	2,084,178	Tronix (TRX)
0x95ad61b0a150d79219dcf64e1e6cc01f0b64c4ce	1,815,190	SHIBA INU (SHIB)

Table 6. Top 10 Ethereum tokens (ERC20 contracts) by popularity, based on number of transactions - Not Complete
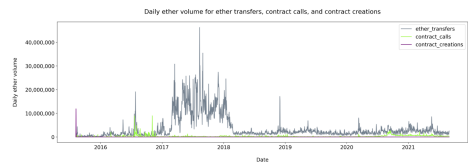

Fig. 15. Daily ether volume for ether transfers, contract calls, and contract creations
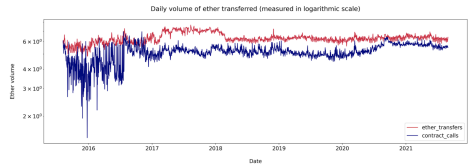

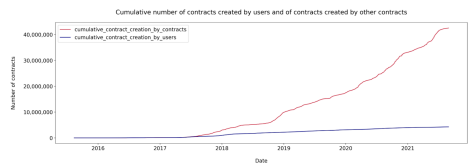Fig. 16. Daily volume of ether transferred (logarithmic scale)

Fig. 17. Cumulative number of contracts created by users, and contracts created by other contracts
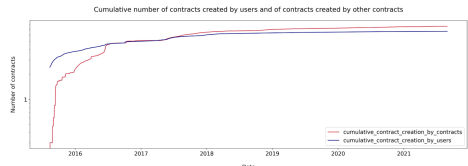

Fig. 18. Cumulative number of contracts created by users, and contracts created by other contracts (Logarithmic scale)

10.2.3 Analytics of accounts

This section presents some observations regarding the activity of the accounts, as measured by the ether volume and the number of transactions in the Ethereum network. Let us identify in which category the top 10 addresses, measured by ether balance, belong. There are some basic categories that have been defined, where all the transactions that are executed in the Ethereum network are displayed. The main categories that are used here are:

- Exchanges
- Decentralized exchanges
- Miners
- Token contracts
- ICO wallets

All the remaining addresses are assumed to be either simple contracts executing specific code or still unlabeled or external addresses. Etherscan identifies if an address is a contract address or a simple external address. There are some clear distinctions among the different categories.

- **Exchanges:** A list of centralized cryptocurrency exchanges which are online platforms that allow customers to buy and sell cryptocurrencies for other assets.
- **Decentralized exchanges:** A decentralized exchange (also known as a DEX) is an exchange market that does not rely on a 3rd party service to hold the client funds, but, instead, trades occur directly between users (peer to peer) through an automated process.
- **Miners:** Accounts that belong to big mining organizations or single individuals.
- **Token contracts:** A Token Contract is a type of smart contract that describes the rules of how the token could be generated and transferred between addresses, if it is splittable/fungible, etc.
- **ICO wallets**

In table 1, the top 10 addresses having the greatest balance of ether are displayed. Names of the address owners have been explored using Etherscan. Most of these addresses are 'exchanges'.

The ranking of addresses based on the ether balance is only one metric of understanding which addresses are important for the Ethereum network. The number of transactions is another important indicator. Transactions are defined by an address which sends the transaction and an address which receives the transaction. The sender can only be an external account (single user, miner, or exchange), but a recipient address can be any address (external or contract address). In table 2, the top 10 addresses, measured by the number of transactions sent, are shown. Based on the data, most of these addresses are either exchanges or mining companies. We

can see miners in this table because they are the ones who validate all the available transactions in the network, but we don't see them in table 13.1 as they do not primarily invest in ether. Their balance is generally from the rewards they receive during the mining process.

It is also important to identify the top addresses, based on the number of transactions received. Table 3, displays these addresses. This is a more diversified sample of addresses since one can find exchanges, token contracts, as well as general contracts. It is surprising that although there are exchanges and token contracts in the table, these are not the same as the ones in table 2.

Again, the accounts that receive most of the transactions in the network just own very few ethers. This makes sense as most of them are contracts and they are not supposed to hold ether but only to execute a sequence of orders. There is one interesting address on this list: CryptoKitties. CryptoKitties is a project utilizing ERC-721 tokens, which focuses on gaming where users can collect virtual cats. Each cat is represented by an Ethereum ERC-721 token, which means that they are all one-of-a-kind and can never be replicated, taken away from the owner, or destroyed. It is surprising to see a gaming platform ranking so high in the number of transactions received.

10.2.4 Analytics of smart contracts

As referred to in section X (E), there are in total 46,376,065 contract addresses. For these contracts, the source of creation is analyzed. There are contracts, which are created directly from an external address, and contracts which are contract-created. In the former ones, the user has to set the recipient address (to_address) to zero. In the latter ones, an external user calls the 'create' function of contract A, which in turn creates contract B. This is a message call (internal transaction) and it is not counted in the number of transactions. So, only the direct contract creations are considered transactions and represent only about 0.3% of the total transactions. The indirect contract creations are reflected in the contract calls. It is very interesting to see in Figs. 17 and 18 that most of the contracts were initially user-created contracts. However, as Ethereum is getting more mainstream and developers start participating and building more complicated contracts, contract-created contracts dominate the first ones since mid-2017. In fact, today there are over 42M contact-created contracts, while the user-created contracts are only over 4M. This means that an important fraction of the activity in the Ethereum network is using more complicated features of the contracts, such as the creation of another contract.

Diving into the contracts, we want to identify who is creating these contracts and how many of these contracts are totally identical. In order to identify the identical contracts, the output of the traces will be compared. The output is the bytecode of the contract whenever a new contract is created. The bytecode is a hex representation of what a contract is produced to execute.

From table 4, it can be clearly seen that a very small number of external users have directly created contracts. What is even more surprising is that only about

49K contracts are responsible for the creation of 42M contracts. Comparing the bytecodes of the user-created contracts, about 10% of them are unique, as there are only about 424K different bytecodes in the total of 4.2M contracts. At the same time, comparing the bytecode of the contract-created contracts, only about 35K of unique bytecodes are used in a total of 42M contracts! In conclusion, analyzing the two types of contracts yields that only around 0.98% of the total contracts are completely unique.

We have also segregated smart contracts based on their type, i.e, ERC20 and ERC721, and analysed them based on their popularity taking the number of transactions as the metric. The results for this are presented in tables 5 and 6.

Table 5 shows that CryptoKitties, an ERC721 collectibles platform, is the most popular. Even after summing up the transaction count of the remaining top 9 contracts, CryptoKitties is still larger than all combined! The contracts in this list are primarily NFT (Non-Fungible Token) platforms or gaming platforms. The utility for cryptocurrencies in these areas is growing so fast and as a result, we can see these contracts having so many transactions.

Similarly, it can be seen in table 6 that Tether: USDT is so much bigger than the other top ERC20 contracts. These ERC20 contracts mainly represent other cryptocurrencies and stable coins built on top of the Ethereum blockchain.

As we see so many transactions happening with smart contracts, they are moving the world towards mass adoption of decentralization. There is a lot of potential in this space and we will surely see other use cases emerging for smart contracts in the near future.

11 INTERPRETATION OF RESULTS

Summarizing all the research that has been done in this work, following conclusions can be drawn:

- The number of transactions and the value of ether transferred are two good indicators to understand how the network behaves. Comparing these two, we have seen that they have not followed the same path during the genesis of the Ethereum blockchain. The number of transactions has exponential growth, while the value of transferred ether fluctuates a lot.

- Looking at ether price, we have concluded that it follows the pattern of the number of transactions. This fact helped us reach the conclusion that the number of transactions may affect the price of ether or vice versa.

- We have concluded that there are 95,429,623 active addresses in the Ethereum network as of 31st August 2021, which includes external addresses, miner addresses, and all contracts.

- Focusing on transactions, we have found that almost 92.7% of them are simple ether transfers and 7.2% are contract calls. Only 0.1% are contract creation transactions.

- There is a lack of diversity in the smart contracts ecosystem. There are just a few creators of the contracts, relative to their number. Moreover, this lack of diversity exists even in the bytecode. This means that most of the contracts in the Ethereum network are identical. More particularly, only about 0.98% of the total 46M contracts are totally unique!

- We have also focused on the analysis of specific addresses, taking into consideration their holdings and the number of transactions. Generally, addresses with high ether balance are either exchanges or external users (individuals). On the other hand, taking into consideration the top 10 addresses that execute the most transactions, we have seen that exchanges are at the top of the list, as expected. However, miners have a dominant position in that list, as well. The main conclusion of this is, that, in most cases, the number of transactions in which an address is involved is inversely proportional to the ether holdings of the account.

This study also describes some key arguments that are worthy of further discussion. In this work, the main analysis is concentrated around addresses that are always measured by ether value (balance) and/or the number of transactions (traffic). We feel that additional research on the Ethereum blockchain can be done by incorporating even more parameters to yield even more reliable results.

12 DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] Njilla L Kamhoua CA Adebayo A, Rawat DB. *Blockchain for Distributed Systems Security*. Wiley, 2019.

[2] Evgeny Allen Day. Ethereum blockchain, 2019.

[3] Suveen Angraal, Harlan M Krumholz, and Wade L Schulz. Blockchain technology: applications in health care. *Circulation: Cardiovascular quality and outcomes*, 10(9):e003800, 2017.

[4] Vitalik Buterin. Ethereum whitepaper, 2022.

[5] Monika Di Angelo and Gernot Salzer. Tokens, types, and standards: Identification and utilization in ethereum. 08 2020.

[6] Xiaomin Du, Beibei Chen, Ming Ma, and Yanjiao Zhang. Research on the application of blockchain in smart healthcare: Constructing a hierarchical framework. *Journal of Healthcare Engineering*, 2021:1–13, 01 2021.

[7] Nabil El Ioini and Claus Pahl. A review of distributed ledger technologies. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 277–288. Springer, 2018.

[8] Stuart Haber and W Scott Stornetta. How to time-stamp a digital document. In *Conference on the Theory and Application of Cryptography*, pages 437–455. Springer, 1990.

[9] ADAM HAYES. BLOCKCHAIN FACTS: What is it, how it works and how it can be used, 2022.

[10] Manish Kumar. *Applications of Blockchain in Digital Forensics and Forensics Readiness*, pages 339–364. 07 2020.

[11] EVGENY MEDVEDEV. ETHEREUM IN BIGQUERY: how we built this dataset, 2018.

[12] Du Mingxiao, Ma Xiaofeng, Zhang Zhe, Wang Xiangwei, and Chen Qijun. A review on consensus algorithm of blockchain. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 2567–2572. IEEE, 2017.

[13] SATOSHI NAKAMOTO. BITCOIN: A peer-to-peer electronic cash system. *Decentralized Business Review*, page 21260, 2008.

[14] Harish Natarajan, Solvej Krause, and Helen Gradstein. Distributed ledger technology and blockchain. 2017.

[15] BAOZHUANG NIU, ZIHAO MU, BIN CAO, AND JIE GAO. SHOULD MULTINATIONAL FIRMS IMPLEMENT BLOCKCHAIN TO PROVIDE QUALITY VERIFICATION? *Transportation Research Part E: Logistics and Transportation Review*, 145:102121, 01 2021.

[16] Shaan Ray. The difference between blockchains distributed ledger technology, 2018.

[17] Anthony Sassano. Why ether is valuable, 2019.

[18] Nachiappan Subramanian, Atanu Chaudhuri, and Yasanur Kayikci. *Blockchain Applications in Manufacturing Supply Chain*, pages 57–66. 05 2020.

[19] Nitin Sukhija, John-George Sample, and Elizabeth Bautista. *Advancing the Cybersecurity of Electronic Voting Machines Using Blockchain Technology*. 07 2019.

[20] Nick Szabo. Formalizing and securing relationships on public networks. *First monday*, 1997.

[21] Kevin Werbach. *The Blockchain and the New Architecture of Trust*. MIT Press, 2018.

[22] GAVIN WOOD ET AL. ETHEREUM: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, 151(2014):1–32, 2014.