International Conference on Innovative Research in Engineering, Management and Sciences
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 4 | Issue 9 | ISSN : 2456-3307

Data Models used in Bitcoin and Ethereum Blockchain Platforms

Tinu N.S[*1]

[*1] Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India
tinuns@newhorizonindia.edu[1]

ABSTRACT

Data analytics has captured attention of both researchers as well as business organizations, since a long time now, as the knowledge or information getting analyzed and evolved is priceless in upbringing the business. Blockchain is the latest technology which is getting adopted at a faster rate due to its unique properties. This paper focuses mainly on data models, and some tools used for data analytics being used in blockchain environment. Public blockchain is an open ledger platform which allows to perform data analytics.

Keywords : blockchain, data models, UTXO, data analytics tools.

I. INTRODUCTION

Blockchain a.k.a Distributed Ledger Technology (DLT) has made its own locus in the technological grade, through its evolution since past decade. The highlight of this technology was its ability to provide tamper proof, cost effective asset transaction among multiple parties without requiring a trusted central authority. Blockchain was devised and made public by Satoshi Nakamoto in his white paper "Bitcoin electronic cash system", released in 2008[1]. The concept was in paper only for a year, after which its first application "Bitcoin" got released in 2009. Bitcoin network came into market along with the inception of first cryptocurrency based on blockchain, called the bitcoin. Currently there are quite a lot blockchain based cryptocurrencies, more than 1600 in count and still growing, called alt-coins. Not only the digital currency world but several other usecases of blockchain came into existence[]. These developments have enkindled public interest in Blockchain technology.

Along with the inquisitiveness about the features provided by the blockchain technology, it got adopted fast in the research area and application development zones related to supply chain, healthcare, e-voting, identity management, asset tracking etc..., to name a few. While it's arduous or too early to predict the longer-term impact of Blockchain, it is safe to mention that it would have several vital impacts on various existing applications. For the usability purpose blockchain technology has evolved into two major platforms, public and private. A hybrid version of public and private blockchain is the consortium blockchain.

Public blockchain are also called the permissionless blockchain, where anyone can participate in the network or leave the network when required. Participants would be anonymous and open to all. In a private blockchain, participation is permissioned, usually used by industry and organization. Read, write operations need permission in a private blockchain network. This paper focuses on data models used by the public blockchains and the data analytics tools used for efficient data analysis. The rest of the paper is arranged into four sections, section II gives an insight to the brief history of blockchain, section III on the data models used by public blockchains, section IV on the data analytics tools used and finally section V providing a conclusion to the topic of this paper.

II. HISTORY OF BLOCKCHAIN IN BRIEF

In 1983 the concept of e-cash protocols that can generate anonymous cryptographic electronic money was introduced by David Chaum and Stefan Brands. The idea of blockchain technology was described as early in 1991,by research scientists Stuart Haber and W. Scott Stornetta, as they introduced a computationally practical solution for time stamping digital documents, so that the documents could not be backdated or tampered with.The system used cryptographically secured chain of blocks to store the time-stamped documents. In 1992 the merkle tree concept were incorporated into the design, the concept was published on 1987, making it more efficient by allowing several documents to be composed into one block[2]. (however, the technology went unused and patent lapsed in 2004). In 2004, computer scientist and cryptographic activist Hal Finney, presented the system called RPoW(Re-usable Proof of Work). The system is operated by receiving a non exchangable or non-fungible hash cash based proof of work token and in return created an RSA-signed token, that can then be transferred from person to person. RPoW solved double spending problem by keeping the ownership of tokens registered on a trusted server that was designed to allow its users throughout the world to verify its correctness and integrity in real time. RpoW can be considered as an early prototype and a significant early step in the history of cryptocurrencies. In late 2008, a white paper introduced the concept of decentralized peer to peer electronic cash system called the Bitcoin, was posted to a cryptography mailing list by a person or a group named Satoshi Nakamoto.

III. BLOCKCHAIN DATA MODELS

The record keeping model in public blockchains can be often broadly categorized into two, unspent transaction output (UTXO) based (e.g., Bitcoin, Bitcoin Cash, Litecoin) and account/balance based (e.g., Ethereum) blockchains[3]. In both types of blockchains, a data block consists of a finite number of transactions. The goal achieved is to keep track of account balances, based on consensus mechanism, but the transactions differ in their characteristics in both models. The two-different types of blockchain transaction data models is briefly discussed below along with its pros and cons.

A. The Unspent Transaction Output Based Blockchain Data

The unspent transaction output (UTXO), as the name depicts is the unspent or leftover cryptocurrency available in the wallet. A wallet is basically a software application that provides an interface to store, access and manage, basically, cryptocurrencies as well as tokens (other assets) with the help of asymmetric keys (private and public keys). Wallets could be classified into five types like, online wallet (hosted and non-hosted),desktop wallet, mobile wallet, paper wallet, hardware wallet[4].These wallets can be used to manage cryptocoins or any UTXO based blockchains are the earliest and most valued blockchains: Bitcoin alone constitutes 45-60% of the total cryptocurrency [5] market capitalization. In UTXO blockchains each data block contains a set of (financial) transactions that encodes the transfer of coins among multiple parties. The basic structure of a transaction is shown in fig 3.1.

Version	Input counter	Inputs	Output Counter	Outputs	Lock Timestamp

Fig 3.1. Structure of a Transaction

The version field is the 4 byte version number of the protocol used by the application. Input counter is the no: of input UTXOs involved in that transaction. Inputs field gives the input UTXO's involved to make the required output UTXOs. Output field includes the output UTXO and output counter gives the no:of Output UTXO's. Lock Timestamp stores the time at which the transaction have been initiated. Each transaction has at least one input and one output. The fig 3.2 shows details of a part of the network transaction block that has 11 addresses(a1 to a11) and 6 Bitcoin transactions(t1 to t6)[]. Block boundaries are not displayed. Each transaction has at least one input and one output. Coins at addresses a7 and a8 remain unspent. The difference between input and output amounts (e.g., 0.2B at t1) are collected as the transaction fee. The equation for a New UTXO is as shown below,

New UTXO = (Sum of UTXOs in the transaction) – (Transaction amount) – (Transaction fee)

The above equation can be made understandable with an example. Let the sender be Alex, who has 100 bitcoins. Though the balance amount is shown as just one value ie 100 here, but how Alex got his balance is actually through several input UTXOs that accounted to his address.The input UTXO's can be four UTXOs worth 25 bitcoin each, two UTXOs worth 50, or a set of UTXOs valuing 34, 18, 43, and 5 bitcoin. The balance amount associated with a particular account is calculated by adding all the input UTXOs to that account. Alex sends 28 bitcoins to Alice, and say 2 bitcoins counts to transaction fees, thus Alex need to spent 30 bitcoins in order to perform that transaction. Thus new UTXO in this transaction is calculated as (100)-(28)-(2), and the value will be 70.The input UTXOs might also be the output UTXOs of some previous transaction. For example in Fig. 3.2, the transaction t4 have got two input UTXO's and three output UTXO's, but the two input UTXO's were the output UTXO's from addresses a2 and a3. There are three rules followed in order to shape data on UTXO blockchains. These rules are according to the design choices by Satoshi Nakamoto in Bitcoin [1]. The three rules are:

Source Rule: Input coins from multiple transactions can be merged and spent in a single transaction (e.g., the address a5 receives coins from t1 and t2 to spent in t4 in Fig. 3.2), or spent separately (e.g., in Figure 3.2, a9 spends coins received from t3 and t4 in t5 and t6) in different transactions.

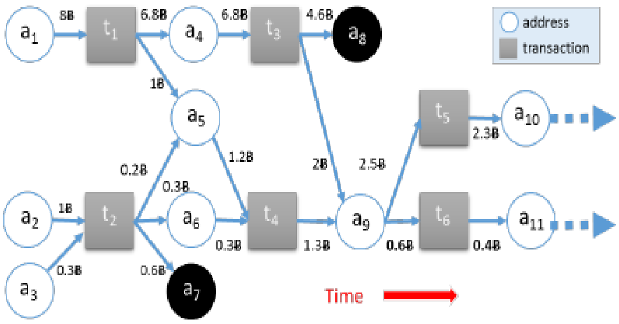

Fig. 3.2. A Portion of Bitcoin Transaction Network

Mapping Rule: Each crypto currency payment must show proof of their account funds by referencing a set of previous outputs. This allows anyone to trace back the history of payments, but it is not always possible to locate where a specific coin originated from. The reason is, each transaction lists a set of inputs and outputs, separately. For example, t2 has two inputs and three outputs, but an explicit mapping between inputs and outputs does not exist. Coins flowing to a9 might have come from either a4, a5 or a6, or from few of them. As a result, a transaction can be reflected as a lake with in-flowing rivers, and out-flowing rivers (i.e., emissaries).

Balance Rule: This rule states that, coins received from one transaction must all be spent in a single transaction. Any amount that is not sent to an output address is considered to be the transaction fee, and rewarded to the miner who validates and creates the block. The difference between input and output amounts (e.g., 0.2B at t2) are collected as the transaction fee. In order to keep the change, or the returned UTXO's, the coin spender can create a new address (i.e., change address) and send the remaining balance to this new address. Another option is to use the spender's address as one of the output addresses, and re-direct the balance. The reuse of the spender's address is discouraged. Thus, most nodes appear in the graph only two times; once when it receive coins and once when it spends. The change address, if created, becomes the new address of the coin owner. Due to these rules, the unspent transaction output based blockchains should be considered as forward branching trees, rather than networks.

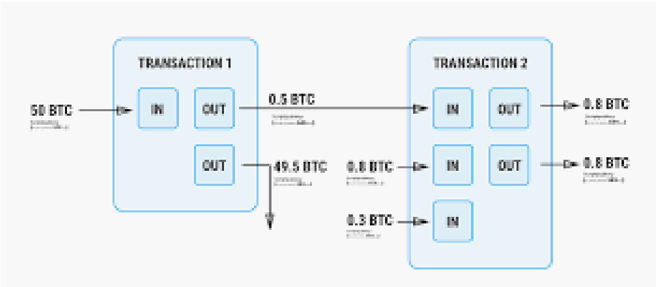

Fig. 3.3. The UTXO or Unspent Transaction Output Model

The UTXO model has potential benefits such as, it simplifies the accounting methods of blockchain, it is easy to keep track of coins, it does not allow double spending or stop from spending non existing coins. This model makes parallel processing easy, thus improves scalability. UTXO model is stateless, it doesn't store the states, thus makes it easy for users to use new addresses for every transaction, this improves privacy to a certain extent. Bitcoin also supports multiple scripting types which allow it to process complex payment logics. it allows for Simple Payment Verifications (SPV) on the network, which allows the wallets to interact with the blockchain environment in a decentralized and trustless manner without having to download the entire Bitcoin blockchain, thus significantly reducing storage.

Disadvantages of UTXO model are revealed when it is applied to a more complex turing complete platforms such as ethereum. This model is weak in programmability and complex computations. As the input UTXOs increases, it becomes difficult to verify the growing scripts and store the witness data. The Bitcoin transaction started with 50 bitcoins, and halves the block reward every 4 years. This geometric series will result in a total of 21 million bitcoins. The global state at any given time is the set of all spendable UTXOs. Most crypto-currencies have the same data model as Bitcoin.

B. Account Based Blockchain Data

Account based blockchain used in ethereum is similar to how accounts are maintained in banks, an address can spend a portion of its coins and keep the remaining balance. Thus a transaction has exactly one input and one output address. Even though address creation is free, mostly a single address is used to receive and send coins several times. Ethereum blockchain was released in 2015 by Vitalik Buterin[6]. The main idea behind ethereum environment is to build and run smart codes(known as smart contracts) on blockchain network. The account based blockchains use two types of addresses; one is externally owned addresses (governed by users) and the other, contract addresses (governed by smart contract code). A transaction to upload the Smart Contract code to a contract address is typically initiated by an externally owned address (i.e., user address), but it can also be initiated by a contract address. The code at the address is stored in the Blockchain and replicated at all Blockchain nodes. In other words, uploading the contract forces other nodes to store the code locally. Account based blockchains have two types of transactions. One is the transactions that happens using cryprocurrency, such as Ether on Ethereum, between two addresses. This can be modeled with a directed edge between the two addresses. The second type, internal transactions, are created when smart contracts change states associated with addresses. Asuume a sell order issued by address a1 to a Smart Contract where the to parameter is a2 and the value parameter is 3 tokens. The Smart Contract creates an internal transaction that transfers 3 tokens from a1 to a2. Internal transactions can be discovered in two different ways: by parsing the transaction's message and updating states associated with a1 and a2 manually, or by running the transaction message through the smart contract code and observing the states and logs created during execution. Another option is to run a full Ethereum node and execute every contract transaction. This is costly in terms of time and resources. The parsing option is easier as it does not require code execution. However, the parsing method cannot discover transaction failures (due to reasons such as insufficient gas), and create internal transactions that do not actually exist.

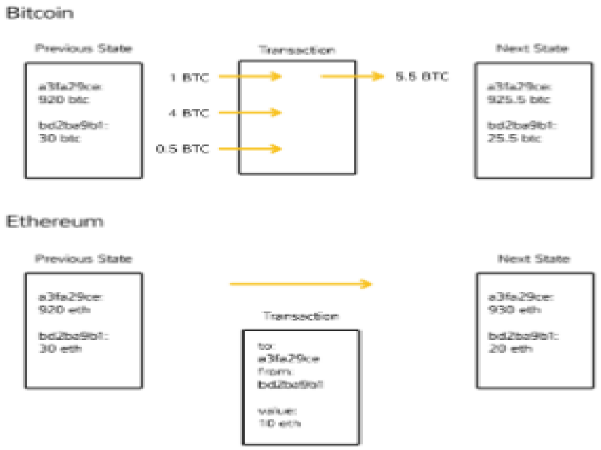

Fig. 3.4. UTXO VS Account Model

Benefits of account based model are simplicity in the development of complex smart contracts that involves multiple parties and state information. This model allows to save space, as it only needs to make one reference and signature that produces one output, contrary to UTXO design. Possible disadvantages are the scalability issues and need for additional evaluation to check for correctness of transactions.

IV. DATA ANALYTICS TOOLS

Blockchain data analysis is gaining interest in the market now, both for business applications and in research area. In this section, prominent data analytics tools are discussed in brief. The most widely used tool in Blockchain data analysis is Blocksci, which is an open source platform, that allows fast and sensitive analysis of data stored [7]. BlockSci's core infrastructure is written in C++ and optimized for speed. It works well with python as well as Jupyter interface. Another tool used is Biva, for network visualization and data analysis. This area is still undergoing updations and rapid developments.

V. CONCLUSION

The blockchain technology has got wide acceptance in the society as it can have major contributions to many applications that can have direct impact. Research works are happening at an intensive level to keep up the expectations from the market. Even many MNC's have invested in this technology foreseeing its impact. The latest contributor is the Google with its BigQuery database, which is a highly scalable enterprise data warehouse used for productive data analysis. This paper mainly focused on the data models used with Bitcoin and ethereum network, its advantages and disadvantages.

I. REFERENCES

[1] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2008.

[2] Ralph C. Merkle. 1987. A Digital Signature Based on a Conventional Encryption Function. In CRYPTO ’87: Proceedings of the 7th Conference on Advances in Cryptology. 369–378.

[3] Cuneyt Gurcan Akcora, Matthew F. Dixon, Yulia R. Gel, and Murat Kantarcioglu, “Blockchain Data Analytics,” IEEE Intelligent Informatics Bulletin December 2018 Vol.19 No.

[4] Tobias Bamert, Christian Decker, Roger Wattenhofer, and Samuel Welten. 2014. BlueWallet: The Secure Bitcoin Wallet. In STM ’14: Proceedings of the 10th International Workshop on Security and Trust Management. 65–80.

[5] F. Tschorsch and B. Scheuermann, “Bitcoin and beyond: A technical survey on decentralized digital currencies,” IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 2084–2123, 2016.

[6] G. Wood, “Ethereum: A secure decentralised generalised transaction ledger,” Ethereum project yellow paper, vol. 151, pp. 1–32, 2014.

[7] H. Kalodner, S. Goldfeder, A. Chator, M. M¨oser, and A. Narayanan, “Blocksci: Design and applications of a blockchain analysis platform,” arXiv preprint arXiv:1709.02489, 2017.