ISSN No:-2456-2165

A Survey on Collaborating Blockchain and Big Data Exchange

Ummu Habeeba K
PG Scholar, Department of Computer Science and Engineering, MES College of Engineering, Kuttippuram

Shajeesh K. U
Assistant Professor, Department of Computer Science and Engineering, MES College of Engineering, Kuttippuram

Abstract:- **Data has become the most important and valuable asset recently. Researchers, analysts and engineers from every field need real dataset in order to analyze market trends, train their systems etc. Existing systems for data exchange can broadly be classified in to two- paid and free. A lot of paid data exchange markets are available. In those paid systems, data owners will provide their dataset and data consumers find their interested dataset. In these systems, it is much difficult to protect privacy and copyrights. Also, special techniques are required to maintain the data services for large organizations who are having big dataset like hospitals, banks, government departments etc. A decentralized system for data exchange will solve most of those issues mentioned above. Collaboration of big data and the trending technology of nowadays called the blockchain will be an innovative solution. In this atmosphere, every members exchange data in a peer-to-peer way. One can audit the use of dataset in order to protect the copyright and privacy as the blockchain records all the documents. No need of a third party is one of the attractive advantages.**

Keywords:- *Blockchain, Data Exchange, Decentralization, Collaboration.*

I. INTRODUCTION

In this modern world, the data is becoming an important and most valuable asset. Researchers, engineers, analysts etc from different field like industry, academy, business etc need real dataset for different purposes like analysis, training etc. Thus, the importance of data exchange is increasing and increasing nowadays. So, data exchange can be considered as a new field. For example, in China, there are a lot of data exchange markets which are run by government itself. Companies are also running such centers in China.

There are three entities in a simple data exchange market. They are data owner, customer and a data exchange centre. A lot of issues are there in the existing big data exchange markets as these markets are following the idea of traditional markets. But data is a special material because it can be reproduced illegally. Some data may be sensitive to privacy and confidentiality [1]. There is need of such a system for data exchange where the transacting parties benefit from the market, transact without the fear of single-point-failure and without the help of an intermediate third party.

The collaboration of blockchain technology and big data exchange will help in building such market for data exchange. Blockchain is a distributed ledger maintained by all members in a peer-to-peer way removing the necessity of third parties. Each and every transaction logs are recorded and saved in blockchain and every user is able to see it. It also cannot be changed and hence anyone trying to change anything in the transaction will not be able to do so. This paper contains the topics studied for the collaboration of big data and blockchain.

II. BIG DATA AND CHALLENGES

Big Data is related to almost all aspects of human activity like simply recording an event to complex procedures like analyzing, researching, designing, products delivery to the final consumer etc. It is also referred to as Data Intensive Technologies. It has become a new technology trend in science, industry, business and almost all fields. Big Data is characterized by following properties: Volume, Variety Velocity, Value and Veracity. Volume is the amount of data being produced and consumed. Velocity is the speed with which the data is being produced and consumed. Volume, Velocity and Variety make up the native/original Big Data properties called as the 3Vs of big data. Value and Veracity are added to it additionally in recent years.

Science already has found different ways to handle the challenges related to the huge volume of data. Scientific research basically aims to validate different hypothesis by collecting data in passive observation or active experiments [2]. Following are the challenges related to handling the big data- searching, sharing, capturing, analyzing, storing, transferring and presenting. A useful solution has been put forward by Google. Google has divided the data and assigned the data to a network having more and more computers connected. Each computer in the network will be performing an assigned task. These computers are also responsible in fetching the results back from each of the other computer. Finally the results are merged to get the final dataset [3].

ISSN No:-2456-2165

III. DISTRIBUTED LEDGER

A distributed ledger helps to eliminate the necessity of a central third party by maintaining a decentralized pattern across different locations and people. It is basically a ledger of any activity like transaction or contracts. Hence it establishes interoperability of networks guarantying transparency, irreversibility, distribution and anonymity. Here, different stakeholders can transact each other without trusting one another. Both the centralized and distributed ledgers are shown in Fig. 1 [4]. In centralized ledger systems one authority is there to control the ledger. They are prone to cyber-attack. In a distributed ledger system, every participant is having a copy of the ledger. With every new transaction, the ledger is updated accordingly and new ledger is produced [5]. But distributed ledgers are difficult to get attacked because attackers need to attack all the copies of ledger which are present in the hands of the members of the network at the same time for that attack to be successful.

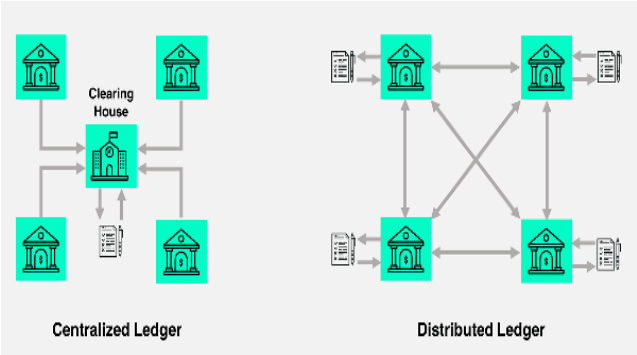

Fig 1:- Centralized and distributed ledgers.

IV. TRANSACTIONS

A transaction is simply an agreement between two parties; they may be buyer and seller, to exchange goods or services. It can also be described as the transfer of ownership between the transacting parties who may be seller and buyer and hence establishes the details on how an organization should work. To document this transfer, blockchain is used. A transaction normally contains an identifier, data and a timestamp. The creation of transaction outside or inside the blockchain network denotes the start of the lifecycle of a transaction. Then, the sender signs the transaction. It is then broadcasted to the network, where the nodes in the network are responsible for verification and validation of the transaction. At last, the buyer receives the ownership of the asset that is being transacted [5].

Only after the process of verification and validation by the members of the network, blockchain records and saves a transaction. Verification is the process of checking whether the nodes involved in the transaction procedure are legitimate or not. After the validation step, network achieves consensus over the state of blockchain. Every validator nodes involved in the validation process receives reward amount based on the protocol implemented. The consensus algorithm decides which block is going to be attached to the blockchain. The final block of consensus is then concatenated with the blockchain by cryptographic means.

V. SECURE ELECTRONIC TRANSACTION (SET)

Transactions have become online nowadays. It is done with the help of debit or credit cards issued from the banks. To ensure the integrity and security of electronic transactions done using these cards, Secure Electronic Transaction or SET is used. SET is not a system to do payment. It is a security protocol applied on those payments which uses different cryptographic techniques to secure payments done using the debit or credit cards online. Different development organizations like Visa, Mastercard, Microsoft support the SET protocol by providing Secure Transaction Technology (STT) and NetScape which provided the SSL technology. SSL stands for Secure Socket Layer (SSL) [6]. SET basically has the following participants, which is shown in Fig. 2 [7]:

1. Card holder- customer
2. Issuer- customer financial institution
3. Certificate authority
4. Merchant
5. Acquirer- merchant financial institution

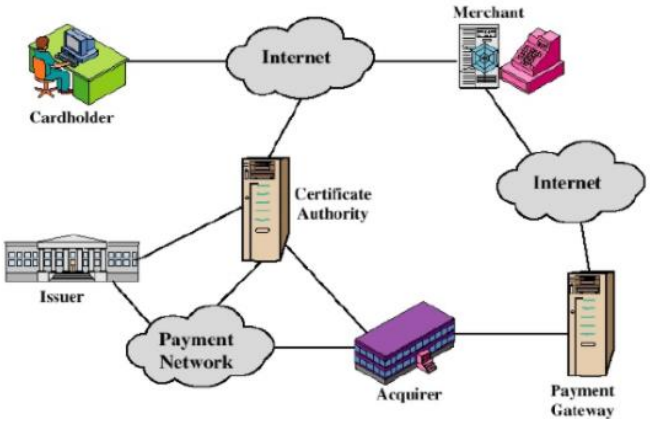

Fig 2:- Participants in SET

Following is the working of SET. Consider, transacting parties have SET enabled servers [8].

1. The customer is having bank account having Mastercard, Visa etc as the certificate authority. as the issuer institution. Mostly, bank will be the issuer of credit or debit cards.
2. The customer gets a digital certificate which can be used for online shopping and other transactions.
3. Third-party merchants also get their public key and public key of bank through those certificates.
4. The customer shops online with the help of mobile phone, from websites, applications etc.

5. Only after confirming the merchant's validity with the help of the certificate, browser sends the order information.
6. This message will be encoded with the public key of merchant, the payment information already encrypted with the bank's public key (merchant is not able to read this), and information that ensures the payment can only be used once for this particular orders.
7. The merchant also verifies the customer identity with the help of the certificate and sends the order message to the bank. This includes the public key of bank, the customer's payment information (which can't be decoded by the merchant), and the merchant's certificate.
8. The merchant and the message are verified by the bank. Using the digital signature on the certificate with the message, bank verifies the payment part of the message. Then, bank signs digitally and sends authorization to the merchant, who can then finish the order.

Then the purchase order is implemented by shipping the requested order and the transaction is completed.

VI. BLOCKCHAIN

A blockchain is actually a growing list of block that consist the details of a transaction, which are linked by cryptographic methods. It was introduced when Bitcoin faced the double-spending problem [5]. Each block in the chain is having a list of a transaction data, a hash to the previous block and a timestamp. Cryptographic hash is used to identify each block and these blocks are referenced by hash of the block that came before it. This creates a link between the blocks creating a chain of blocks and hence called as blockchain. A typical blockchain is shown in Fig. 3. The design and nature of blockchain reveals that it is immutable and tamper resistant. And they are also resistant to modification of data [9].

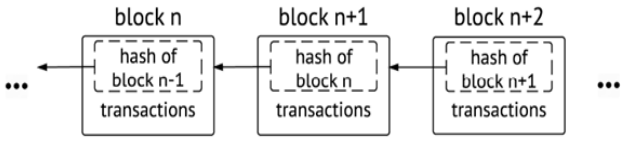

Fig 3:- A typical blockchain

The concept of blockchain was first given by a person or a group of people who work under the name Satoshi Nakamoto in 2008. Looking to the following steps one can easily understand the working of the blockchain:

1. A pair of private/ public keys is used by the users to interact with the blockchain. Each member in the blockchain is addressed by their public keys and they use their private key in order to sign their transactions. Authentication, non-repudiation and integrity are achieved by using this asymmetric cryptographic system then, this transaction is broadcasted to the network.
2. The transactions are validated by the neighboring nodes and invalid transactions are discarded when one is found. And this transaction is spread across the network.
3. A timestamped candidate block is created which is called mining. This candidate block is actually the one that was collected and validated by the network using the process above during an agreed-upon time interval. This block is broadcasted back to the network by the mining node.
4. The nodes verify that the block contains valid transactions and add the block to their chain. If the block is found to be invalid, it is discarded on the spot. This denotes the end of a cycle.

Note that this is a repeating process. Fig. 4 represents the working of blockchain [10].

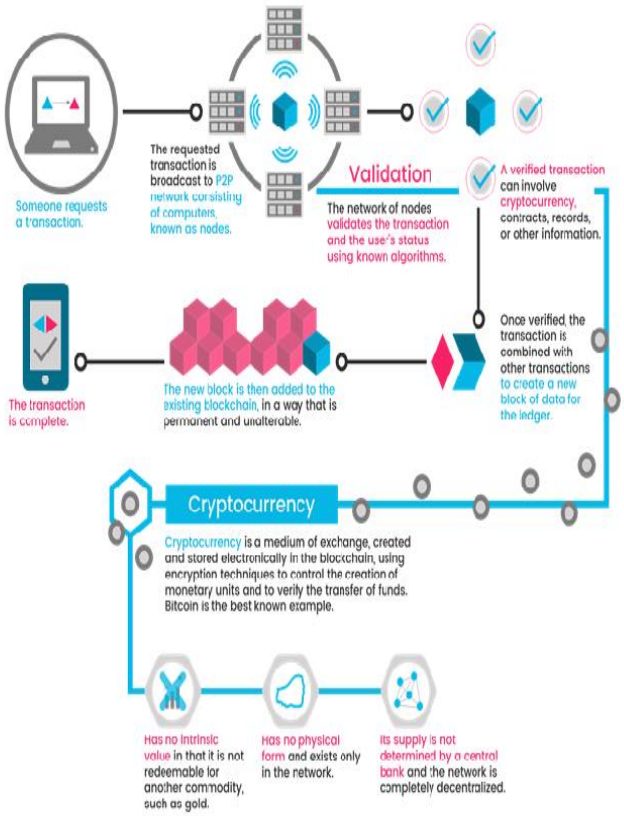

Fig 4:- Working of blockchain

The third party agency which may be causing single point of failures can now be removed from the system thereby avoiding the risk of leaking of data by these intermediaries. A smart contract can also be incorporated to this which can automatically guarantee the rights and benefits of both participators of every transaction [11].

There are broadly two types of blockchin; public and private. A public blockchain has no access restictions. Examples of these kind of blockchain are Ethwereum blockchain and bitcoin blockchain. Private blockchain is

ISSN No:-2456-2165

permissioned; that is, one can enter in to the blockchain nly after receiving invitation from the administrators of the network. The administrators also assign the validator and participant works to the members of the network.

VII. COMPARISON TABLES

The following table illustrates the differences between the centralized database and the distributed ledger.

Centralized database	Distributed ledger
Internal and external reconciliation required	Consensus on data
Restrictionless	immutable
Single-point-failure	Distributed across nodes
Centralized control	Decentralized control
Middle man and third parties involved	Peer to peer relationship
Cryptography as an after thought	Cryptographic verification
Actions taken on behalf of others in case of disputes	Cryptographic authentication and authorization
Backups must be provided manually	Availability and resiliency increases with node count

Table 1:- Comparison of Centralized Database and Distributed Ledger

The following table illustrates the differences between a private and public blockchain in detail.

Public blockchain	Private blockchain
No restrictions in joining to the network	Only invited members can join the network
Transmission power is equal for each node	New transactions can only be created by some nodes
Transaction accomplishment speed is low	High speed of transaction accomplishment
Transaction cost is higher	Comparatively, transactions cost is cheaper
Adding a new block based on proof-of-work, proof-of-stake consensus protocols	Adding of new block is initiated by pre-approved participants
Anonymous	Nonymous
Members no need to trust each other	Requires trust among the members
Consumes more energy	Consumes less energy

Table 2

VIII. CONCLUSION

The working of current existing systems revolves around a central third party. In those systems, data exchange markets act as the central third party. Those having the dataset will display the details of their dataset in to the market. Customers will also be submitting the details of the dataset that they need. Market will cross match those details submitted by both data owners and customers and help the customer find the required dataset. The existing system suffers from the following serious issues:

1. Fully dependent on a trusted third party
2. Security is entirely dependent on the third party
3. Single point of failure
4. Expensive
5. Bottle neck problem because every transaction is going through the trusted third party
6. Rely on public judicial systems in case of disputes

Hence there is need of such a system for data exchange where the transacting parties should be able to transact without above mentioned problems. Introduction of the trending technology called the blockchain into this system is the suggested solution. This will provide a decentralized and secure solution for big data exchange.

REFERENCES

[1]. Jinchuan Chen, Yunzhi Xue, "Bootstrapping a Blockchain Based Ecosystem for Big Data Exchange", IEEE 6th International Congress on Big Data, 2017.
[2]. Yuri Demchenko, et. al, “Defining Architecture Components of the Big Data Ecosystem”, IEEE International Conference on Collaboration Technologies and Systems(CTS), 2014.
[3]. Pradeep S, Jagadish S Kallimani, “A Survey on Various Challenges and Aspects in Handling Big Data”, IEEE International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2017.
[4]. "The Difference between Blockchain & Distributed Ledger Technology" [Online]. Available: https://tradeix.com/distributed-ledger-technology/.
[5]. Daniel Burkhardt, Maximilian Werling, Heiner Lasi, “Distributed Ledger: Definition & Demarcation”, IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), 2018.
[6]. "Secure Electronic Transaction (SET) Protocol" [Online]. Available: https://www.geeksforgeeks.org/secure-electronic-transaction-set-protocol/.
[7]. "Participants in the SET System" [Online]. Available: https://www.slideshare.net/mobile/AgnChomentauskait/afp-summit-secure-electronic-transactiom-setstohner.

[8]. Shiyong Lu, Scott A. Smolka, “Model Checking the Secure Electronic Transaction (SET) Protocol”, International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2009.
[9]. K. Christidis, M. Devetsiokiotis, "Blockchains and Smart Contracts for the IoT ", Electrical and Computer IEEE Conference, May 2016.
[10]. "Smart Contracts: The Blockchain Technology That Will Replace Lawyers" [Online]. Available: http://blockgeeks.com/guides/smart-contracts/.
[11]. Guy Zyskind, Oz Nathan, Alex ’Sandy’ Pentland, “Decentralizing Privacy: Using Blockchain to Protect Personal Data”, IEEE CS Security and Privacy Workshops, 2015.