Blockchain-based Secure Big Data Storage on Cloud

Manikandan D, Valliyammai C, Karthika RN

Abstract: ***In the cryptocurrency era, Blockchain is one of the expeditiously growing information technologies that help in providing security to the data. Data tampering and authentication problems generally occur in centralized servers while sharing and storing the data. Blockchain provides the platform for big data and cloud storage in enhancing the security by evading from pernicious users. In this paper, we have discussed the exhaustive description of blockchain and its need, features and applications. Analysis of blockchain is done for different domains such as big data, cloud, internet of things and mobile cloud where the differences V's are compared with big data and blockchain. SWOT (Strength Weakness Opportunities Threats) analysis is performed to address the merits and limitations in blockchain technology. The survey in aspects of data security, data storage, data sharing and data authentication through blockchain technology is done and the challenges are discussed to overcome the problem that leads in big data and cloud storage. The detailed comparative analysis proves that the blockchain technology overcomes the problems in big data storage and data security in cloud.***

Keywords **: *blockchain, bigdata, cloud, SWOT, data security***

I. INTRODUCTION

On rapid development of information technology in security, blockchain leads the major role in which it undergoes decentralized peer-peer system. Blockchain technology is a distributed public ledger [10] in which it records all the transactions details held, by avoiding the third party intermediate [24] and it stores the huge amount of data in one single block and it also provides more security to the data by hashing techniques where data loss may not occur. Blocks are interconnected to form a chain structure to frame a blockchain technology in which every block consists of both hash value and previous hash. Blocks are verified by the previous hash of the current block [22] with the hash value of the previous block and it helps in identifying whether the block is [26] malicious or not. Data in the blockchain undergoes immutability features in which once the data is updated into the block, it cannot be changeable so data changes will not occur. Blocks are open access [14] to the user connected to the blockchain network in which the user data is not open. Blockchain maintains consensus algorithms such as proof of work and proof of stake to store sensitive data. Big data and cloud undergo some issues in security, storage, sharing and authenticating the data. Blockchain is analyzed to resolve these issues where the research challenges are identified from the survey. In this paper we have contributed the analysis of the blockchain network in perspective of big data storage and cloud security where SWOT analysis is also taken to know benefits and loss of credits. This paper is organized as the detailed description of blockchain network evolution, needs and features SWOT analysis to understand the working of blockchain and blockchain based big data storage using cloud computing and addresses the challenges in blockchain.

Revised Manuscript Received on September 25, 2020.

***** Correspondence Author

Manikandan D, Computer Technology, MIT Campus- Anna University, Chennai, Tamil Nadu, India. Email: pugalanthimanikandan40@gmail.com

Valliyammai C, Computer Technology, MIT Campus- Anna University, Chennai, Tamil Nadu, India.
Email: cva@mitindia.edu

Karthika RN*, Computer Technology, MIT Campus- Anna University, Chennai, Tamil Nadu, India.. Email: karthirn@mitindia.edu

II. HISTORY OF BLOCKCHAIN NETWORK

Online transaction techniques help to transfer the money where it is made easier to use with devices like Mobile, IoT devices by avoiding personal contact with the bank. Every transaction through online is maintained secretly, even hackers hack the transaction detail of the user. In 2019 RBI releases, 71,500 crore had made fraud through online transactions and net banking. Blockchain technology helps in providing security to the data in which BPay undergoes the fair transaction and no amount is detected from the user for the transaction. The history of blockchain is depicted in fig1.

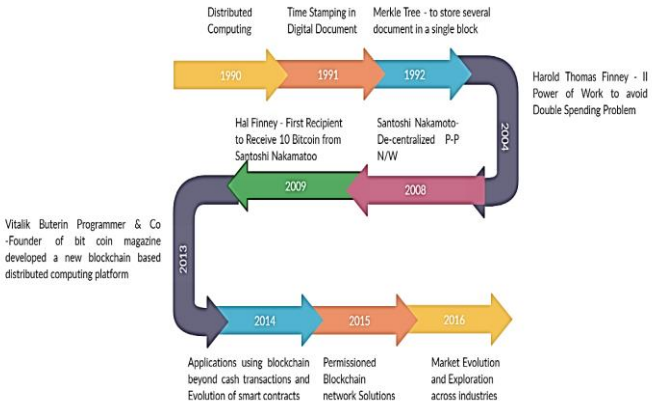

Fig 1. History of Blockchain

III. NEED FOR BLOCKCHAIN

A. High Transaction Fee

Centralized server acts as intermediate between both sender and receiver. Consider an online transaction where the sender sends the money to the receiver in which the third-party intermediate such as the bank consumes the huge transaction fee to verify the transaction detail and it gets stored and organized by the central server organized by a third party.

B. Double Spending

Double spending is one of the issues that arises while online money transactions in which this happens when the user sends the money to the two receivers without knowing the amount of that account. Fig 2 depicts about an online transaction in which the user A has two digital tokens, A transfers two digital tokens to user B (there is some problem in transaction) and again A transfers again those two digital tokens to user C and then the double spending occurs.

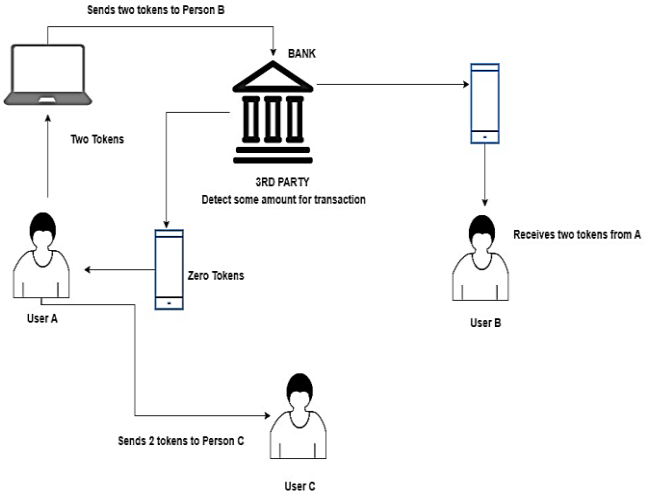

Fig 2. Double Spending Problem

C. Net Frauds

All third-party intermediaries follow centralized servers where unauthorized users can steal the details of the user while online transaction through internet banking. Data theft is easily hacked by the hacker. In this world, fraud works mostly occur through the internet where data security is not maintained properly in the third-party server. Data modification and Data tampering may occur while data sharing through the internet.

D. Poor Data Recovery

Data recovery is an advantage for the user to gain the lost data where in centralized server, data loss may occur frequently while storing and sharing the data. Data gets destroyed and it can't recover easily. Data recovery is poor in the centralized server.

IV. BITCOIN Vs BLOCKCHAIN

Bitcoin is one of the digital currencies in which this helps to send and receive the money across the world in which it undergoes decentralized peer-peer system. Transaction fees in bitcoin are minimum, when compared to other third-party intermediaries such as banks and government sectors. It doesn't include a third-party system and it is easy to transfer the coins very fast and cheap. Personal information of the user is hidden in the bitcoin transaction and it provides security to the data by cryptographic system.

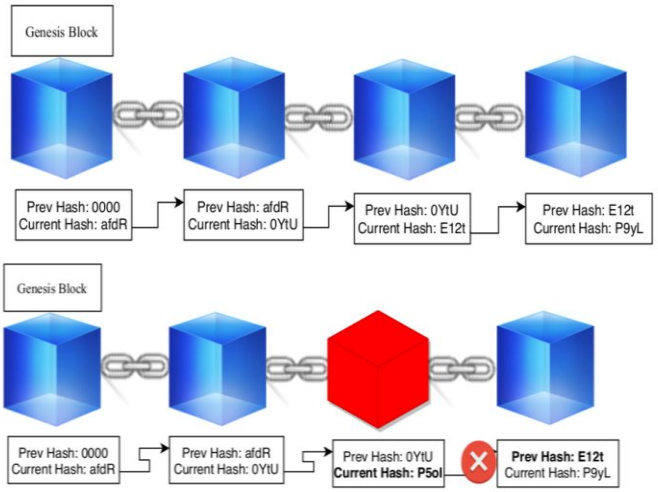

Fig 3. Overview of Blocks in a Blockchain

The information about bitcoin transactions is stored inside the blockchain. In which it maintains immutable records of public distributed databases where the records contain transaction details and these records are stored as a block in blockchain. The figure 3 shows the overview of blocks in blockchain. Each block in the blockchain has both hash and previous hash value where hash comprises alphanumeric value to identify the block in blockchain. For the same file hash value will be the same so it can be easily identified and tampered Nonce is a random value used to vary the hash value based on the given data. The first block of the blockchain is named as genesis block in which the parent block doesn't have the hash value and it is denoted as hash which is shown in fig 4.

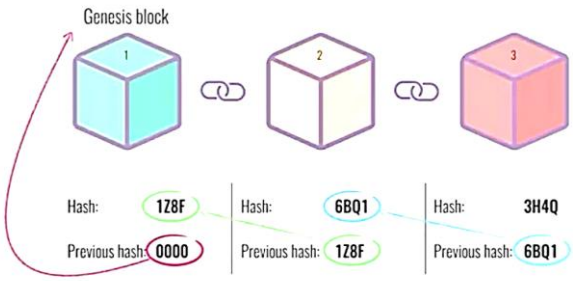

Fig 4. Genesis Block

V. BLOCKCHAIN FEATURES

A. SHA 256 Encryption

Secure Hash Algorithm takes a user data as an input and it returns an alphanumeric output of 64 bit. Hashing is the one-way function once the data is hashed then it can't be modified or changed where decrypting back of the hash value is impossible. Hash value is unique for the same data, so encryption algorithm is also performed in which it can be any message or file and it is impossible to hack the original data. In blockchain while the hash value of date is tampered while mining current block check with previous hash value if it doesn't satisfy then the block in the blockchain gets rejected.

B. Private and Public Key

Cryptographic technique consists of both public key and private key in which it helps to encrypt and decrypt the message. In which a private key is kept secure with the user in the blockchain network. Public Key is shared to all the users in the network in which except the user, all are considered as miners. Transaction details of the user are collected and it gets hashed and then it is encrypted by using the AES encryption algorithm. After hashed data is encrypted then it is initiated in the network then the mining process takes place in the blockchain network. After the data gets mined, it gets converted into the block. Once the data is converted into blocks, it can't be hacked easily until it attains a 51% malleability attack.

C. Distributed Public Ledger

Blockchain is a distributed public ledger in which it attains a decentralized system where digital data is recorded and it is open to all the users connected in the system. Data gets verified by each nodes (i.e. miners) connected to the network to prevent tampering of data. Consider 3 users in the network A,B,C in which A sends 5 bitcoins to user D who is not connected in the network. In which user sends the information of both sender and receiver address with transaction detail are hashed using SHA 256 algorithm and the hash value is open to all in the network (i.e. B, C) node mines the hash value and it converts the hash data into blocks.

D. Proof of Work

Proof of Work is one of the main consensus algorithms that help in mining the user data. It helps to determine the nonce value for the block using a hash algorithm. User data (i.e. transaction data or sensitive data) are distributed across the blockchain network. PoW produces the puzzle to mine the data in which miners in the network try to crack it. Miners who solve the puzzle first will be rewarded. After the puzzle gets solved, miners mine the data by identifying nonce values to generate the hash value for the data. Target value is fixed by the miner where hash value may start with 0 to n in which it takes a huge amount of computing power to mine the data where every block takes at least 10 minutes to mine one single block due to difficulty level of puzzle. Once the data is verified by miners, it gets converted into a block and added in the blockchain network.

E. Mining

Mining is a one of the main processes held in a blockchain network in which user encrypted data is decrypted and miners validate the hash value of the block. Consider there are 'n' numbers of the miners in the network, if user A sends data to the user B then the transaction detail is automatically generated and the data is hashed and initialized by the user. Miners mine the data by verifying the hash value of the block by solving the puzzle in which miners who solve the puzzles will get the reward of 12.5 bitcoins. Once all the miners mine the data, then the block gets converted where it validates by finding the current hash value of the block with previous hash value of next or previous block.

F. Immutable

Blockchain works on the immutability feature in which the data modification is not possible. Once the user's data is converted into blocks then the data cannot be accessed by unauthorized users and data privacy and security is maintained by the hash value and it cannot be solved by the malicious user.

VI. BLOCKCHAIN IN DIFFERENT DOMAINS

Blockchain network is connected with the 'n' number of miners who mines the transaction. Proof of Work (PoW) introduces the puzzles to mine the block. Miner who solves the puzzle will mine the block in which this process goes in chronological order where blockchain helps in different domains.

A. Blockchain in Big Data

Big data deals with the huge amount of data and it obeys the four V's such as velocity, veracity, volume and variety. Data consumption of user increases day by day in which large volume of data increases an elasticity problem arises this decreases the performance of the system where blockchain performs 'n' number of transactions at one block in which it satisfies some v's of big data. Velocity is the speed of the system which reacts to that data in which it gets measured by time. Big data faces an issue in the speed of data and it tries to solve some cases where the data is initialized by the user, the data get automatically converted into blocks very fast and it gets added as the blockchain. Big data face an issue in the speed of data and it tries to solve some cases where the data is initialized by the user and it get automatically converted into blocks very fast and added as the blockchain. Veracity is one of the advantages of big data in which it helps to prove the trustability of data. Big data stores a huge amount of data by the multiple sources, where trustability of sources is the challenging task and it is very difficult to identify accurate data from the users or from the sources. Blockchain solves the veracity problem by verifying each and every block with the help of hash value in which the malicious data and users are removed while mining the blocks. Volume is the enormous amount of data collected to form the big data such as company data and it may consume high storage. Blockchain stores tremendous amounts of data in one single block such as 'n' of transactions or huge volume of data. Big data faces elasticity problems while storing the data. Big data consists of a large variety of data in which it can be structured, unstructured and semi-structured. Big data tries to solve these problems by preprocessing the data such as cleansing, validating and by formatting the data where the blockchain has the capacity to store the 'n' variety of data in which it can also be unstructured and semi structured. Big data undergoes data mining technique to process with the data and it may have intrinsic value and the data may not have the accurate value. Data cleansing techniques were used to solve the inaccurate data. Blockchain stores undergo the mining process to identify whether the data is malicious or not.

B. Blockchain in Cloud Computing

Cloud is a pool of virtualized shared resources in which it can be accessed anywhere at any time and it provides everything as a service such as security, storage, etc. Integration of blockchain with cloud solves the storage issue in blockchain and it provides high security to data.

C. Blockchain in Internet of Things

Internet of Things is a device which senses the information of devices with one another connected with the internet through wired or wireless. Integration of iot infrastructure with blockchain provides more security to the sensory data and it helps in providing the security and privacy to the data.

D. Blockchain in Mobile Computing

Mobile computing is one of the most popular technologies in which humans interact with another human through the device and its application such as images, videos and documents are stored in the cloud in which the records stored in the mobile cloud can be changed by the unauthorized user. Implementing mobile cloud using blockchain improves the security for the mobile data.

VII. SWOT ANALYSIS

SWOT is an analysis is done for the blockchain in which it helps to find the strength, weakness, opportunities and threats where it is briefly analyzed below.

A. Blockchain Strength

Immutability: Immutability is one of the special features of blockchain in which it helps the data from hacking. Once the data gets initialized by the user, it cannot be changed or modified.

Transparency: Blockchain undergoes 100% transparency for the user in the network where user data is hidden from the unauthorized user. So, data tampering is not possible in the blockchain network.

Privacy: Blockchain provides more security to user data in which privacy is maintained. Data privacy is unharmed because the user data cannot be shared with authorized and unauthorized users in the network except the current use.

Avoids Third-Party: Blockchain avoids the centralized servers such as cloud, bank servers and other servers. The centralized server can be easily hacked by the hacker in which it undergoes an immutable decentralized system and it avoids all 3rd party intermediates.

B. Blockchain Weakness

Storage: Blockchain consists of a huge amount of transaction data in which one block consumes huge storage space where ‘n’ number of blocks are interconnected to form the blockchain and this leads to storage issues.

Scalability: Scalability is another issue in which blocks in blockchain so on increases, in which systems cannot survive to increase the system speed. System performance level will be decreased.

C. Blockchain Opportunities

KYC Database: In banks, KYC is maintained for every customer where the data are maintained in the bank server where KYC is unique identification of customers. Centralized databases of banks can be hackable where data theft may occur. Blockchain technology provides a unique hash value for every user where the different hash value is provided for the user.

Speedup the Bank Process: Blockchain technology helps to speed up the bank processes (i.e. fast payment transfer and increase in quality of service and product) and provides more security in storing customer detail.

D. Blockchain Threats

High Investment: Compared to normal payment in the bank where the BPay undergoes the high investment, it cannot be done with the user system and it requires GPU to mining the block and also it consumes more time.

Disappearance of Existing System: By implementing the blockchain technology in this world leads to disappearance of the bank system in which all payment can be done with BPay without using online or offline payment and the customer data and details gets uploaded in blockchain than in bank centralized servers. All data of the user are distributed throughout the network as the hash value.

VIII. APPLICATION OF BLOCKCHAIN

A. Voting System

Many issues arise in voting a person, consider voter elections, in which many malpractices are done. This situation also arises even after introducing electronic voting machines due to invalidation of authenticated users. Blockchain helps to resolve this problem by maintaining the separate hash value for each fingerprint based on that votes are recorded and it cannot be hacked until it attains the 51% error.

B. Supply chain Management

Supply Chain applications with blockchain help many industries and organizations to find the defect and to resolve it. Consider one industry that builds the product which undergoes ‘n’ number of processes to make as a product. By this process many defects and some activities in between the process may lead to damage in the product and these defects are detected by using blockchain technology. Each and every details of the product is verified and assured for the next process.

C. Insurance

Blockchain helps insurance companies to verify the data given by the consumer is true or false in which it undergoes record keeping technology where records and transactions are stored in blockchain because it maintains immutable ledger. It is very difficult to apply for false or fraud records because once the data is converted into blocks then it is difficult to remove the already uploaded records in the blockchain network.

IX. SURVEY OF BLOCKCHAIN

In the digital era, the quantity and diversity of digital data generated by users, sensors, IoT devices, etc. Many benchmark providers have failed to store, organize and secure Big Data. The Blockchain technology with cloud provides significant input and the following table 1 explores the techniques, challenges and outcome of blockchain with cloud, big data and IoT.

Table 1. Explores the literature survey in blockchain.

Authors	Technique/ Scheme	Challenges	Outcome
Yinghui et al [26]	BPay technique helps to find a secure and fair payment without relying on 3rd party. Provable Data Possession helps to identify whether the server and client is malicious or not.	Blockchain based payment doesn't rely on third party.	Robust fairness of payment and Soundness. Data security is maintained.
Bao-Kun et al [3]	Paillier Cryptosystem technique was used to maintain confidentiality of data in blockchain network.	Protect multiparty data.	Improves efficiency. Data confidentiality is maintained.
Jin Ho Park et al [10]	Secure bitcoin protocol was used to adopt blockchain security in cloud. Elliptic Curve Digital Signature algorithm was used to verify the transaction.	Large amount of data can be transferred.	Removes user information in cloud. Double Spending is avoided. Security is maintained while data gets shared.
Ruiguo Yu et al [18]	Text Encryption Protocol is used to prevent from malicious user while mining seed. RSA algorithm was also used prevent the data from the unauthorized user.	Authenticate the user data.	Data privacy is maintained in social network.
Shangping et al [19]	Attribute Based Encryption was used to solve the privacy issues. Interplanetary File System used to chunk large files and stored in different node.	Failure in system when some data shared.	Avoid single set of failure. Large Data throughput. Maintains Data privacy.
Joanna et al [11]	Asymmetric key cryptography was used as security for every block.	Managing the data in cloud.	System resilience Fault tolerance.
Turesson et al [23]	Proof of Useful Work (PoUW) was used to preserve privacy in machine learning technique.	Harms the full computational potential.	Maintains data privacy.
Tanzir et al [22]	Integration technique helps to provide security to IoT device through blockchain.	Data loss while data processing in IoT device.	Provides security to IoT data.
Niranjanamurthy et al [14]	Asymmetric cryptographic technique helps to avoid untrusted environment in blockchain. ECDS algorithm was used to authenticate transaction.	Synchronization problem occurs in distributed database.	Accurate block in blockchain. Overcomes load sharing.
Valentina et al [24]	Hyper cycle was implemented to triggered phase. Smart contract reads all information and analyzes the data.	Decision making is impossible in insurance system.	Identify the frauds during client processing.
Mauro et al [13]	Compressed Sensing technology was used to covert analog signal into information.	Blocks consumes huge amount of data.	Provides security while sensing IoT data.
Chunchi et al [4]	Legal supervision scheme was used to search the encrypted data in blockchain. Public Encryption Key Search was decentralized public keyword search.	Steals the private key where the data was stolen	Efficient transaction handling. Prevents data from 3rd party.
Franceso et al [7]	Computation intensive blockchain algorithm helps to improve energy stringent in IoT.	It enforces contract between two parties.	Avoids data tampering.
Wenliet al [25]	Consensus algorithm was implemented to store all transaction in block.DAG was implemented to maintain the information of transaction.	Allow to access big data.	Identify the trustworthy of internet service. Reduce the traffic delay.
Rongyue et al[17]	Database organization methods supports open sharing of future data.	Sharing of big data.	Modified record cannot be tracked by unauthorized user.
Zehui et al [27]	Mobile blockchain was implemented in edge computing to maximize profit	Mobile device cannot participate in mining.	An economic approach for resource management.

X. BLOCKCHAIN BASED SECURITY

Blockchain is a public distributed ledger in which it maintains the immutable feature where each block is connected to form a blockchain network by consensus algorithm. Miners in the network mine the data where blockchain helps in providing security to the cloud data in which it provides storage space for the data and blockchain technology also helps in sharing and authenticating the data by maintaining privacy. In this paper, we have analyzed data security, storage and sharing and authentication and we compared the analysis with small data and large data.

A. Data Security Analysis with Research Challenges

Blockchain provides more security to the data and it also helps in secure payment without trusting third party intermediates in which it avoids double spending problems and net frauds through online. Blockchain Payment is implemented for the secure transaction in which it maintains a public ledger [7, 10] to store the detail of the consumer where a data possession scheme [26] was introduced to find the server is malicious or not in which illegible transactions are impossible. After the transaction gets completed, blockchain removes [10] the user information from the cloud. In the blockchain network each block provides the security by hashing technique [22] in which every block possesses both hash value and previous hash value in which the data cannot be easily hacked by the malicious user. Digital Signature Algorithm [10] helps to identify whether the data is from authorized users or not. Elliptic Curve Digital Signature Algorithm [10, 26] used to authenticate the transaction and to protect the public key of the user. Endorsement strategies [9] are followed to guide the peers and it helps to identify whether the transaction gets approved or not.

Asymmetric key [11] encryption techniques were used to provide security for every block in the blockchain. Blockchain helps in IoT [1, 2, 4, 13, 16, 22] by providing security to the information gathered from the device where attribute-based encryption techniques [19] were used to encrypt the user data and also provide the security to [18] social network data. Blockchain undergoes consensus algorithms [28] to solve database synchronization and the chaining process improves privacy by compressing [13] chain length.

Research Challenges:

- *Data Representation:* User data is converted into a block and it gets represented as a hash value in which each block in the network has a unique hash value and it provides the security to the data and it cannot be hacked by a hacker.
- *Robust Fairness:* Blockchain based payment maintains the security to the transaction data in which it provides fast and efficient service against eavesdropping and malleability attack and it reduces the computation time to complete the transaction.
- *Soundness*: Blockchain hides the user data to the unauthorized user but it is opened to all who are connected to the system in which the feature helps to provide security to the user data by finding the malicious user and it avoids replying on remote servers.
- *Information Security:* Blockchain provides security to the information in which it keeps only the last modified record in which it is hidden to all in the network where the source of the modified record cannot be tracked.
- *Protects multi party data:* Blockchain helps to protect the multiple 3rd party data by providing security in which it protects the centralized server such as the bank, government sector, etc where it avoids the data tampering.

B. Data Storage Analysis with Research Challenges

Blockchain network maintains the decentralized [19, 28] peer-peer system in which it is used to store large amounts of data in distributed ledger where ‘n’ number of transactions are verified and possesses the consensus algorithm [25] to store the data as block in blockchain. IoT devices detects the data in which it continuously stores the information in the system where it consumes the large amount of data where attribute based encryption [19] were used to access the user data securely in which Interplanetary file system is used to store the large amount of data. Big data comprises an enormous amount of data in which it leads to the synchronization problem where blockchain avoids the problem by mining processes. Paillier cryptosystem [3] in blockchain helps in providing the confidentiality to the data storage. User data are stored in the cloud in which it possesses the centralized form [10, 19, 20] where the data is outsourced from the cloud where the data tampering happens frequently. Blockchain provides more security to the user storage in which the data gets stored in the distributed ledger where the blocks are distributed to all the peers [15, 25] in the network but the data is not distributed.

Research Challenges:

- *Data Representation:* Big data is consumed by the users frequently where the data may be text, audio, video files and it possesses large storage [11] to store the data. Blockchain helps to store the huge data into block to provide security to data.
- *Data Replication:* Data Replication problem mostly occurs in the centralized server in which the single data file may be stored for the ‘n’ number of times. Blockchain avoids the problem by consensus algorithm where Synchronization Byzantine Fault Tolerance (SBFT) [28] which was used to maintain the data consistency in cloud.
- *Information Retrieval:* Data stored as blocks in the blockchain can be retrieved [14] by the user easily using the unique hash value of the block and by traceability feature where the huge amount of data can be retrieved in the blockchain by reducing the time complexity.

C. Data Authentication Analysis with Research Challenges

Blockchain helps in authenticating the data [6,16] in which the process helps to protect the data from the harmed user. Authentication problems mostly occur on the social media [18] in which one’s account can be hacked without getting permission from the authorized person. Blockchain improves the authentication to avoid data tampering and to prevent the privacy [3,23] of the data in which it helps to identify whether he/she is an authenticated user or not. Authentication process takes place in blockchain where it is the first process for transaction [26] in which after the user or server gets authenticated then the communication takes place between two ends.

Punishment reward mechanism [12] was introduced to broadcast messages from vehicles using the internet of things where the message can be deleted or edited by the malicious user where the problems get solved by authenticating the message from the vehicle.

Encryption technique helps in authenticate the data where text-based encryption [19] is used to authenticate a social network using blockchain technique and attribute-based encryption [18] technique helps in providing privacy to the block. Elliptical Curve Digital Signature Algorithm [14] helps in authenticating the transaction in blockchain based payment system.

Research Challenges:

- *Data Representation:* Each block in the blockchain is represented as the hash value where the encrypted data gets hashed in which the hash value cannot be hacked by the malicious user because it consumes the large alphanumeric and it varies for every individual block and it is difficult to solve.
- *Data Encryption:* Blockchain data is encrypted using an algorithm in which the data is hidden and it sends the data to the network where authenticate users such miners are connected in the network to mine the block.
- *Maintains privacy:* Blockchain maintains the privacy of the user data by authentication process in which it helps to find whether the user is malicious or not. After the user is authenticated then the data gets accessible to the user and it helps to reduce the privacy issues that mostly occur in social networks.

D. Data Sharing Analysis with Research Challenges

Sharing the user data may lead to the privacy issue in which the blockchain resolves the problem and protects the data from the malicious user where paillier cryptosystem [3] is introduced to maintain confidentiality while sharing the data. Data Privacy protocol is maintained in the blockchain to check whether the correct user encrypts and decrypts the data. Data is mostly shared using the Internet of Things in which the information from the devices are shared with one another and it may lead to privacy problems. Data sharing techniques [12] also get followed on the Internet of Vehicles in which one vehicle message gets shared to the other vehicle without leaking the privacy using blockchain. Clinical and medical data [16] follows privacy between patient and the specialist to the data history of the patient using blockchain technique. Business Information Modeling (BIM) [17] audits for historical modification with big data.

Research Challenges:

- *Data Representation*: Blockchain performs sharing mechanism in which the block gets shared through all the peers connected to the network where it maintains the privacy.
- *Load Sharing:* Blockchain possesses the load sharing mechanism in which each block undergoes the big data where these blocks consume the loads and it gets represented as a hash value.

XI. RESEARCH CHALLENGES IDENTIFIED FROM SURVEY

Block (size) requires a massive amount of data in which blockchain stores the big data as the block. Security providence of data specifies that the blockchain provides security to the data by consensus algorithm. Avoids centralized server and multi-party protection determines that the blockchain accepts all third-party data where it undergoes the decentralized system and it avoids malicious users. Record tracking and robust soundness indicates the blockchain helps in the payment system in which it maintains the transparency and faster transaction and stores the details. Reliability in the blockchain provides the accurate data for the user where Information retrieval describes the blockchain data can be retrieved by the user whenever it is necessary. Load sharing indicates that the block loads get shared by implementing new mechanisms. Automatic detection of malicious user determines that blockchain possesses the decentralized system and it avoid malicious user in the network. Blockchain plays an important role in providing security to the big data and cloud storage and helps to hide the data from the malicious user while sharing and also in authentication process. The identified research challenges in blockchain are listed below.

A. Block Size

Big data consumes the huge amount of data in which blockchain helps to store the data as the block where each block is linked with the previous block and next block. Blockchain possesses the consensus algorithm in chaining the block in which it specifies how many numbers of the blocks get added as one block. First block in the blockchain specifies the genesis block where it doesn't provide previous hash value.

B. Security providence of the data

Blocks in the blockchain have a unique different hash value. Consider there are three blocks in the blockchain in which every block has a unique hash value where blocks have different hash values for different blocks. Hash value cannot be hacked by the malicious user where the security levels of blocks are improved in which the large data gets converted into the block and it is specified with the hash value.

C. Information Retrieval

Data can be retrieved by the user using a hash value specified for the block. Consider there are 3 blocks in which the user needs to retrieve the 2nd block. Each block has the unique hash value where the 2nd block hash value helps to retrieve the content that gets stored in that block 2.

D. Maintains Privacy

User privacy is maintained in the blockchain where the data cannot be stolen by the unauthorized person. Blockchain undergoes the immutability feature, in which once the data is converted into blocks, it cannot be modified. User data in the blockchain network is represented as the hash value.

E. Data Hash Representation

Blockchain consists of 'n' number of blocks and it gets specified as the hash values. It helps to avoid the malicious user, Consider there are 2 blocks in which the unique block has both hash and previous hash value where 1st block hash value is taken as the previous hash value of the 2nd block. This hash representation helps in chaining the block in blockchain.

F. Automatic Detection of Malicious user

Every block in blockchain has its own hash value and it helps to identify the malicious user automatically. Consider there are 3 blocks in which malicious user change the hash value of the 2nd block in which the blockchain automatically checks the previous hash value of the block with the hash value of the previous block. After satisfying the condition, blocks get added in the blockchain.

G. Avoids Centralized Server

Blockchain undergoes decentralized peer-to-peer system. In current situation, the centralized server is used to store the user data where the data can be easily hacked by the hacker and it leads to the loss of data in which it can be overcome by blockchain technology.

H. Robust Fairness

Blockchain network helps in the payment process and its performance is faster when compare to online transaction and avoids eavesdropping. BPay is safe and it reduces the computation cost where the transaction held in BPay is small and it maintains privacy and secures transactions without any loss.

I. Multiparty Protection

Blockchain supports multiparty data in which blocks accept all kinds of data where it can be text, audio, video, image etc. Blocks accept the multiple 3rd party and centralized server (i.e. government sector) into decentralized and it helps to protect the data.

J. Reliability

Blockchain technology helps in authentication of data and it is used to prevent the user privacy information in social networks because many of the privacy information is leaked due the improper authentication in which it often occurs in the social media. Blockchain resolves this problem by maintaining the privacy and data accuracy.

K. Fault Tolerance

Blockchain possesses a distributed system in which it undergoes fault tolerance characteristics in which one system fails then the other system has the capacity to overcome by providing data where in the blockchain network if one system fails to mine the data, the other system has the capacity to mine the data into block.

L. Record Tracking

Blockchain has the capacity of tracking the record continuously in which the user data in the network is specified by hash value. Consider there are three blocks in which the user converts the third data into block where the block in the blockchain can be tracked using hash value and it is unique and different from other hash values of the block.

M. Load Sharing

Blockchain undergoes a load sharing feature in which one single block can consume a large amount of data and this may overload the system where many new mechanisms are introduced to share the loads and increase the performance.

N. Soundness

Blockchain undergoes the special feature called soundness in which it helps to hide the user information in the network and this provides security to the data. Soundness features are mostly useful in blockchain based payment systems to hide the transaction details in the network from unauthorized users.

XII. CONCLUSION

This paper explored the blockchain technology security providence for big data in the cloud. While analyzing big data; security, storage, sharing and authentication are the challenging tasks in which we surveyed on how blockchain overcomes. The blockchain leads the major role in different domains such big data, cloud, internet of things and mobile cloud. SWOT analysis of blockchain is performed. The blockchain network in perspective of data security, data storage, data sharing and data authentication were analyzed and addressed the research challenges to avoid data tampering. The challenge of the blockchain to overcome the big data and cloud issues and all the challenges are mapped with blockchain technology explored. Our survey addresses how the blockchain provides security to the big data in the cloud for sharing and authentication of data which resolves the storage problem in cloud.

REFERENCES

1. Ali Dorri, Salil S. Kanhere, Raja jurdak and Praveen Gauravaram (2019), 'LSB: A Lightweight Scalable Blockchain for IoT security and anonymity', Journal of Parallel and Distributed Computing, Vol. 134, pp. 180-197.
2. Ana Reyna, Cristian Martín, Jaime Chen, Enrique Soler. and Manuel Díaz (2018), 'On blockchain and its integration with IoT Challenges and opportunities', Future Generation Computer Systems, 88, pp.173-190.
3. Bao-Kum Zheng, Lie-Huang Zhu, Meng Shen, Feng Gao, Chuan Zhang, Yan-Dong and Jing Yang (2018), 'Scalable and privacy-preserving data sharing based on blockchain', Journal of Computer Science and Technology, No. 33(3), pp.557-567.
4. Chunci Liu, Yinhao Xiao, Visheshjavangula, Qin Hu and Shengling Wang (2018), 'NormaChain: A Blockchain-based Normalized Autonomous Transaction Settlement System for IoT-based E-commerce', IEEE Internet of Things Journal, Vol. 6, NO. 3, JUN 2019.
5. CuneytGurcanAkcora, Matthew F. Dixon, Yulia R. Gel and Murat Kantarcioglu (2019), 'Blockchain Data Analytics', Journal OF IEEE Intelligent informatics, Vol. 20, No.1, JANUARY 2019
6. Dipankar Dasgupta, John M.Shrein and Kishor Datta Gupta, (2019), 'A survey of blockchain from security perspective', Journal of Banking and Financial Technology, No.3(1), pp.1-17.
7. Francesco Restuccia, Salvatore D'Ore, Salil S.Kanhere , Tommaso Melodia and Sajal K. Das (2019), 'Blockchain for the Internet of Things: Present and Future', arXiv preprint, arXiv:1903.07448.
8. Huikang Cao, Li Ruixuan, Wenlong Tian, Xu Zhiyong, and XiaoWeijun (2020), 'Blockchain-based accountability for multi-party oblivious RAM', Journal of Parallel and Distributed Computing , Vol.137, pp. 224-237.
9. Jian Chen, ZhihanLv and Houbing Song (2019), 'Design of personnel big data management system based on blockchain',Future Generation Computer Systems, Vol. 101, pp. 1122-1129.
10. Jin Ho Park, and Jong Hyuk Park (2017), 'Blockchain security in cloud computing: Use cases, challenges, and solutions', *Symmetry*, No. *9*(8), pp.164-177.
11. Joanna Kolodziej, Andrzej Wilczynski, Damián Fernández-Cerero and Alejandro Fernandez-Montes (2018), 'Blockchain secure cloud: a new generation integrated cloud and blockchain platforms– general concepts and challenges', European Cybersecurity Journal, Volume 4, issue-2.
12. Lei Zhang, Mingxing Luo, Jiangtao Li, Man Ho Au, Kim-Kwang Raymond Choo, Tong Chen and Shengwei Tian (2019), 'Blockchain based secure data sharing system for Internet of vehicles', Vehicle Communications, Vol.16, pp. 85-93.

13. Mangia M., Marchioni A., Pareschi F., Rovatti R. and Setti G., (2019), 'Chained Compressed Sensing: A Block-Chain-inspired Approach for Low-cost Security in IoT Sensing', IEEE Internet of Things Journal.
14. Niranjanamurthy M., Nithya B.N. and Jagannatha S., (2018), 'Analysis of Blockchain technology: pros, cons and SWOT', Cluster Computing, pp.1-15.
15. Paula Fraga-Lamas and Tiago M. Fernández-Caramés, (2019), 'A review on blockchain technologies for an advanced and cyber-resilient automotive industry', IEEE Access, No.7, pp.17578-17598.
16. Peng Zhang, Jules White, Douglas C. Schmidt, Gunther Lenz and S.Trent Rosenbloom (2018), 'FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data', Computational and Structural Biotechnology, Vol.16, pp. 267-278.
17. Rongyue Zheng, Jianlin Jiang, Xiaohan Hao, Wei Ren, Feng Xiong and Yi Ren Y (2019), 'bcBIM: A Blockchain-Based Big Data Model for BIM Modification Audit and Provenance in Mobile Cloud', Mathematical Problems in Engineering.
18. Ruiguo Yu, Jianrong Wang, Tianyi Xu, Jie Gao, Yongli An, Gong Zhang and Mei Yu (2017), 'Authentication with blockchain algorithm and text encryption protocol in calculation of social network', IEEE Access, No. 5, pp.24944-24951.
19. Shangping Wang, Yinglong Zhang, and Yaling Zhang (2018), 'A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems', IEEE Access, 6, pp.38437-38450.
20. ShaoanXie, Zibin Zheng, Weili Chen, Jiajing Wu, Hong-Ning Dai and Muhammad Imran (2020), 'Blockchain for cloud exchange: A survey', Computers and Electronic Engineering, Vol. 81, pp. 106526.
21. Tanzir Mehedi S.K., Abdullah Al Mamun Shamim, and Mohammad Badrul Alarm Miah (2019), 'Blockchain-based security management of IoT infrastructure with Ethereum transactions', Iran Journal of Computer Science, pp.1-7.
22. Turesson H., Roatis A., Laskowski M. and Kim H., (2019). 'Privacy-Preserving Blockchain Mining: Sybil-resistance by Proof-of-Useful-Work', arXiv preprint, arXiv:1907.08744.
23. Valentina Gatteschi, Fabrizio Lamberti, Claudio Demartini , Chiara Pranteda and Victor Santamaría (2018). 'Blockchain and smart contracts for insurance: Is the technology mature enough?', Future Internet, No.10(2), pp.20.
24. Wenli Yang, ErfanAghasian, Saurabh Garg, David Herbert, Leandro Disiuta and Byeong Kang (2019), 'A Survey on Blockchain - based Internet Service Architecture: requirements, challenges, trends and future', IEEE Access.
25. Yinghui Zhang, Robert H. Deng, Ximerg Liu, and Dong Zheng (2018), 'Blockchain based efficient and robust fair payment for outsourcing services in cloud computing', Information Sciences, 462, pp.262-277.
26. ZehuiXiong, Yang Zhang, Dusit Niyato, Ping Wang and Han Z. (2018). 'When mobile blockchain meets edge computing', IEEE Communications Magazine, N0. 56 (8), pp.33-39.
27. Zhiqin Zhu, Guanqiu Qi, Mingyao Zheng, Jian Sun and Yi Chan (2019),' Blockchain based Consensus checking in decentralized cloud storage', Simulation Modeling Practice and Theory, p.101987.
28. https://bitcoin.org/bitcoin.pdf
29. https://www.livemint.com/industry/banking/rbi-says-bank-fraud-touche s-uprecedented-rs-71-500-crore-in-2018-19-1559552056883.html

AUTHORS PROFILE

Manikandan D, Completed his M.E in MIT Campus-Anna University. He is doing his intern in Honeywell and his area of interest is Big data, Blockchain and IoT. Email: pugalanthimanikandan40@gmail.com

Valliyammai C, received her Ph.D in Computer Science and Engineering at Anna University. She has eighteen years of teaching experience. Her area of interest includes Big Data, Cloud Computing, and Databases. She has around 85 publications in National and International conferences and journals. E-mail: cva@annauniv.edu

Karthika RN, M.Tech., a Research Scholar in the Department of Computer Technology, MIT Campus, Anna University, Chennai, India. She received her B.Tech. and M.Tech. degree in Information Technology discipline. Her areas of interests include Blockchain, Cloud Computing, Big Data Storage.
E-mail: rannar@gmail.com