iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: III Month of publication: March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3147

Blockchain based Secure Data Sharing and Sensitive Information Hiding in Cloud Storage Environment

Nihila Fathima M[1], Remitha M. S[2], Rithanya R. K[3]
[1, 2, 3]Department of Computer Science and Engineering, University College of Engineering, Nagercoil

Abstract: Cloud computing is the dynamic delivery of Information Technology services over the network. With increased popularity of cloud services, data owners are motivated to store their huge amount of personal data on remote servers. Nevertheless, it brings many security issues since the data stored in cloud is vulnerable to attack. To ensure confidentiality, data owners encrypt their data before outsourcing to cloud. However, to achieve data sharing with sensitive information hiding with other users of cloud, encrypting the whole shared file achieves sensitive data hiding but the shared file will not be accessible to others. In cloud based storage systems like Electronic Health Record system, organization struggle to keep patient data secure, up-to-date, auditable and accessible to all parties together with hiding sensitive data. Existing system of sensitive information hiding made use of sanitizer for sanitizing the data file before uploading to cloud and third party auditor for auditing the integrity of the sensitive data stored a in cloud. This scheme requires to control important database that hire many people and design many process to prevent database from tampering. Unavoidably, all this requires great amount of money and time. Therefore in this paper, Blockchain technology is used in which third party organization can be replaced with distributed database where sensitive data is maintained securely in blocks, lock down by clever cryptography. Hence users can trust that the transactions can be executed exactly as the protocol commands removing the need for a third party thus ensuring integrity of the sensitive data.

Keywords: Data Sharing, Sensitive Information Hiding, Data Integrity, Blockchain, Cryptography

I. INTRODUCTON

With increased development of cloud computing, cloud storage services are used by various organizations and individuals since the data is remotely stored, maintained, managed and backed up. Hence data owners have no burden of overhead resulting from cost, storage and maintenance. The service allows users to store files online and access them back from any location via Internet. However storing enormous amount of data with Cloud Storage Provider (CSPs) raises concerns about data protection. Data Integrity and privacy can be lost due to the migration of data from one location to another by the cloud admin or by any malicious users. The CSP is an external entity that cannot be fully trustworthy. There can be a chance of data tampering / alteration by any unauthorized entity. Therefore the data owners encrypt their data before outsourcing to cloud and to make sure whether the data is secure, unmodified by others, data owners opt for data storage correctness. Therefore to ensure the integrity of remote data, various data integrity auditing schemes have been proposed, in which the verification of cloud data is performed by a Third Party Auditor (TPA). This encrypted data is accessible back by the data owner .All these schemes consider only the privacy checking of the personal data stored in cloud. But the user data might contain sensitive and non sensitive data. If outsourced data have to achieve sensitive information hiding as well as data sharing of non sensitive data with other users of cloud, encrypting the whole file achieves data hiding but the shared file won't be accessible to others. Therefore in [6], an identity based remote data integrity auditing scheme is proposed in which the data block of sensitive information is sanitized be the sanitizer. The signature generated for the sanitized block of data is used to verify the integrity of sanitized file using TPA on behalf of data owners. Thus the file stored in the cloud is able to be shared and used by others under the condition that the sensitive information can be protected while the other information can be published. However, data privacy against TPA is also needed since it is not guaranteed to be trustworthy as they know the data's real authenticators, and they can try to manipulate the data for their own benefits[4].Also an unusual activity in TPA can cause entire cloud system to go down or reduce the performance. Therefore to address these issues BlockChain Technology is used in which third party organization can be replaced by a distributed database where sensitive data is kept securely in blocks thus ensuring the integrity of the sensitive data while the non sensitive data can be directly uploaded to cloud.

Blockchain is a chain of block ordered in a decentralized network of untrusted peers. The immutability behavior of Blochchain makes the sensitive data stored in blocks free from further alteration. Each block of data will be verified independently by the miners (nodes) on the network via a consensus model which often requires some resource in the form of computing power to create blocks and to show the proof of validation. Further the validated block will be added to the Blockchain environment and connected to the cloud.

A. A Descriptive Example For Data Sharing And Sensitive Information Hiding

Electronic Healthcare Record (EHR) systems are crucially significant to many healthcare organizations. EHR is designed to store health records in a secure location say cloud giving accessibility to practitioners and admin staff that require regular access or the researchers for research purposes. EHR contains two kinds of information both sensitive and non-sensitive. There are two sections of sensitive information. One is the patients' personal details such as Name, Address, Blood group etc while the other is hospitals' sensitive information like hospital name, address etc. The non-sensitive information includes details about the patients' disease, prescription provided for that disease. Basically, this non sensitive information can be encrypted and directly uploaded to cloud enabling only the authorized users to view it and research about it .While the sensitive information such as the patient as well as the organization's personal information can be hashed and fed to the Blockchain environment. Miners (nodes) on the Blockchain network select the sensitive information transaction from a pool of transactions that is eligible to process. In simple terms miner can be viewed as people who run mining operation to add transactions to Blockchain. Every miner constructs their own block of transactions. A block is a permanent store of records which, once written, cannot be altered or removed. To add the transaction to the block and to add that block to blockchain network, the miners have to find the eligible signature for the data and the block based on Proof-of-Work working standards. This process is known as Mining. The miner who first find the eligible signature will broadcast this information to the other miners on the network. The other miners do a confirmation on that block of transaction by validating the generated signature based on the working principle of Consensus model and allow the block to be added to the blockchain environment. Finally, the blockchain network will be connected to cloud. Under the condition that the sensitive information is kept secured in blocks, immutable lock down by a clever cryptography.

B. Related Work

Data integrity, data confidentiality in cloud storage is the key privacy issue that has to be considered. Checking data for correction is called data integrity. To overcome the challenges associated with data integrity, various schemes have been proposed. In [1], G.Atniese et al. proposed a technique in which the client send public key along with the file to the server and delete the file from his local storage. Then check for response from the server for data possession of a section of data of the file by challenging the server with proof-of-possession. In [2], Y.Li et al. proposed Fuzzy-identity based auditing protocol in which Biometric is used as fuzzy identity. The cloud user encrypts the file using the private key generated by measuring the property of fuzzy identity. The TPA is used to check whether the data stored is intact or not or anyone knowing the cloud user's identity is able to check the data integrity on behalf of the cloud user. In [3], Yang et al. proposed an efficient shared public data integrity auditing scheme which not only protects the identity privacy but also identity traceability of the group members. Most Existing protocols can support data integrity features with the help of a third-party auditor. In [4], W. Shen et al. proposed a Remote data possession checking scheme with privacy-preserving authenticators named Homomorphic Invisible Authenticators (HIA) invisible to both cloud and TPA since there is a chance for modification of data by the third party who knows the data's real owners. This guarantees that no private information is leaked to third party verifiers. In [5], J. Shen et al. proposed a public auditing protocol with blockless verification to guarantee data owner that the cloud has securely stored data.

To share data across multiple users, Zhang et al. W.Shen et al. [6] proposed a remote data integrity auditing scheme is used to achieve data sharing with sensitive information hiding in identity-based integrity auditing for secure cloud storage. Integrity of the shared data is audited using a Third party Auditor.

In general these schemes require a third party to verify the integrity of the shared data which is untrustworthy with Possibilities of security attacks due to the presence of third party. D. Randall et al. [7], suggests that the decentralized and programmable nature of the blockchain applications can be used to change health information technology. A.Alketbi et al. [8] reviews the potential use cases and application of blockchain to enable government services. Du Mingxiao et al. [9] review the principles of consensus algorithms and analyzed the performance of different consensus mechanisms. Nima Zahadat et al. [10] explore the origin of Blockchain, how centralized data management can be replaced with decentralized nature of blockchain ensuring confidentiality, integrity. A.Gervais et al. [11] introduced a novel quantitative framework to analyze the security and performance suggestion of various consensus and network parameters of PoW blockchains. To implement privacy- preserving resource distribution and sharing in a decentralized network [12] proposed Transaction-based Access Control (TBAC) platform which integrates the standard attribute-based access control (ABAC) model and the blockchain system. C. Decker et al. [13] analyze how Bitcoin uses a multi-hop broadcast to propagate information and blocks through the network to update the ledger of blockchain history. In this paper, we explore how to achieve data sharing with sensitive information hiding in Blockchain environment for secure cloud storage.

II. PRELIMINARIES

This section reviews some preliminary knowledge used in this paper, including System model and Design goals.

A. System Model

As illustrated in Fig. 1, we consider a blockchain based data sharing with sensitive information hiding consisting of five entities: the cloud, the data owner, the admin, Miners, and the Researcher.

1) The cloud provides enormous amount of space to the data owner to store their data in the cloud.
2) The data owner owns large amount of data that will be outsourced to the cloud.
3) The Miners are in the blockchain network responsible for performing the mining process to create the block and to place that block in the blockchain environment.
4) The admin monitors the details about the data owner and the researcher.
5) The Researcher fetches the data from the cloud and uses that data for their research purpose.

As for the relationships among the five entities, a brief explication will be provided here. The data owner outsources his data files consisting of both sensitive and non sensitive data. The non sensitive data is encrypted and directly outsourced to the cloud. The researcher can fetch that data from the cloud and use it for research purpose. The sensitive data is hashed and fed to the blockchain network. The miners on the blockchain network pickup these transacted data and perform the mining process. For that they perform the Proof- of -Work and consensus principle to create a block for the hashed data and to place that block in the blockchain network. Finally, this blockchain environment is connected to the cloud. The admin monitors the details of the data owner and researcher.

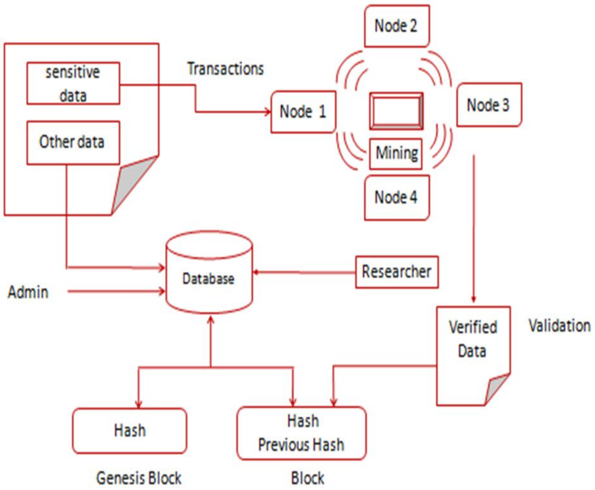

Fig 1: System Model

B. Design Goals

1) To achieve data sharing with sensitive information hiding in which the sensitive information of a document will be protected securely in blocks of blockchain with eligible signature after efficient validation while the other information can be published to the cloud enabling authorized user to access it.
2) As the sensitive data of the file is kept in blocks of blockchain, it achieves the immutability property thus ensuring integrity.
3) The non sensitive information of the file is uploaded to cloud under the knowledge of the data owner allowing only authorized user to view and research about it

III. THE PROPOSED SCHEME

A. An Overview

In order to achieve data sharing with sensitive information hiding, encrypting the whole file before outsourcing to the cloud secures the data but the shared file cannot be accessed by all. Therefore, to achieve this, BlockChain Technology is used in which third party organization can be replaced by a distributed database where sensitive data is kept securely in blocks thus ensuring the integrity of the sensitive data while the non sensitive data is encrypted and it can be directly uploaded to the cloud giving access to all authorized users. Each user owns a secret key generated by a Private Key generator. Finally the blockchain network is connected to the cloud under the condition that the sensitive data is kept secured in blocks, immutable thus ensuring integrity

The blockchain is a sequence of blocks, which holds a complete list of transaction or records. Each block contains block header, previous hash value and transactions. The first block is known as genesis block. A block can contain maximal number of transactions depending on the block size and the size of each transaction. The miners on the blockchain environment pickup these hashed data and perform the mining process i.e. Proof-of-Work. In simple terms, miner can be viewed as people who run mining operation to add transactions to the large distributed public ledger of existing transactions called Blockchain. Every miner constructs their own block of transactions. A block is a permanent store of records or transactions which, once written, cannot be altered or removed. Before adding the transactions to their block, the miner checks whether the transaction is eligible to process or not. After checking, to add the transaction to the block and to add that block to blockchain network, the miners have to find the eligible signature for the data and the block that is unique to each block of transactions. For this, miners need to solve a large complex mathematical problem which requires a huge amount of computing power. The block header contains a nonce and miners would change the nonce frequently to get different hash values until they find the respective signatures. The consensus requires that the calculated value should be smaller than or equal to a definite given value. When one node meets the target value, it would broadcast the block to other nodes on the network for mutual confirmation on the correctness of the hash value. If the block is validated, the miners would append this new block to their own blockchain which provides a non-repudiable, permanent, and transparent record of transactions. Every other block that is added on top of it will do a confirmation on this new block.

B. Non sensitive data encryption and decryption

The non sensitive data of the user encrypted for secure cloud storage consists of the following steps: *setup, keygen, encrypt, decrypt.* The algorithms are detailed as follows:

1) Setup (k—>p, msk): It is run by the Private Key Generator (PKG). It takes a security parameter k as input to initiate a cryptographic scheme and a master secret key and system parameters as output.
2) Keygen (p,msk—>sk) : Generates shared secret keys. It takes system parameters p, master secret key msk, as input and shared secret key sk as output.
3) Encrypt (p,sk,m—>c) : Encrypts the data using AES. It takes the message or plaintext m, shared secret key sk, as input and outputs cipher text c.
4) Decrypt (p,sk,c—>m) : Decrypts data. It takes cipher text c and its corresponding shared secret key as input and outputs the resultant plaintext message m.

$\text{Decrypt}_{sk}(\text{ Encrypt}_{sk}(\text{ m })) = \text{m}$

C. Sensitive Data Transaction

The sensitive data of the file containing personal information is hashed using a cryptographic hashing algorithm before bundling them together into blocks. In electronic healthcare records, the data in is classified based on the level of sensitivity of data contained in it. When the data owner set the sensitive data for transaction to blockchain environment, it involves the following process:

1) Sensitive data are bundled together into blocks.
2) Miners verify that the transaction within each block is legitimate.
3) To do so, miners solve a complex mathematical puzzle known as Proof-of-Work problem.
4) Miners who first solve the problem inform other miners for further verification.
5) Verified block of transacted sensitive data will be stored in the blockchain.

Sensitive data transaction say T= {T1, T2,… Tn} transmitted to the Blockchain network claims some output by providing a proof of ownership. As transactions are broadcasted, when a node receives a transaction, it is verified whether it is eligible to process or not. This process includes:

a) *Add Transaction (From Add, To Add, M):* The data owner bundles the transaction of data to blockchain containing the address of the source ,destination and the sensitive data m.
b) *Hash Transaction:* The transactions are taken as input and run through a cryptographic hash algorithm which gives output of a fixed length. Sha256 (fromadd, toadd, sensitive data, timestamp) ——>Hash (m)
c) *Key Generation (D, H):* Takes the input parameter k and generate matching private-public key(d,H) based on elliptic curve cryptography.
d) *Sign Transaction (D, Hash (M)):* The data owner signs the transaction by taking private key d, hashed data m and returns signature(r, s).The signature ensures that the source of transaction is legitimate. This signature is stored inside the transaction object and later stored on the blockchain. Sign (d, hash (m)) ——> Signature
e) *Verify Transaction(R, S, H, M):* Every node receiving the signed transaction verifies if signature(r,s) on message m and public key H is valid. A malformed transaction will not go past one node. Verify (H, Signature) ——> True/False

D. *Mining a Block*

Each block of transaction has its own mathematically complex puzzle to solve and becomes part of the puzzle for next block of transaction, creating a chain Every miner compete to solve the problem to find the respective hashed output for the hashed input and signature of the block based on cryptographic hash algorithm. The solution so obtained is called Proof-of-Work. To add a block to blockchain, the SHA256 hash computed for a block header must be less than or equal to the target hash. The hash is a random number between 0 and 2**256-1.To find this is a computationally intensive process in which the difficulty level to solve the Proof-of-Work is adjusted based on the time it took to find the previous blocks so that an average a block is solved in 10 minutes. The average time to find a block can be estimated by calculating:
Time = difficulty * 2 **32 /hash rate The Mining algorithm used by the miners to perform Proof-of-Work is Dagger Hashimoto [14-15], an ASIC (Application Specific Integrated Circuit) resistant memory-hard algorithm to equally tally the computing power of all machines involved in the mining operation.

1) *Block Creation:* In blockchain, the transactions are ordered in the form of blocks in a linear chain, which are linked to each other. Hence, Miners solve have to solve the complex mathematical puzzle, form a new block and confirm the transaction. Every miner constructs their own block of transaction. First block is called the genesis block.

Block contains
a) Index (first block: 0)
b) Previous hash(determines previous block)
c) Timestamp(current time)
d) Data(data that the block finder needs to include in the blockchain)
e) Hash(hash taken from content of the block)

2) *Signature Generation:* For a block to be a part of the blockchain, it needs to have a valid hash that is unique for each block. In blockchain, this is created by a cryptographic hash function SHA256 that always gives the same output for the same input, but always a different output for different input. This gives blocks their signatures. Signature or Proof of Work (PoW) symbolizes that a miner has spent a plenty of time and resources to solve very complicated mathematical problems. Each block of transactions gets non-identical mathematical problems to solve, so every miner will work on different problem that is unique to the block they built. SHA256 (index +previous hash +data + timestamp)——>Hash
3) *Signing a Block:* A signature is not always enough. A block will be only accepted on the blockchain if its signature starts with a continuous number of zeroes. The proof-of-work consists of finding a byte string, called nonce, that combined with the block header results in a hash with a given number of leading zero- bits, or target. Finding such a nonce can only be done by actually calculating the hash of the block for all possible nonce until a valid solution is found. It is therefore difficult to find an input that produces a solution, but easy to verify it. The hash is also used as the block's identity. A block now contains;

a) Transaction data,
b) The signature of the previous block,
c) Difficulty and nonce

The process of repeatedly changing the nonce to find an eligible signature for the block is called mining and the nodes attempting to find a solution to the proof-of-work are often called miners.

4) *Adding Block To Blockchain:* The miner, who first finds the qualified output or signature for the next block to be added in the blockchain, will broadcast this block and its generated signature to other miners.

E. Validation and Verification

In this step, the other miners will examine the validity of the signature launched by the miner by hashing the data string of the broadcasted block and comparing the hash out with the signature. If both match, the miners give their confirmation on its validity. The block is now ready to be added in to the blockchain, and is spread across all other nodes on the network. A blockchain is said to be valid if:

1) The blockchain contain only valid blocks.
2) All the transaction contained in the blocks is valid.
3) The blockchain starts with the genesis block.

F. Results and Monitoring

The admin of the organization is responsible for monitoring the sensitive and non sensitive data outsourced to the cloud enabling only the authorized users to have access to it for further research purposes.

IV. SECURITY MODEL

A. Security Objectives

To enable secure data sharing with sensitive information hiding for cloud data, the proposed design should achieve the following objectives:

1) *Integrity:* immutability provides integrity. Thus eliminating the use of third party authority to check the data for any compromise ensures integrity.
2) *Immutability:* The hashing process of a new block always includes meta-data from the previous block's hash output. This makes the chain unbreakable making it impossible to modify or delete data after it has been validated and placed in blockchain
3) *Confidentiality:* The transaction and the identities of participating nodes should be protected. Transaction details must be invisible to the person who is not involved in that particular transaction until the participating entities reveal their information
4) *Access Controllability:* In Access Controllability the knowledge owner will be able to perform wholly the selected restricted actions on the cloud data. Only the legal users are allowed to access the information from the cloud and the illegal users can't be able to get the information from the respective cloud without permissions

The way in which sensitive data propagated to the blockchain network, will be secured is explained with a simple example as follows:

The distributed network of blockchain is monitored by the admin of the organization. Hence any unauthorized or malicious users trying to alter the data or trying to access the block of data will completely change the chain, which is impossible. Let the number of blocks chained to form blockchain be b1, b2, b3 ….bn. When a hacker attacks b3 and tries to change the data, due to the properties of hash functions, a slight change in data will change the hash drastically. This means that any slight change changes made in b3 will change the data and the hash of b2 will result in changes in b1 and so on and so forth. Thus sensitive data transacted achieves immutability thus ensuring integrity.

V. PERFORMANCE EVALUATION

The performance of the proposed scheme is demonstrated in this section.

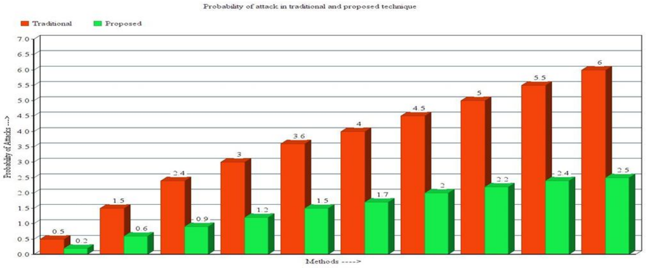

Fig 2: Probability of attacks

Blockchain based sensitive information hiding achieves more security compared to traditional system of data hiding. In the traditional system of data hiding, the sensitive information is stored in a centralized database so if an attacker attacks the database, he can fetch all the information from it. But in blockchain, each block of sensitive data transaction is connected to all the blocks back and forth. Hence it is very difficult to tamper with a single data because a hacker would need to change the block containing that data as well as those connected to it to avoid detection. The data is kept secure in blockchain through cryptography. In fig 1, the probability of various kinds of attacks is calculated using the request response time and throughput. Network participants have their own private keys assigned to transactions they make and it acts as a personal digital signature. If any record is altered, the signature becomes invalid and the peer network will get to know that something has happened.

VI. CONCLUSION

In Blockchain based secure cloud storage of data sharing and sensitive information hiding, the file or the document containing non sensitive information can be outsourced to cloud enabling only authorized user access while the sensitive information of the file is secured using blockchain technology and later, it will be connected to the cloud. The immutable behavior of blockchain eliminates the need for third party authority to check the integrity of the sensitive data.

REFERENCES

[1] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kiss- ner, Z. Peterson, and D. Song, “Provable data possession at untrusted stores,” in Proceedings of the 14th ACM Conference on Computer and Communications Security, ser. CCS ’07, 2007, pp. 598–609.

[2] Y.Li, Y.Yu, G.Min, W.Susilo, J.Ni, K.K.R.Choo,“Fuzzy Identity-based data integrity auditing for reliable cloud storage systems,”IEEE Transactions on Dependable and Secure Computing, vol.14,issue.8,2017

[3] G. Yang, J. Yu, W. Shen, Q. Su, Z. Fu, and R. Hao “Enabling public auditing for shared data in cloud storage supporting identity privacy and traceability”,journal of system and software,vol.113,Mar 2016.

[4] W. Shen, G. Yang, J. Yu, H. Zhang, F. Kong, and R. Hao, “Remote data possession checking with privacy- preserving authenticators for cloud storage,” Future Generation Computer Systems, vol. 76, 2017

[5] J. Shen, J. Shen, X. Chen, X. Huang, and W. Susilo, “An efficient public auditing protocol with novel dynamic structure for cloud data,” IEEE Transactions on Informtion Forensics and Security, vol.12, issue.10, 2017

[6] Wenting Shen, Jia Qin, Rong Hao and Jiankun Hu, “Enabling Identity-Based Integrity Auditing And Data Sharing With Sensitive Information Hiding For Secure Cloud Storage”, IEEE Transactions on Information Forensics and Security, vol.14, issue 2, Feb 2018

[7] David Randall, Pradeep Goel and Ramzi Abujamra, “Blockchain Applications and Use Cases in Health Information Technology”, Journal of Health and Medical Informatics, vol.8, issue.3, July 2017

[8] Ahmed Alketbi, Qassim Nasir, “Blockchain for government services—Use cases,security benefits and challenges”, IEEE Conference on Learning and Technology , 2018

[9] Du Mingxiao, Ma Xiaofeng,Zhang Zhe, Wang Xiangwei, Chen Quijun, “A Review On Consensus Algorithm Of Blockchain”,IEEE International Conference on System,Man andCybernetics,Oct2017.

[10] Nima Zahadat* and Whitney Partridge,” Blockchain: A Critical Component to Ensuring Data Security”, Journal of Forensic Sciend and criminal Investigation,July 2018.

[11] Arthur Gervais,Ghassan O.Karame, Kurl Wüst,”On the security and performance of Proof of Work Blockchain,”,in the Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security,Oct 2016.

[12] Yan Zhu, Yao Qin, Guohua Gan, and Shuai Yang, William Cheng-Chung Chu,”TBAC: Transaction-Based Access Control on Blockchain for Resource Sharing with Cryptographically Decentralized Authorization”, 42nd IEEE International Conference on Computer Software & Applications, 2018.

[13] Christian Decker, * Roger Wattenhofert,” Information Propagation in the Bitcoin Network”, 13-th IEEE International Conference on Peer-to-Peer Computing

[14] T.Dryja, “Hashimoto: I/O bound proof of Work,” 2014

[15] V.Butein,”Dagger: A Memory-Hard to comput, Memory-Easy to verify scrypt Alternative,”2013