2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT) | 979-8-3503-1212-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/IConSCEPT57958.2023.10170327

Decentralized Data Storage Solutions using Hyperledger Fabric

Swati V. Jadhav, Sahil P. Patil, Siddhesh B. Patil, Dirgh D. Patodia, Aryan Pokharkar

Department of Computer Engineering
Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India

Abstract — **As geospatial data grows exponentially with the various advanced photographic capture technologies and satellite imagery being readily available and accessible, the need for secure and scalable storage systems grows larger. Blockchain-based distributed data storage and access systems can help provide an apt solution to this ever-looming issue. This paper talks about how Private Blockchain using IPFS (Interplanetary file System) and Linux Hyperledger technologies can be utilized for the Storage & Retrieval of Geospatial data. This paper explains the flow of the data from the point of generation to the point of retrieval from the Blockchain. The image is stored into IPFS Decentralized servers where we are provided with a CID (Content Identifier) for every unique entity. The CID is stored into the Hyperledger for the additional layer of blockchain-based security. For data retrieval, we simply enter the specific CID into the Hyperledger server through the authorized nodes with Keys to the private blockchain. This study aims to make the technology for the safe storage and reading of such data available.**

Keywords — Blockchain, IPFS, Hyperledger, CID, Geospatial Data, Docker.

I. INTRODUCTION

In these technologically advanced times, the data available grows at an unimaginable rate. This data needs to be handled with caution as in the wrong hands it can lead to some devastating consequences. Geospatial data is one such data which is very important for national security as it contains the Topographical and Geographical information about any given location which can lend an upper hand in case of wars or other disturbances. To avoid such things happening we need to store the data in secure locations. One of the best ways to store the data is in Private Blockchains.

A distributed digital ledger with transaction-containing blocks is known as a blockchain. On a blockchain, a user's transaction must first be signed with a private key before being broadcast to all network nodes for inclusion in a block. The transaction is subsequently placed in a candidate block with other transactions. The candidate block must then undergo validation through mining to be included in the blockchain. The mining procedures vary depending on the blockchain's consensus mechanism, but they all often include calculating a value based on the transactions in the block. Following the mining operation, the node will broadcast the block to every other node in the network so that they can all quickly confirm that the value produced is accurate. The block is added to the blockchain if everything is accurate; otherwise, it is removed. Once a block is added into the ledger, further modifications cannot occur without invalidating that specific copy of the blockchain since the nature of the blockchain is immutable.

To store the data on decentralized servers we use IPFS which returns a hash code CID corresponding to every unique entity in the database. Digital material is stored using IPFS, which offers high levels of integrity and universal accessibility. Ethereum smart contracts are used to control, manage, and offer traceability and insight into the history of digital material from its inception to the most recent version in a way that is decentralized and accessible from anywhere in the world with high integrity, resilience, and transparency.

To create and maintain a private blockchain we decided to proceed with Hyperledger technology. The Linux Foundation initially founded this open-source project known as Hyperledger [9]. The initiative was unveiled in December 2015 with the goal of advancing and promoting Blockchain technology while broadening the scope of potential business applications. The architectural approach adopted by Hyperledger projects places a strong emphasis on interoperability, very secure solutions, a token-based strategy without a local coin, and the development of apps with rich and intuitive user interfaces. Developers may experiment with making numerous sorts of components thanks to the modular architecture. They can provide pieces that can be put together to form distributed general ledger systems that satisfy various needs.

This research has been influenced by multiple other works by renown researchers in their own rights. We would now like to discuss those works in detail.

Cottrell et. al. [1] implemented a permissioned blockchain, Distributed Hash Tables (DHT), and a peer-to-peer network of PingER monitoring Agents to create a blockchain-based data storage system, basically keeping just the metadata of the files on the blockchain while the actual files were saved through DHT at various places (MAs). For those who are not acquainted, PingER is an end-to-end

Internet performance assessment system created and run by the SLAC National Accelerator Laboratory in the United States. Through implementation, they were able to successfully eliminate a single point of failure. For this, we make use of the distributed hash tables (DHT) and permissioned blockchain concepts.

Lazar et. al. [3] took upon themselves the challenge of determining whether the presence of Hyperledger fabric for applications in practice was justified. They concluded that Hyperledger's primary objective was to elevate business transactions to a whole new dimension that would include significant global processes after analyzing the Hyperledger frameworks that were developed as part of an initiative to promote and encourage Blockchain technologies. Hyperledger Fabric, something that we have utilized, has been created to adapt and provide a means to the maximum number of use cases. Modern businesses have a solid foundation present in Hyperledger technologies to utilize, and as time progresses those foundations are certainly to increase much more in their scale, and term of impact.

In a 2022 study on the Interplanetary File System (IPFS) by Trautwein et. al. [2], a set of assessment procedures were implemented that enabled in-depth analysis of the currently available features, upon which the performance and efficacy of IPFS was assessed. The researchers discussed their observations made while working at Protocol Labs, the IPFS project's biggest backer and employer of most supporters. Single points of failure may be present in the conventional centralized system. The failure of the domain provider or the security provider may occasionally result in significant data loss and create uncertainty for the commercial objectives of an organization or a person. [3] The decentralized web, a developing movement with the goal of promoting user emphasis on data security and privacy, is bringing about change by encroaching on a variety of spheres that are a part of users' everyday life. The user-uploaded content is initially imported into the Interplanetary File System (IPFS), after which a CID is assigned to it. The fundamental unit of detaching a name for each unique piece of material is called a Content Identifier, or CID. CIDs profoundly advance the notion of decentralizing content delivery.

Wang et al [7] in their paper discuss the implementation of a blockchain network for efficient picture storage and watermarking without the need for a third party to arbitrate. safe image security utilising zero-watermark algorithm based on blockchain network. The network uses Ethereum blockchain, Smart contract and IPFS technologies. To overcome the blockchain data growth challenge, the framework created integrates blockchain and zero-watermark technologies and employs an interplanetary file system. Furthermore, the image owner can verify the image and ensure the image's copyright traceability. Furthermore, the keyword search feature of the system's stored images is implemented using a smart contract, which eliminates the problem of a lack of trustworthy third parties.

Protocol Labs has created Filecoin [5], a decentralized storage network that transforms cloud storage into an algorithmic marketplace. The platform operates on a blockchain and includes a native protocol token, also known as "Filecoin," which miners can obtain by offering storage to users. Clients in Filecoin rely on miners to store or share their data. Just like Bitcoin miners, Filecoin miners compete to discover blocks with significant rewards. However, unlike Bitcoin, the mining power of Filecoin is linked to active storage, which benefits users. This contrasts with Bitcoin mining, which is primarily focused on maintaining blockchain consensus. This incentive structure provides Filecoin miners with a strong motivation to accumulate as much storage as possible and offer it for rent to clients.

Satashi Nakamoto famously invented Bitcoin [6], a peer-to-peer electronic money that used a blockchain network with cryptographic hashes and digital signatures to ensure the security and privacy of financial transaction data. Blocks in chain format are securely linked together using the hash function. It did away with the idea of a trusted third party for transaction record verification and storage in favour of employing participant network nodes/blocks as the verifying authority, implying that every transaction would only be recognised as legitimate after validation by every other node/participant. Each node maintained a copy of the distributed digital ledger, which was used to record each transaction in the network. Users may mine, store, and exchange bitcoins on the decentralised Bitcoin network via a time-consuming computer operation.

Zheng et al [10] solved the problem of huge data volume in blockchain by storing transactional data on the IPFS network and packaging the returned hash of the transaction into the block in their study about IPFS-based storage architecture. As a result, data compression grows as new node synchronisation speeds up. The storage solution in this plan is IPFS because of its reasonable distributed storage performance.

Dobre et al [15] worked on developing a signature-based image authentication that would resist JPEG compression and used blockchain based technology to store the image signatures in a decentralized network. They used the transaction hash of the image signature as a payload and embedded this hash into the JPEG file’s header. We greatly appreciated the innovative approach made by Dobre et al.

Tang et al [14] explored the problem that arose with medical data leakage in most, if not all medical institutions, which compromised the patient’s private medical data. They asserted that PACS (Public archiving and communication systems) alone could not ensure the necessary secrecy, which was of the highest significance to the patients, and they emphasised that the patients had no authority over their own medical information at the time of treatment. Using the system of smart contracts for credit ratings was their model's fundamental innovation.

Dong [13] noted the variety of infringement disputes in intellectual property rights that run rampant today, ranging to all forms of creative mediums such as films and television, and other entertainment mediums such as games. Dong

proposed that blockchain technology could serve as an innovative tool in this field of IP protection and serve as a permanent solution. Dong's suggested model served key in the initial stages of research we conducted in looking for intuitive Ethereum-based secure storage models.

II. Methodology/Experimental

A. Flowchart

The program starts with the data/image being uploaded on the IPFS servers which decentralized the data information and then returns a unique CID (Content Identifier) with respect to every distinct data for example: "". After receiving the CID for the data, we send the data to Hyperledger servers (private blockchain) through the Hyperledger Fabric API's. This is done through three special Hyperledger Fabric nodes that are:
1. Endorsing Nodes – Nodes that can execute chain code, 2. Committing Nodes – Nodes that have all the copies of the Ledger,
3. Ordering Nodes – Nodes that maintain the sequence of the transactions. We have defined the Private Blockchain using Chain Code (Smart Contract for Hyperledger) which works as the control code of the ledger. Running separately from the endorsing peer process in a protected Server container is how Chain code operates. Through transactions made by apps, Chain code creates and controls the ledger state.

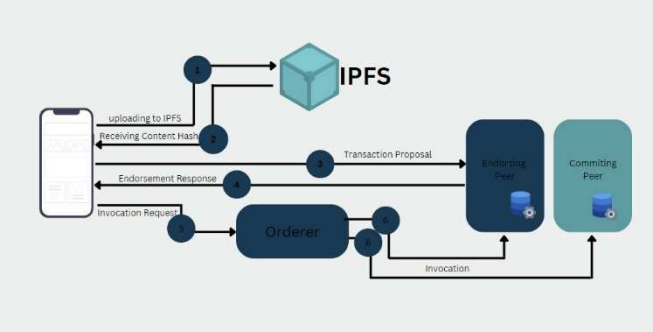

Fig 1) Network Workflow

The client initiates a transaction to the ledger which checks for authentication using the MSP (Membership Service Provider) identifying the user, moreover it checks the Access Control List for authentication of the user for the read access. There is an invocation request sent to the Hyperledger servers which runs by through the endorsing nodes that execute the instructions and then sends an endorsement response back to the user, furthermore after receiving the response another request is forwarded to the ordering nodes which queue up the request and then forward the request to the committing and the endorsing nodes and there are changes made to both the peers whilst returning the copy of the ledger thus returning the image (which is in TIFF form) which is converted to JPG format through the Convert API and then displayed to the user, thus adding multiple layers of security to the storage and browsing of the Images.

B. Method

For storage of the images, we are using IPFS, a decentralized file system, which is based on the backbone of content addressing, Merkle DAG's and DHT. The image submitted by the user is sent over the IPFS network wherein the image is passed via the SHA-256 hashing algorithm.

The SHA-256 is the only relevant algorithm in our research work. It is a one-way function that generates a 256-bit hash. A hash is nothing but a random alphanumeric string which cannot be back traced i.e., the input generates the hash but what input was initially given to the hashing algorithm can never be predicted based solely off the hash. Based upon the generated cryptographic hash, the corresponding CID of the file uploaded over IPFS is generated.

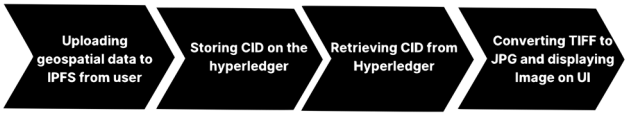

Fig 2) Real-time Flow of the Project process

CIDs (Content Identifiers) usually consists of two parts: Codec and Multihash. Codec interprets the data and multihash represents the hash type and hash value. The actual file is broken down into blocks of fixed size into a Merkle DAG. A Merkle DAG (Directed Acyclic Graphs) is a data structure wherein the traversal flow is one way, and no branches are linked together, or no loops exist. Every node of a Merkle DAG contains the part of data and CIDs of the children's nodes. Every node here acts as a root node for its own sub-Merkle DAG. Thus, retrieving data via IPFS is faster since multiple sub parts of the data can be retrieved from different peers through these sub-Merkle DAG. Which peer of the network contains the requested CID is maintained through DHT (Distributed Hash Tables). DHT is a distributed system for mapping keys and values. DHT acts as the fundamental component of the content routing since it maps what user is looking for to the peer which is storing the matching content. So basically, we can think about DHT as huge table which contains the answer to "who" has "what" data. Based upon this mapping from the DHT, the network establishes a connection between the user peer and broadcasting peer. In this way, the CID is retrieved and then appended to the IPFS gateway URL. Now, the IPFS allows multiple ways to retrieve the content hosted on the network. We can either set up an IPFS node network locally or we can use IPFS gateways that provide an HTTP based service to access IPFS content. In our research, we are using Protocol Lab's public gateway here to access the file uploaded by the user. Since geospatial image data exists in TIF (Taggable Image Format), which does not come with in-browser support, we are required to convert the image into a more common .jpg (Joint Photographic Expert Group) format to make the image

viewable on the browser. For conversion purposes, we are using ConvertAPI, where the response received from the public gateway is passed onto the API's endpoint. The API then responds to the request with the URL that contains the image uploaded by the user in JPG format. The received URL is appended on the DOM node via creating an image container. In this way, we have managed to implement successful IPFS content hosting and retrieval.

To felicitate our private/permissioned blockchain we have utilized Hyperledger fabric and Hyperledger based framework.

We chose Hyperledger because of various points, such as:

Permissioned: As the Hyperledger Fabric platform is permissioned, members know one another rather than being anonymous and hence fully untrustworthy, as is the case with a public permissionless network.

Modular: Hyperledger's modular nature is preserved. With Hyperledger, everything is created as modules that we may import as needed or remove when not in use.

Scalable: Fabric may employ consensus techniques that don't require a native coin to support the execution of smart contracts or to compensate for costly mining.

Security: The security feature consists of the Channel, the Chain-code, and the MSP (membership service provider).

Chaincode: A transaction proposal is used to trigger the execution of chaincode functions, which operate on the current state database of the ledger.

MSP: In Fabric, the Membership Service Provider (MSP) is the trusted authority responsible for managing the identities of all participants in the hidden transactional networks. These networks are designed in such a way that the identities of all participants are known. Therefore, before recording a transaction in the ledger, the MSP confirms it. MSP confirms the identities by associating Certificate Authority as identities.

III. Results and Discussions

The frontend accepts the JWT token from the user as the authentication medium to upload/access images from the IPFS platform. If the user is accessing the platform for the first time, then he is requested to enter the username to login. After successful login into the Hyperledger network, the user can upload images to the IPFS wherein the CID generated is automatically appended to the IPFS gateway URL to retrieve the data. The response received from that endpoint is then sent to API for converting its file format into a browser supported format like JPG so as the image can be displayed on the browser itself. There are no such restrictions enforced upon the count of images that can be uploaded on the network at one instance.

A. Electronic Image Files

The process of Image Conversion from TIF format to JPG format can be seen below. As mentioned previously, ConvertAPI has been utilized for the same.

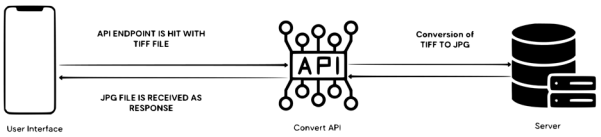

Fig 3) Data flow between the Client-Server architecture

Let us go through the entire process of accessing the proposed Decentralized Storage System, step-by-step.

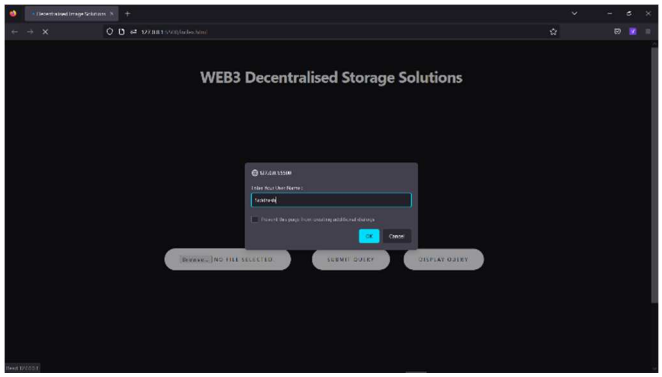

Figure 4.1) User is prompted for their Username as authentication (Frontend)

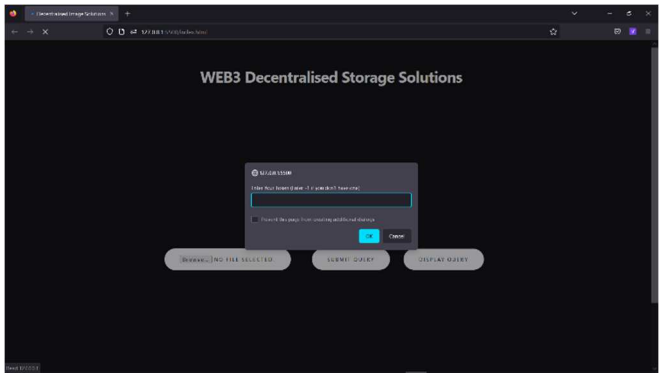

Figure 4.2) User is prompted for their Token as authentication (Frontend)

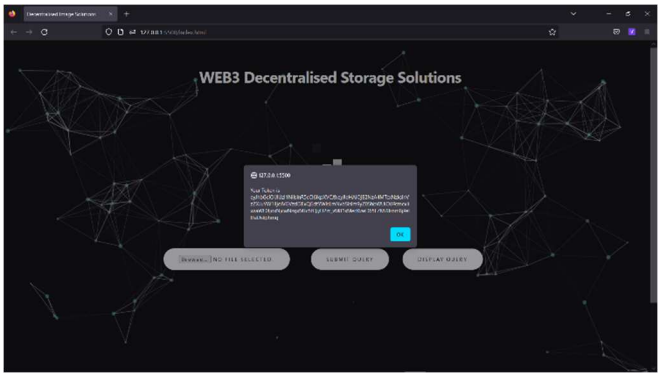

Figure 4.3) Successful Token Generation for the end-user on first login (Frontend)

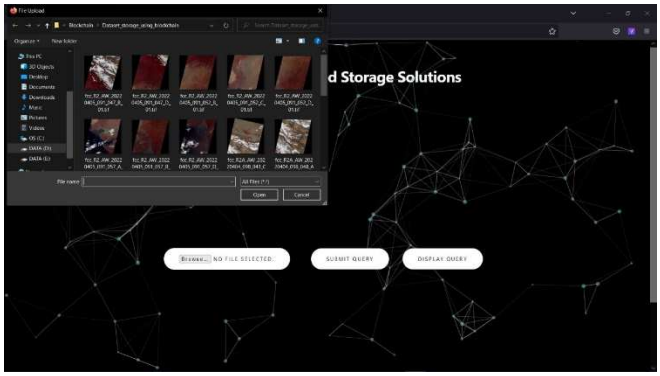

Figure 5.1) Selecting the required geospatial images for upload.

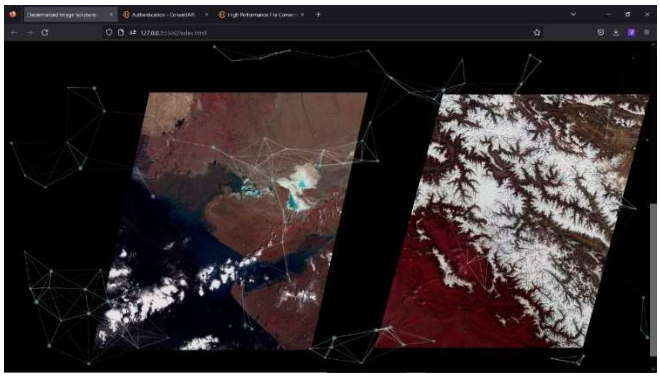

Figure 5.2) Uploaded images can be accessed and viewed on the website.

Upon further review, there are some key features and some drawbacks to an otherwise innovative decentralized storage solution. Through this work we have aimed to provide enhanced privacy, security, and enabled resistance towards censorship.

The usage of CIDs (Content Identifiers) via IPFS has added a level of accessibility that traditional URLs cannot provide, as they are tied to specific locations on the internet, while the CIDs can be accessed from any node which we require from.

Our implemented Hyperledger framework ensures that only verified authors and viewers may access the respective data. The resultant security provided ensures that no unauthorized observers may view or access the sensitive end-user data. Stored data is decentralized. Data loss due to server failure, faults, hardware losses, etc. is minimized compared to traditional centralized storage solutions as in our project, the data is stored over a network instead of singular, specific locations.

IV. Future Scope

The project in its current state works smoothly and independently, but there can be many more improvements to this project that can be done in the future to make this a deployable and secure storage solution for the public too.

The scope of data formats that can be stored in our decentralized storage system could be broadened from simply storing geospatial data. Other alternatives could be storing crucial health records, safeguarding necessary and extremely sensitive data that is critical for regional safety.

The utilized API for the format conversion (ConvertAPI) could be replaced with an in-house manufactured API as the existing model is used from a free service provided by a third-party. This retains a level of minor centralization that we seek to eliminate, as user data and records could be accessed by unauthorized entities. However, the ConvertAPI service is GDPR-compliant hence it does not raise any glaring security issues for our users and their crucial data.

The website's User Interface (UI) is not extremely user-friendly, which can be treated to create an easily usable and understandable decentralized storage space for all.

V. Conclusion

Concluding this paper, in today's world which overflows with endless data and information every minute, all sensitive and critical data is to be safeguarded at all costs. We set out with the aim to create a secure and stable storage system for such data. Through modern technologies such as Hyperledger and IPFS we have demonstrated the possibility of creating a private blockchain for satisfying this protection need.

Acknowledgment

We would like to express our gratitude to our university, Vishwakarma Institute of Technology, Pune, for giving us the chance to work on this project. We appreciate Dr. Sandip Shinde, the head of our department. We also want to express our gratitude to Prof. Swati Jadhav, who served as our project and research manager during the whole process.

References

[1] S. Ali, G. Wang, B. White, R. L. Cottrell, "A Blockchain-based Decentralized Data Storage and Access Framework for PingER," in 17th IEEE International Conference on Trust, Security And Privacy In Computing And Communications, pp. 1-8, 2018.

[2] D. Trautwein, A. Raman, G. Tyson, I. Castro, W. Scott, M. Schubotz, B. Gipp, Y. Psaras, "Design and Evaluation of IPFS: A Storage Layer for the Decentralized Web," in ACM SIGCOMM 2022 Conference, pp. 1-15, 2022.

[3] M. S. Krstić and L. J. Krstić, "Hyperledger Frameworks with a Special Focus on Hyperledger Fabric," pp. 639-663, 2020.

[4] P. Han, A. Sui, T. Jiang, C. Gu, "Copyright Certificate Storage and Trading System Based on Blockchain," in 2020 IEEE International

Conference on Advances in Electrical Engineering and Computer Applications (AEECA), pp. 1-6, 2020.
[5] Protocol Labs, “Filecoin: A Decentralized Storage Network,” Tech. rep., 2017.
[6] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System”, 2008. (Available: https://bitcoin.org/bitcoin.pdf) [Accessed: 10 December 2022].
[7] B. Wang, S. Jiawei, W. Wang, P. Zhao, “A Blockchain-based System for Secure Image Protection Using Zero-watermark,” in 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems, pp. 1-6, 2020.
[8] IPFS Ecosystem directory. https://ecosystem.ipfs.io/. 2022.
[9] Hyperledger, “Hyperledger-fabricdocs Documentation,” 2019. [e-book]. Available: https://buildmedia.readthedocs.org/media/pdf/hyperledgerfabric/release-1.2/hyperledger-fabric.pdf [Accessed: 11 December 2022].
[10] Q. Zheng, Y. Li, P. Chen, X. Dong, “An Innovative IPFS-Based Storage Model for Blockchain,” in 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 726-732, 2018.
[11] A. M. Athreya, A. A. Kumar, S. M. Nagarajath, H. L. Gururaj, V. R. Kumar, D. N. Sachin, and K. R. Rakesh, “Peer-to-Peer Distributed Storage Using InterPlanetary File System,” 2021.
[12] T. Guggenberger, J. Sedlmeir, G. Fridgen, A. Luckow, “An In-Depth Investigation of Performance Characteristics of Hyperledger Fabric,” 2021.
[13] Dong, Xueying, “A method of image privacy protection based on blockchain technology,” in *2018 International Conference on Cloud Computing, Big Data and Blockchain (ICCBB)* (2018): 1-4.pp, 2018.
[14] H. Tang, N. Tong, J. Ouyang, “Medical Images Sharing System Based on Blockchain and Smart Contract of Credit Scores,” 2018.
[15] R. A. Dobre, R. O. Preda, C. C. Oprea, and I. Pirnog, "Authentication of JPEG Images on the Blockchain," in Proc. IEEE 14th Intl. Conf. on Networking, Sensing and Control, pp. 1-6, 2018.
[16] A. Kosba, A. Miller, E. Shi, et al., "Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts," in Proc. IEEE Symp. on Security and Privacy, pp. 839-858, 2016.
[17] D. Eastlake Rd and P. Jones, "US Secure Hash Algorithm 1 (SHA1)," Internet Engineering Task Force, RFC 3174, 2001.
[18] R. Zhang, R. Xue, and L. Liu, "Security and Privacy on Blockchain," ACM Comput. Surv., vol. 52, no. 3, pp. 1-34, 2019.
[19] J. Mattke, C. Maier, and A. Hund, "How an Enterprise Blockchain Application in the U.S. Pharmaceuticals Supply Chain is Saving Lives," MIS Quarterly Executive, vol. 18, no. 4, pp. 246-261, 2019.
[20] M. Kuzlu, M. Pipattanasomporn, L. Gurses, and S. Rahman, "Performance Analysis of a Hyperledger Fabric Blockchain Framework: Throughput, Latency and Scalability," in Proc. IEEE Intl. Conf. on Blockchain, pp. 536-540, 2019.
[21] J. Swati, P. Nitin, P. Saurabh, D. Parikshit, P. Gitesh and S. Rahul, "Blockchain based Trusted Secure Philanthropy Platform: Crypto-GoCharity," 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA, Pune, India, 2022, pp. 1-8, doi: 10.1109/ICCUBEA54992.2022.10011026.
[22] L. Balduf, S. Henningsen, M. Florian, S. Rust and B. Scheuermann, "Monitoring Data Requests in Decentralized Data Storage Systems: A Case Study of IPFS," 2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS), Bologna, Italy, 2022, pp. 658-668.