# Data Storage in Blockchain Based Architectures for Internet of Things (IoT)

Munavwar Shaikh, Charles Shibu, Enrico Angeles, Deepa Pavithran

*Information Security Engineering Dept, Abu Dhabi Polytechnic*, *P.O. Box 111499*,

Abu Dhabi, United Arab Emirates.

{munavwar.shaikh, charles.gabriel, enrico.angeles,deepa.pavithran}@adpoly.ac.ae

2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) | 978-1-6654-4067-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/IEMTRONICS52119.2021.9422573

***Abstract*— Data storage in IoT systems describe how data is pushed from the sensor and where it will be stored. In Bitcoin, there is a limit to the number of transactions that can be stored in the block, similarly in Ethereum it is limited by the gas limit in block. An IoT infrastructure have several applications that includes a sensor collecting data from physical environment. Storing a large amount data in a blockchain faces several challenges due to the distributed nature of the blockchain, high transaction processing time, and reduced scalability. Hence it is crucial to identify what data should be stored and how to store it in a secure way. In this paper, we provide how data is being stored in various blockchain based IoT applications and provides data storage compliance in treating IoT data in a blockchain environment.**

***Keywords—blockchain, Internet of things, data storage, sensors, data compliance, cloud.***

## I. INTRODUCTION

Blockchain is a recent technology that can provide data security and privacy in IoT Systems. It uses a distributed ledger design that stores data in multiple nodes within the network. In a decentralized system, computing and storage tasks are redundant, which means every node in the decentralized network must store a complete copy of all data and perform the same operations of data. This is how blockchain maintains trust among nodes and provide security to the data. However, when the number of participating nodes increases or the amount of data to be stored is high, it will not be practical to replicate the entire data into all participating nodes and to perform concurrent computations on it.

The existing internet of things systems uses either offline or cloud-based information-sharing technology. It has several shortcomings including the high maintenance cost, low efficiency, and lack of an effective mechanism to ensure IoT equipment Identity, information validity, authenticity, consistency, and integrity of information in different systems [1].

Blockchain systems do not rely on trusted third parties. All nodes make decisions together to verify the legitimacy of the transaction. Even if some nodes are attacked or destroyed, it will not cause damage to the entire blockchain system.

## II. BACKGROUND

### A. Blockchain

Blockchain is a distributed ledger where blocks are connected with cryptographic hashes and are distributed over all the nodes in the system. Generally, there are two types of blockchain: public and permissioned. In a public blockchain, anyone can join the network or participate in the consensus mechanism. An example of a public blockchain is Bitcoin. In a permissioned or private blockchain, only certain entities will have access to participate in the blockchain process. This is typically used in applications that involve multiple entities in a restricted way. It will have a certain control on who can access the data and who can participate in the consensus. The consensus is an agreement on how and what data are added to the blockchain.

### B. Internet of Things

As defined in [2] Internet Of Things (IoT) is an emerging global Internet-based information architecture facilitating the exchange of goods and services in the global supply-chain network. The IoT data comes from a large number of devices generating billions of data objects. The IoT network has to collaborate with different devices to sample, process, and make this data useful for analytical and decision making. Realizing the full potential of IoT has not yet been achieved. This is due to the lack of standards, heterogeneous nature of devices, diversified communication protocol, and security [3].

## III. DATA STORAGE IN BLOCKCHAIN BASED IOT USECASES

A majority of blockchain-based IoT Application is in the field of healthcare, Smart home, smart city, Industrial IoT and agricultural applications.

### A. IoT Devices employing blockchain

The blockchain-based design for IoT data storage takes a distributed access control and data management approach. In this scenario, data ownership is transferred to users instead of following the traditional centralized access control system based on trust modeling. With the use of blockchain-based functions, secure and fail-safe data management with a verifiable and distributed access control layer on the storage layer is practiced. The storage of time-critical IoT data is facilitated at the edge of the network via a location-based decentralized storage system, which in turn is managed with blockchain technology [4]. The parallelism of large storage systems is suggested by Quanqing Xu[5] in order to shorten the execution time of many basic data analysis tasks, whereby blockchain can be used as intelligent contracts to facilitate the negotiation of a contract in the IoT and also to enforce it. OSD-based smart contract (OSC) approach is used in which IoT devices interact with

978-1-6654-4067-7/21/$31.00 ©2021 IEEE

such blockchains. For data analysis applications, the IoT device processors perform application-specific operations. That way, only the results are returned to the clients, rather than the data files they read [5].

### B. Healthcare

Since the advent of smart IoT devices in the Healthcare industry, promising technologies such as cloud computing, ambient assisted living, big data, and wearables are being applied in various platforms within Healthcare [6]. E-health regulations related to data and policies worldwide to determine how they assist the sustainable development of IoT and cloud computing in the healthcare industry evolved as well with the thriving data in all these platforms [6]. The Medical data of a patient is treated with utmost care and secrecy. Hence the IoT devices in healthcare systems are devised to ensure the security of patients' healthcare data, realize access control for normal and emergency scenarios, and support smart deduplication to save the storage space in big data storage system. The medical files generated by the healthcare IoT network are usually encrypted and transferred to the storage system, which can be securely shared among the healthcare staff from different medical domains leveraging a cross-domain access control policy. A smart secure deduplication method could be followed as proposed by Tang et al. [7] to ensure the medical files or data can be accessed by all the data users after deduplication and authorized by the different original access policies.

### C. Industrial IoT(IIoT)

The main goal of industrial IoT is to improve operational efficiency, increase the scale of production and better management of industrial components and processes through product customization, condition of machines getting monitored in an intelligent way, smart monitoring applications for production workshops as well as predictive and preventive maintenance of industrial equipment [8].
In industrial usage of IoT devices, storage and retrieval are done by integrating fog computing and cloud computing [9] and a flexible and economic framework for data processing was made by the authors. The data in these devices are preprocessed in the edge server with the addition of timestamps and stored locally. The remaining data gets sent to the external storage medium and away from the local storage for information retrieval and data mining [9]. This set of data is not time-sensitive.

### D. Smartcity/Smart Home

Smartphone applications are used to control and monitor home functions using wireless communication technologies. Domb in [10] examines the concept of the smart home and the data associated with it, incorporating IoT services and cloud computing It also discusses embedding intelligence in sensors and actuators, networking intelligent devices using the appropriate technology, saving storage space and improving data exchange efficiency, and facilitating data interaction with intelligent things using cloud computing and storage for easy access to various locations. A precise, fast, open and shared information system is the basis for Smart City applications. In view of this massive, distributed, heterogeneous and complex state data, the storage and management of traditional data will encounter great difficulties. The traditional infrastructure uses a centralized approach, an expensive large server, hard disk storage hardware, and a relational database management system. This leads to poor scalability of the system, higher costs and, for the most part, difficult adaptation to the demand for higher reliability of real-time status data from smart city applications [11].

### E. Agricultural Applications

Some IoT-based smart farming surveillance systems use a two-tier approach to data storage to store and secure large amounts of data from any IoT device. Tier-1 focuses on collecting data from various sensors and storing them locally using the SD card. Tier-2 uses a cloud server to store the large volume of IoT sensor data [12]. Gill et al. [13] explains the benefits of a cloud-based autonomous information system for the delivery of Agriculture-as-a-Service (AaaS) using cloud and big data technologies. Accordingly, the Aaas system collects information from various users via pre-configured devices and IoT sensors and processes it in the cloud with the help of big data analysis and provides the users with the necessary information automatically.

## IV. Data Storage Architectures in IoT Blockchain

### A. Storage in cloud

Permissioned blockchains use cloud servers to store encrypted data blocks. All transactions in a different block and create a combined hash of each block using Merkle Tree and transfer it to the distributed network. This way, any changes in cloud data can be easily detectable. Doing the storage in this manner also preserves decentralization to some extent [14]. Fig.1 shows blockchain based IoT architecture where a cloud storage is being used to store the IoT Data. Cloud storage uses users' data in an identical group. The Hash of the data stored in the cloud is sent to the overlay network. The hierarchical storage structure in storing the majority of the blockchain in the clouds helps faster upload of one day's blockchain to the cloud, it maintains the most recently created blocks in a blockchain overlay network. The blockchain connector and the cloud connector are the two software interfaces that are defined to create the blocks and coordinate to the clouds but one disadvantage of this method is a need to be implemented in a more real IoT application to find out if this technology is reliable [15].

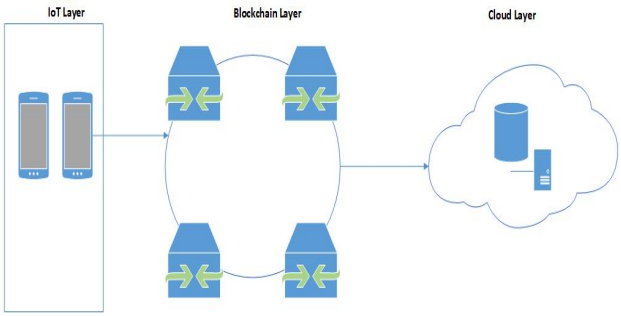

Fig:1 Storage in cloud

*B. Storage in Decentralized way*

This is a fully decentralized way of storing data. Liu et al. [16] integrates decentralized blockchain network and distributed storage network. Data produced by the IoT device is stored in the distributed storage nodes in the peer-to-peer network, whereas the reference to data that serve as the identifier is stored in the Blockchain. Blockchain only stores the digest of data but not data itself. Hence the amount of data on the blockchain is greatly reduced. The use of blockchain technology in storing data and with the use of management hub including the integration of IoT and decentralized access control to the blockchain solves the issues in managing several controlled IoT devices, it was able to cope up with different IoT scenarios but there are possibilities of threat and vulnerabilities in the IoT since it is no longer part of the blockchain technology [17]. Fig.2 shows a blockchain architecture that uses decentralized storage for storing IoT data. Decentralized cloud architecture was used in which small-scale data centers meet low latency and high bandwidth because these data centers are located closer to the users. The locality-aware data storage and the processing development provide its full potential with the decentralized access control layer. Data streams that are chucked at pre-defined lengths show reasonable results but this one is still in the initial stage and needs improvement[4]. Sapphire system which is a large-scale blockchain storage system used for data analytics in the IoT it uses an object-based storage interface that provides richer semantic information for the stored object to optimize its performance more effectively than other storage systems[5]. A secure structure for IoT data storage and protection which is based on blockchain technology incorporating edge computing helps manage data storage and assist small IoT devices to accomplish computations. The authentication system adapted the certificate-less cryptography, but it needs to improve the authentication scheme for the blockchain-based systems[18]. The integration of IoT, blockchain, and cloud technologies allows observation of vital signs of the patient in a protected approach. The storage and access of data used in the blockchain technology data sharing significantly increase overall throughput but need to be tested using various IoT frameworks[19]. Blockchain technology solves the problem of IoT information sharing security but performance needs to be enhanced when applied to a specific industry [1].

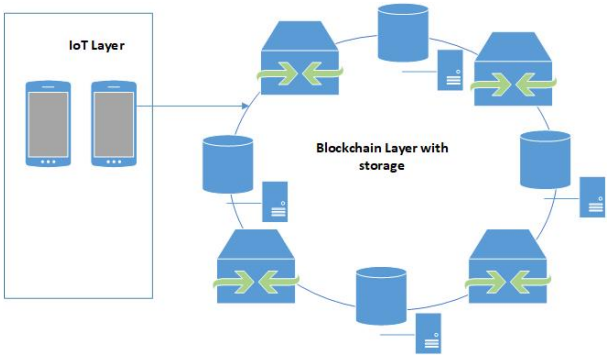


Fig:2 Storage in Decentralized way

*C. Storage in Hybrid way*

In this method data will be stored in a hybrid way, where processed/ or raw data from the sensor is stored in a cloud server and a reference to the data will be stored in the blockchain. Si et al. [1] propose a double-chain model combining data blockchain and transaction blockchain. IoT data is divided into lightweight data and multimedia data, where multimedia data is compressed and integrated to reduce data capacity and to improve data quality. The processed data is divided into account book data and outsourced storage data (multimedia data) stored in a fog node that can be easily downloadable. Fig.3 shows blockchain architecture that uses hybrid storage. A comparison of various storage options in IoT based blockchain architectures is provided in Table1.

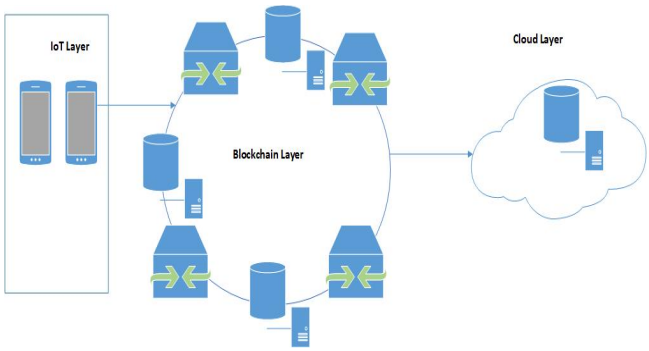


Fig:3 Storage in a Hybrid Way

Table 1: Comparison of various data storage options in IoT Blockchain

| Type | Storage Design | Advantages | Disadvantages | Reference paper |
|---|---|---|---|---|
| Cloud | Hierarchical storage structure | • Faster upload of one-day's blockchain to the cloud | • Not yet implemented in a more real IoT application | [15] |
| Decentralized | Blockchain system with Management Hub | • It performs best with the use of management hub node | • It might face different threat and vulnerabilities since IoT is no longer part of the blockchain technology | [20] |
| Decentralized | Data streams are chucked at pre-defined lengths | • Initial results show reasonable overhead | • Still at the initial stage and needs for improvement | [4] |
| Decentralized | Object-based Storage interface | • Has richer semantic information for the stored object to optimize its performance more effectively than other | | [5] |

| | | | | |
|---|---|---|---|---|
| | | storage systems | | |
| Decentralized | Blockchain system combined with Edge Computing | • Secure scheme for IoT data storage and protection | • Need to improve authentication scheme for blockchain-based system | [18] |
| Decentralized | Data Sharing using Blockchain technology | • The overall throughput increased significantly | • Need to be tested using various IoT frameworks | [19] |
| Decentralized | Blockchain technology | • Solve the problem of IoT information sharing security | • Performance needs to be enhanced when applied to a specific industry. | [1] |

## V. Compliance on Data Storage

To start with knowing the objectives of the organization. This interprets to knowing which regulations will apply to the organization directly related to the regional legal and regulatory requirement and which types of data are expected to manage or not, how long to retain it and how to protect it. In the organization the compliance officers can be the ones to provide information.

Organizations need to focus on data classification and data mapping. Data classification and data mapping are crucial in discovering the types of information that are being held in storage systems and how they are being moved across the network. Not only are they essential factors in determining how regulated information is stored, but also a solid step in establishing compliant policies.

**Continual monitoring**: Storage compliance is not a set-and-forget affair. Continual monitoring is key in ensuring that regulated data is properly cared for during its lifecycle. Procedures must be in place to make sure this monitoring happens regularly.

**More than security**: Security and storage often go hand-in-hand, but storage compliance takes it to another level. Encryption will help, and many storage solutions feature support for the security-enhancing technology, giving both storage and security professionals one less thing to worry about.

**Critical:** Testing and audits will ensure compliance with policies and IT mechanisms are up to the task. It's best to work out the kinks now before having to explain to investigators why sought-after emails or transaction records have gone missing. There are many requirements based on the type of information and data the organization has. Some might involve using what is called DAR (Data Encryption at Rest), which encrypts the storage device so that if removed from the system, the data is nearly or totally impossible to access (the degree of difficulty depends on the encryption algorithm and the size, complexity and entropy of the key or keys for the device).

Understanding what is required from a governance point of view for the organizational data or the resulting information is based on things like best practices for the industry or regulations and agencies like the U.S. National Bureau of Standards (NIST), ISO, HIPAA, SEC, GDPR in Europe. And the resulting architectural or procedural changes are the types of things that will be needed to address as part of the architecture. Compliance is not easy, nor is it free. The cost depends on lots of factors but trying to force compliance after the architecture is planned and built is always far most costly than doing it beforehand.

**Decentralization**: The main benefit of using blockchain would be that no single authority would have control over the data generated by the IoT devices. That way, a distributed peer-to-peer network is born that permits the parties that don't know or trust each other to collaborate more smoothly. This type of network will also make it possible to unify IoT devices and streamline the distribution of updates throughout the network.

**Security:** The current security architecture of IoT has its shortcomings. When the data is managed by a central authority, the system is more susceptible to a single point of failure. Blockchain's unique security protocol normally described as transparent and immutable is a good solution to the largest issue of IoT development. Blockchains will store unalterable data history that can be consulted for each unique address. This lays the foundations of a platform that provides improved identification and authentication in IoT. The robust level of encryption that blockchain guarantees won't let the hackers overwrite data records.

**Transparency**: Anyone with authorization could track the transactions made on the network to follow up on what has happened in the past. This feature is useful to identify any leakage and take action.

**Autonomy:** Blockchain will reinforce the machine-to-machine economy that IoT is based on by offering a safe way to store information on different transactions. That way micropayments for services and data can be processed in a straightforward way. IoT devices that rely on blockchain can execute digital agreements automatically when the terms are met. Automating transactions between devices improves machine-to-machine communication.

**Reduced costs:** Integrating blockchain to the organizational processes would allow IoT companies to reduce costs. Eliminating the massive overhead costs related to IoT gateways will help them to reduce the strain on the company budget.

When defining compliance requirements, one should be looking to the future rather than the present because of the cost and challenge of shoehorning things in after the fact. That means that someone needs to be continuously studying compliance requirements in the given industry to which the organization belongs, along with best practices. Data will only become more important in the future, and we need to be up to the challenge.

## VI. Conclusion

Due to the heterogeneity of applications, the amount of data generated by IoTs differs. Identifying what data to be added to blockchain is a primal task. Healthcare data should be treated in a confidential way and hence utmost care should

be taken in sharing and storing of data. The IIoT data use external storage medium away from the local storage for information retrieval and data mining. This is achieved through Edge/Fog computing. Agricultural applications either store data locally or to cloud if the data generated is large. The decentralized nature of the blockchain requires the data to be stored in multiple locations. The most common method of storing data is in still in the cloud. However, decentralized ways integrates blockchain with distributed storage network. Hybrid storage is gaining more popular which divides the data into two, the one to be added to the blockchain network and the other to the cloud data storage. This is a more economical way that implements the concepts of blockchain while maintaining the network latency. The IoT will enable electronic objects to exchange huge amounts of data; storing it in a reliable way will be challenging. Organizational compliance group will know best how long data or information is required, but there are many other requirements that will have to address to ensure that the business objectives in the areas of performance, availability and data integrity, all of which need to be address for the life of the data and information.

## REFERENCES


[1] H. Si, C. Sun, Y. Li, H. Qiao, and L. Shi, "IoT information sharing security mechanism based on blockchain technology," *Futur. Gener. Comput. Syst.*, vol. 101, pp. 1028–1040, 2019, doi: 10.1016/j.future.2019.07.036.

[2] R. H. Weber, "Internet of Things - New security and privacy challenges," *Comput. Law Secur. Rev.*, vol. 26, no. 1, pp. 23–30, 2010, doi: 10.1016/j.clsr.2009.11.008.

[3] S. A. Bragadeesh and A. Umamakeswari, "Role of blockchain in the Internet-of-Things (IoT)," *Int. J. Eng. Technol.*, vol. 7, no. 2, pp. 109–112, 2018, doi: 10.14419/ijet.v7i2.24.12011.

[4] H. Shafagh, L. Burkhalter, A. Hithnawi, and S. Duquennoy, "Towards blockchain-based auditable storage and sharing of iot data," *CCSW 2017 - Proc. 2017 Cloud Comput. Secur. Work. co-located with CCS 2017*, pp. 45–50, 2017, doi: 10.1145/3140649.3140656.

[5] Q. Xu, K. Mi, M. Aung, Y. Zhu, and K. L. Yong, "A Blockchain-Based Storage System for Data Analytics in the Internet of Things Quanqing," *New Adv. Internet Things*, vol. 715, pp. 119–138, 2018, doi: 10.1007/978-3-319-58190-3.

[6] L. Minh Dang, M. J. Piran, D. Han, K. Min, and H. Moon, "A survey on internet of things and cloud computing for healthcare," *Electron.*, vol. 8, no. 7, pp. 1–49, 2019, doi: 10.3390/electronics8070768.

[7] Y. Yang, X. Zheng, W. Guo, X. Liu, and V. Chang, "Privacy-preserving smart IoT-based healthcare big data storage and self-adaptive access control system," *Inf. Sci. (Ny).*, vol. 479, pp. 567–592, 2019, doi: 10.1016/j.ins.2018.02.005.

[8] W. Z. Khan, M. H. Rehman, H. M. Zangoti, M. K. Afzal, N. Armi, and K. Salah, "Industrial internet of things: Recent advances, enabling technologies and open challenges," *Comput. Electr. Eng.*, vol. 81, p. 106522, 2020, doi: 10.1016/j.compeleceng.2019.106522.

[9] J. S. Fu, Y. Liu, H. C. Chao, B. K. Bhargava, and Z. J. Zhang, "Secure Data Storage and Searching for Industrial IoT by Integrating Fog Computing and Cloud Computing," *IEEE Trans. Ind. Informatics*, vol. 14, no. 10, pp. 4519–4528, 2018, doi: 10.1109/TII.2018.2793350.

[10] M. Domb, "Smart Home Systems Based on Internet of Things," *Internet Things Autom. Smart Appl.*, vol. 11, no. 2, pp. 260–267, 2020.

[11] H. Y. Shwe, T. K. Jet, and P. H. J. Chong, "An IoT-oriented data storage framework in smart city applications," *2016 Int. Conf. Inf. Commun. Technol. Converg. ICTC 2016*, pp. 106–108, 2016, doi: 10.1109/ICTC.2016.7763446.

[12] M. S. Ahmad and A. U. Zaman, "IoT-Based Smart Agriculture Monitoring System with Double-Tier Data Storage Facility," pp. 99–109, 2020, doi: 10.1007/978-981-15-3607-6_8.

[13] S. S. Gill, R. Buyya, and I. Chana, "IoT based agriculture as a cloud and big data service: The beginning of digital India," *J. Organ. End User Comput.*, vol. 29, no. 4, pp. 1–23, 2017, doi: 10.4018/JOEUC.2017100101.

[14] A. D. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A decentralized privacy-preserving healthcare blockchain for IoT," *Sensors (Switzerland)*, vol. 19, no. 2, pp. 1–17, 2019, doi: 10.3390/s19020326.

[15] G. Wang, Z. Shi, M. Nixon, and S. Han, "ChainSplitter: Towards blockchain-based industrial IoT architecture for supporting hierarchical storage," *Proc. - 2019 2nd IEEE Int. Conf. Blockchain, Blockchain 2019*, pp. 166–175, 2019, doi: 10.1109/Blockchain.2019.00030.

[16] S. Liu, J. Wu, and C. Long, "IoT Meets Blockchain: Parallel Distributed Architecture for Data Storage and Sharing," *Proc. - IEEE 2018 Int. Congr. Cybermatics 2018 IEEE Conf. Internet Things, Green Comput. Commun. Cyber, Phys. Soc. Comput. Smart Data, Blockchain, Comput. Inf. Technol. iThings/Gree*, pp. 1355–1360, 2018, doi: 10.1109/Cybermatics_2018.2018.00233.

[17] O. Novo, "Blockchain Meets IoT: An Architecture for Scalable Access Management in IoT," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1184–1195, 2018, doi: 10.1109/JIOT.2018.2812239.

[18] R. Li, T. Song, B. Mei, H. Li, X. Cheng, and L. Sun, "Blockchain for Large-Scale Internet of Things Data Storage and Protection," *IEEE Trans. Serv. Comput.*, vol. 12, no. 5, pp. 762–771, 2019, doi: 10.1109/TSC.2018.2853167.

[19] D. H. Wang, "IoT based Clinical Sensor Data Management and Transfer using Blockchain Technology," *J. ISMAC*, vol. 2, no. 3, pp. 154–159, 2020, doi: 10.36548/jismac.2020.3.003.

[20] O. Novo, "Blockchain Meets IoT: An Architecture for Scalable Access Management in IoT," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1184–1195, 2018, doi: 10.1109/JIOT.2018.2812239.