Secure Distributed Storage System for Large-scale IOT Data Using Blockchain

Renuka Dhole[1], Pooja Jadhav[2], Shrishti Dixit[3], Harshada Dhaigude[4], Dr. K.S. Wagh[5]
AISSMS IOIT, Computer Department,
SPPU, Pune, Maharashtra, India.

Abstract: Internet of Things is stated as things or objects that are connected to each other over the internet. IoT can connect a variety of physical objects, to reach common goals. In some of the IoT applications data can be stored in distributed hash tables while the DHTs stored address can be stored in Blockchain. IoT devices have low computational powers coming and they are not capable of conducting complex computations. In a traditional cloud-based IoT structure, a centralized cloud server collects and controls all the data, which brings two drawbacks: The cloud server needs very high storage capacity to store the IoT data and Sensitive data can be easily leaked from the server. Development of the IoT has made progress in recent years. Storing and protecting this huge volume of IoT data has become a significant issue. Traditionally, cloud based IoT structure were used but they lead to high computation and storage demands on the cloud servers. Due to centralized servers, there were many trust issues. To solve these problems, we have proposed a distributed data storage scheme employing Blockchain and certificate less cryptography. Blockchain serves as an unchangeable ledger that allows transactions take place in a Decentralized manner. To the best of our knowledge, this is the first work designing a secure and responsible IoT storage system using Blockchain.

Keywords: Certificateless Cryptography, Blockchain, IoT, Proof of Work.

I. INTRODUCTION

The Internet of things is the network of physical devices embedded with electronics, software, sensors, actuators which enables these things to connect, collect and exchange data. With the help of IoT technology, the physical devices can communicate and interact over the Internet.

Based on this, IoT data can be stored in two ways first is in centralized manner and the other can be in Decentralized manner. A centralized storage is a database that is located, stored, and maintained in a single location. This location is mostly a central computer or database system, for example a desktop or server CPU. Whereas in Decentralized storage it is, a database that does not have any centralized server, in this IOT data is stored in a distributed way, which will be easily available to be access by a particular entity. There are some drawbacks of centralized storage such as delay in sending information to the entities, higher risk of loss of data, increase in handling cost, not suitable for storing large-scale data etc. So to overcome these drawbacks decentralized storage systems are used.

A Blockchain is a growing list of records, called blocks, which are linked together using cryptography. Each block contains a hash value of the previous block, a timestamp, and data. We can say that the blockchain is a decentralized technology. Blockchain technology is used to manage the database that records Bitcoin transactions. That means, its network, and not any central authority manage Bitcoin. Blockchain works as a peer-to-peer network without a trusted third party. In this network, a user creates a transaction and sends it to a peer-to-peer (P2P) network. The users in the P2P network will utilize an algorithm to determine how to write transactions into an empty block, and the transactions will be validated only after they are written into a block. As time goes on, there will be more and more blocks forming a blockchain.

Here, we are creating a Blockchain interface for storing the data and securing it from intruders. We are taking the raw data of transport for London which is live data and storing this data into our blockchain server. The data is all about the condition of roads in London, whether the roads are good or closure. It is live data which is updating continuously. So, we are giving the URL and the data will come automatically and get stored in blockchain server. Firstly, we are collecting the raw data which will show the status of all the roads. Than we will filter the data as we are focusing on the roads which are not in good condition. So, our filter data will show the details of the roads which are not in good condition. The data is in the dictionary form, which is a key-value pair. After collecting the data we will create the blocks. There can be any number of blocks. The blocks will contain the filter data. We are encrypting our data using hashlib, which contains different hashing algorithms, and making a fixed size of key i.e, of 256 bit. The size of our data does not matter as we are storing the key which is of fixed size. Once, the blocks are created than we can connect each block with a chain and all the blocks are linked with each other, therefore, blocks cannot be changed, and this is what we call blockchain. Every block is dependent on the previous block. The first block in the blockchain is called as genesis block it is also referred as block zero. The second block to be added on top of zero block would referred as block one. A particular block will have a nonce value i.e. the number used only once is a unique value set for each block, it will also have the hash values of previous and next block of the particular block. The genesis block will have previous hash value as 0's as it is the first block in the blockchain. We have done the risk analysis to some extend and we observed that the merging of chain is difficult. And if key is stolen than data accessing is difficult.

[1] M. Andrychowicz, S. Dziembowski, D. Malinowski, and L. Mazure. Bitcoin is a decentralized digital currency. Its features include lacking of central authority that control transactions, list of transactions is publicly available, syntax allows more advanced transactions than simply transferring the money. This paper shows how these properties can be used in the area of secure multiparty computation protocols (MPCs). Firstly, it is shown that the Bitcoin system provides an attractive way to construct a version of "timed commitments". Secondly, a concept of multiparty protocols that work "directly on Bitcoin" is introduced. It is observed that the Bitcoin system can be used to go beyond the standard "emulationbased" definition, by constructing protocols that link their inputs and the outputs with the real Bitcoin transactions. The mentioned protocols guarantee fairness for the honest parties no matter how the loser behaves.

[2] R. Li, T. Song, N. Capurso, J. Yu, J. Couture and X. Cheng. IoT is changing human lives by connecting everyday objects. RFID tag can be attached to each product which, when placed into a smart shopping cart, can be automatically read by a cart equipped with an RFID reader. As a result, billing can be conducted from the shopping cart itself, which prevents customers from waiting in a long queue at checkout. Another benefit of this kind of system is that inventory management becomes much easier, as all items can be automatically read by an RFID reader instead of being scanned manually. To validate the feasibility of such a system, design requirements of a smart shopping system are identified, a prototype system is built to test functionality, and design a secure communication protocol to make the system practical.

[3] T. Song, R. Li, B. Mei, J. Yu, X. Xing, and X. Cheng. The IoT has developed extraordinarily in recent years. There are quite a few smart home systems that have been developed by major companies to achieve home automation. In proposed scheme, data transmissions within the smart home system are secured by a symmetric encryption scheme with secret keys being generated by chaotic systems. Message Authentication Codes (MAC) scheme is incorporated to guarantee data integrity and authenticity. A detailed security analysis and performance evaluation in comparison with the previous work in terms of computational complexity, memory cost, and communication overhead.

[4] I. Stojmenovic and S. Wen. Fog Computing is a paradigm that extends Cloud computing and services to the edge of the network. Fog provides data, compute, storage, and application services to end-users. The motivation and advantages of Fog computing, and analyze its applications in a series of real scenarios, such as Smart Grid, smart traffic lights in vehicular networks and software defined networks is elaborated. Security and privacy issues are further disclosed according to current Fog computing paradigm.

[5] S. Yu, C. Wang, K. Ren, and W. Lou**.**Cloud computing is an emerging computing paradigm in which resources of the computing infrastructure are provided as services over the Internet. The problem of simultaneously achieving fine-grainedness, scalability, and data confidentiality of access control actually still remains unresolved. This paper addresses this challenging open issue by, on one hand, defining and enforcing access policies based on data attributes, and, on the other hand, allowing the data owner to delegate most of the computation tasks involved in fine grained data access control to untrusted cloud servers without disclosing the underlying data contents.

[6] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu. Edge computing is the practice of processing data near the edge of your network, where the data is being generated, instead of in a centralized data processing warehouse. With the push from cloud services and pull from IOT, we visualize that the edge of the network is changing from data consumer to data producer as well as data consumer. In edge computing each device in a network plays its own role in processing data. Moreover, the energy consumption could also be reduced by 30% to 40% by cloudlet offloading. There's a lot of hype about edge computing, but processing data at the edge presents some tough problems. Some solutions are emerging. In edge – based systems, which some call "near cloud", the goal is to extend the boundary of the cloud to be closer to the edge.

[7] S.-H. Seo, M. Nabeel, X. Ding, and E. Bertin**.** In p2p networks, a valid signature scheme plays an important role in identification of users, authenticity and integrity of sharing information. In this paper, we propose an efficient and secure threshold signature scheme without parings and a trusted party based on CL-PKC, which fits the characteristics of p2p networks. This will solve the risk of single-point failure caused by almost signature schemes. Moreover, this scheme is more computationally efficient than other schemes built from pairings.

[8] G. Zyskind, O. Nathan et al**.** This paper describes a decentralized personal data management system that ensures users own and control their data. The strong dependencies on the centralized servers bring significant trust issues. To solve this problem a decentralized data storage scheme employing blockchain has been proposed. A protocol that turns a

blockchain into an automated access-control manager that does not require trust in a third party has been implemented. Unlike bitcoin, transactions in our system are not strictly financialthey are used to carry instructions, such as storing, querying and showing data. Finally, possible future extensions to block chains that could tackle them into a well-rounded solution for trusted computing problems in society have been stated. They have combined blockchain storage to construct a personal data management platform focused on privacy.

[9] L. Jiang, L. Da Xu, H. Cai, Z. Jiang, F. Bu, and B. Xu**.** Because of RFID and sensor network technology, common physical objects can be connected. Such a network brings a series of challenges for data storage and processing in a cloud platform. IoT data can be generated quite rapidly, the volume of data can be huge and the types of data can be various. To solve these potential problems, this paper proposes a data storage framework not only enabling efficient storing of huge amount of IoT data, but also integrating both structured and unstructured data. This data storage framework is able to combine and extend multiple databases and Hadoop to store and manage diverse types of data collected by sensors and RFID readers.

[10] TK. Christidis and M. Devetsikiotis**.** Motivated by the recent explosion of interest around blockchains, this paper examine whether they make a good fit for the IoT sector. Blockchain allow us to have a distributed peer-to-peer network where nontrusting members can interact with each other without a trusted intermediary, in a variable manner. We review how this mechanism works and also look into smart contracts scripts that reside on the blockchain that allow for the automation of multi-step processes. It describe how a blockchain-IoT combination: facilitates the sharing of services and resources leading to the creation of a marketplace of services between devices and allows us to automate timeconsuming workflows. It also point out certain issues that should be considered before the deployment of a blockchain network in an IoT setting: from transactional privacy to the expected value of the digitized assets traded on the network.

[11] Wei Wang , Peng Xu , Laurence T. Yang. Investigated a secure cloud-assisted IoT data managing method to keep data confidentiality when collecting, storing and accessing IoT data with the assistance of a cloud with the consideration of users' increment.

[12]Madhusudan Singh, Abhiraj Singh, Shiho Kim. The blockchain technology is explained and infrastructure of IoT which is based on Blockchain network and a model has been provided for the security of internet of things using blockchain.

[13] Ata Ullah , Iqra Sehr , Muhammad Akbar , Huansheng Ning. Proposed Secure De-duplicated Data Dissemination (S-DDD) scheme for healthcare IoT scenario using FoG servers at the edge of the network also proposed a lightweight de-duplication mechanism that includes adaptive chunking algorithm (ACA). Moreover, A symmetric key based encryption mechanism for healthcare data exchange from smart devices towards a collector node.

[14] K. Mohanram , T. T. Mirnalinee. This paper propose new techniques for the new secure and effective storage of transport data consisting of vehicles, registrations and payments in the e-Services Web Portal of Transport Department.

II. SYSTEM ARCHITECTURE:

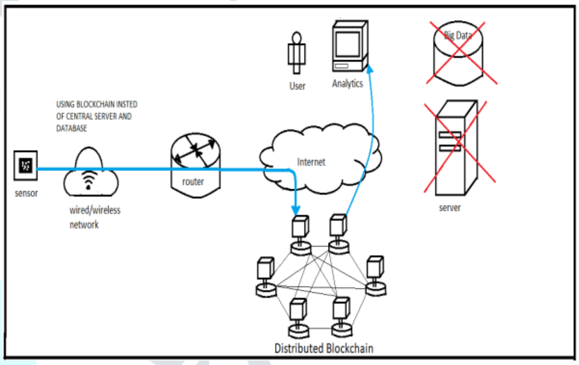

Fig 1. System Architecture

The architecture; the flow of our system is shown in Fig 1. The data will be collected from the sensors via internet. Further, the data will be then encrypted and stored in blocks. As time goes, the process will continue and the data will be added to the blocks creating chain of blocks that is blockchain. As we are using Blockchain so there will not be any centralized servers hence there will not any risk of hacking of data or no issue of any third party. The data stored in blocks will be immutable unchangeable and other users will not be able to see the data as it will be in encrypted form. We are eliminating the drawbacks of centralized servers, by using the decentralized storage system.

III. ALGORITHM

Proof of Work Algorithm

In Proof of Work algorithm, an actor is elected as a leader and chooses the next block to be added. Actors need to find a solution to a particular mathematical problem. When a new block is *mined*, that *miner* gets rewarded with some currency. Validity of the other nodes is done by other nodes. This is done by checking that the hash of the data of the block is less than a pre-set number. Due to the limited computational power, miners are also incentivized not to cheat. Attacking the network would cost more because of costly hardware, energy, and potential mining profits missed.

Challenge-response protocols assume a direct interactive link between client and the server. Fig.2 shows the implementation of the protocol

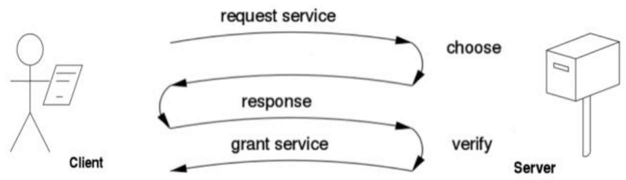

Fig 2. Challenge and response working

Solution-verification protocol helps problem to be self-imposed before a solution is sought by the requester, and the provider must check both the problem choice and the found solution. Fig.3 shows how the solution-verification process is carried out.

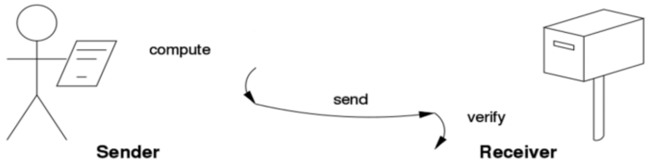

Fig. 3. Solution verification process

The transactions are gathered by the nodes. The nodes then form the blocks. In this each block must refer to the previous block. This process repeats every 10 minutes. But multiple miners trying to solve for the same block. To select a miner to be assigned who to solve a block of transactions, we create a mathematical puzzle that is hard to solve, i.e. it takes a lot of work. The puzzle often is a long number (hash value) where the computer tries to guess each number. The miner who solves the puzzle first gets to mine the block.

We have created an almost immutable blockchain. We could create new blocks really quickly.

Advantages of POW

Complete decentralization achieved. Easy to implement. Huge cost and computing power required to destroy the system.Protection from DDoS-attacks (Distributed Denial of Service)

Disadvantages of POW

Waste of energy and resource.For resolving a proof of work problem miner must have a computer, which is expensive.

IV. RESULT AND DISCUSSION

A blockchain is a decentralized ledger that contains connected blocks of data. The ability to create/store/transfer digital assets in a distributed, decentralized and tamper-proof way is of great practical value for future systems. A key challenge in the deployment of Blockchain as a Service (BassS) for system is the hosting environment. Edge devices are often too constrained regarding computational resources and available bandwidth leading to cloud as potential hosts. This paper evaluates the use of the cloud as possible platforms. Analysis of performance indicates that the network latency is the dominant factor.

The data acquisition model gathers data from the site and the blocks are created successfully. The individual blocks are then combined in the next module to form a blockchain. This provides the security to the data and prevents it from getting tampered.

V. CONCLUSION

Thus we have studied the proposed system and the technology Blockchain in deep and done all the documentation work. This project has given a brief overview of data analysis modelling in IoT. This project is proposed by comparatively analyzing the different techniques used for secure communication of IoT. It discusses applications in IoT and used techniques to solve the current security and trust issues in IoT such as vulnerable third-party interference and costly attack provision. For future work, need more research efforts at this stage to explore the primary study and new opportunities. Future scope of this work is tried to develop an algorithm to prevent trust issue in IoT data storing.

ACKNOWLEDGEMENT

We are using this opportunity to express gratitude to thank all the people who contributed in some way to the work described in this paper. Our deepest thanks to our project guide for giving timely inputs and giving us intellectual freedom of work. We express thanks to head of computer department and to the principal of AISSMS Institute of Information Technology, Pune-01 for extending his support.

REFERENCES

[1] “IoTOne: Integrated Platform for Heterogeneous IoT Devices”, Nathaniel Gyory, M. Chauha,2017

[2]1S.-H. Seo, M. Nabeel, X. Ding, and E. Bertino, “An efficient certificateless cryptography scheme without pairing,” in Proceedings of the third ACM conference on Data and application security and privacy. ACM, 2013, pp. 181–184.

[3] “Management Architecture for Heterogeneous IoT Devices in Home Network”, Cu PHAM, Yuto LIM, Yasuo TAN,2016

[4**]**W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge computing: Vision and challenges,” IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637–646, 2016.

[5]) L. Jiang, L. Da Xu, H. Cai, Z. Jiang, F. Bu, and B. Xu, “An IoT oriented data storage framework in cloud computing platform,” IEEE Transactions on Industrial Informatics, vol. 10, no. 2, pp. 1443–1451, 2014.

[6]TK. Christidis and M. Devetsikiotis, “Blockchains and smart contracts for the internet of things,” IEEE Access, vol. 4, pp. 2292–2303, 2016.

[7] R. Li, T. Song, N. Capurso, J. Yu, J. Couture, and X. Cheng, “IoT applications on secure smart shopping system,” IEEE Internet of Things Journal, vol. 4, no. 6, pp. 1945–1954, 2017.

[8] **“**Multi-Tenancy in Decentralised IoT, Sylvain” Cherrier, Zahra Movahedi, Yacine M. Ghamri-Doudane,2015

[9] “Achieving secure, scalable, and fine-grained data access control in cloud computing ”- S. Yu, C. Wang, K. Ren, and W. Lou, 2010

[10] I. Stojmenovic and S. Wen, “The fog computing paradigm: Scenarios and security issues,” in Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on. IEEE, 2014, pp. 1–8.

[11] **“**A Distributed Publisher-Driven Secure Data Sharing Scheme for Information-Centric IoT”, Ruidong Li, Hitoshi Asaeda, Jie Li,2017