A Comprehensive Study on Securing Data through Blockchain and AI

Vikram Tilekar[1] Kaustubh Jadhav[2] Yogesh Mehetre[3] Azhar Khan[4] Prof. Sarika Patil[5]
[1,2,3,4,5]Department of Computer Engineering
[1,2,3,4,5]Keystone School of Engineering, Pune, India

Abstract— Now a day in most of the advanced country doctors is using data seeking a method of immediate and effective treatment for their patients. Because of this the method the patient's medical condition is easily diagnosed and thereby a right treatment can be started immediately. In the paradigm of data seeking most of the times data vendors are involved in the process of data providing. Oftenly these data venders are acting as the mediators in between the data seeker and the data providers. However, it's become difficult to trust on the data vendors as they leak the crucial information about the providers or they lack the data itself. So to overcome this Blockchains are considered as the one of the best methods to provide the security for the data over the network. And secondly providing the data access control at the data provider's end is always a headache, so this paper studies most of the past work to understand their flaws to introduce a new methodology.

Keywords: Blockchain, Artificial Intelligence, Data provider, Data Vendor

I. INTRODUCTION

Ever since the inception of the human race, data has reigned supreme. Data has allowed humans as a species to dominate planet earth and has facilitated a plethora of technological advancements that have been reliant on a lot of data. Data is a highly essential component that can help facilitate growth and help the human race new heights. Most of the information technologies that have been designed rely on a large amount of data to keep a constant growth and provide services to the users. This is one of the most interesting propositions that have fueled the modern technology and information systems.

Data has many forms, going back to the very nascent stage of humans when the race evolved to start comprehending the surrounding environment. At this stage, humans started recognizing patterns and grasping other information about their environment to survive and thrive on this planet. This is the data that was inscribed on the cave walls that dwellers used at that time to stay safe from the various environmental elements such as rain, cold other predators, etc. these cave walls have paintings of various different animals and the hunting practices of that era. This information was vital for the survival of the race, therefore, it was taught to the young ones by drawing on the walls, this made sure the information was not lost but transferred to the younger generation to improve upon.

This was the first encounter of humans with data management and data preservation. This data kept improving generations upon generations to lead them to discover various metals and other materials that constantly improved over time to help shift the human race from a hunter-gatherer lifestyle to an agricultural one. This was far more sustainable than the previous lifestyle as this reduced the traveling and foraging around which was highly dependent on parameters out of the human race's control. Changing to an agriculture-based society was beneficial as this allowed them to settle down.

This transition also took a lot of data to be made possible, several generations studied and grew crops to understand the needs and other critical components that are needed for growing it in the best way. This was all passed down as word of mouth after the language had evolved enough to communicate complex messages. This is how data was proliferated until humans developed scripts and other forms of writing. Then it was slowly transitioned from folk songs and other word of mouth techniques to books that were concise and could be stored for long periods of time and passed down.

Books are another form of textual data this is very useful and is still being used widely all over the world. But earlier the books were handwritten and copied manually until the invention of the printing press by Gutenberg. This was a tedious process and the invention of the printing press ameliorated this effect and books could be printed faster and more efficiently with fewer mistakes as they were all identical. Therefore, books became quite a common form of data for the human race for a long time and still is a predominant format for textual information all across the globe.

It wasn't much longer until electronic forms of storage became more widespread and easily available post the invention of the personal computer. This led to an introduction of a more efficient and faster means of data collection and transfer which gets better every day. but modern information systems have greatly influenced by the introduction of the internet. This was a turning point in the data generation and management paradigm. The internet was created to allow for collaboration between various researchers which would allow them to share resources without actually physically being present at that location. This was revolutionary at that time and in its initial phases it was used to bolster research at United States and it was used by the military to communicate between themselves.

The internet was eventually opened for public use and what we know is history from that point forward. The internet created another set of services and other products that increased the convenience of the average human being by a large margin. There are still a lot of new features and other applications that are being added every day. This process generates a lot of data, with very high velocity as modern internet speeds with the addition of the fiber optic transfer medium has skyrocketed. Social Media also generates a large volume of data in a large variety every day. All of this data is highly useful and can provide a lot of insight through the use artificial intelligence.

Artificial intelligence, when combined with Data Mining, can help unleash a different perspective on the data generated and help a lot with the variety of businesses online as well as offline. It has the capability to predict various different scenarios with the help of analyzing previous data. but there is a problem with that aspect, as more and people are coming online and generating the data, are being highly

conscious of sharing their data with other data stakeholders citing security reasons. This is a problem as most of the data stakeholders gain access to a lot of data and if kept unchecked can lead to serious data leak scenarios.

This is a problematic occurrence as the Data stakeholders are responsible for collecting storing management of the data that is used for such Machine Learning applications. There is a need for a better security management system that helps ameliorate this effect as this kind of practice hampers data proliferation and sharing which would lead to scarcity of data and the artificial intelligence systems would starve and deplete. To achieve higher security the Blockchain paradigm is a suitable choice due to its inherent tamper-proof nature that would further increase the security of the system.

The Blockchain paradigm consists of two components, the block, and the chain. The Block is actual that is supposed to be protected and the chain is the special formation of the blocks that discourage tampering or modifications to the data once it's stored. The Blockchain achieves this through the ingenious use of hash keys, which are nothing but data about the data stored on the block. This hash key can be successfully used to validate the data that is being stored on the block. Therefore, the hash key of the current block is stored in the immediately next block and then the hash key from the next block is stored in the consecutive block, this forms a chain of blocks known as Blockchain.

Any modification done to the data stored on any of the blocks would lead to its hash key being modified, this would cause the next block to cease to identify the previous one modified block and would break the chain. The point of breakage would signify the point of modification and thus provide excellent security in such an application of data sharing with Data Stakeholders. This would allow a much more secure system for data sharing that would remove all inhibitions with people sharing their data online due to security reasons.

This paper dedicates section 2 for analysis of past work as literature survey and section 3 concludes the paper with feasible statement of the literature study.

II. LITERATURE SURVEY

This section of the literature survey eventually reveals some facts based on thoughtful analysis of many authors work as follows.

1) W.Dai [1] explains the conventional data trading ecosystem in a very simple way the seller will sell the data directly to the buyer by eliminating the middleware such as third parties. The seller will send the data to the trusted platform and set the proper selling price. The buyer will select a data product of interest and order, similar to other e-commerce transactions. The ability to handle and access large amounts of data easily is decisive, as such data can inform decision making in both public sectors and private sectors. The proposed paper analyzes the data trading market and sets the new rule for the trading market so they implemented secure data trading platform (SDTP) on Ethereum& Intel's Software Guard Extensions (SGX) and perform an in-depth analysis.
2) G.Magyar [2] defines the blockchain as consisting of blocks of records, where the blocks are linked together using an unchangeable key referencing mechanism. No central entity is responsible for governing and managing the data recorded in the blockchain. Blockchain is working in single network independent entities, which form together with a point-to-point network. The latest technology solves a crucial problem of accessing data without threatening personal privacy. The proposed paper introduces a new approach to an integrated health information model using the blockchain.
3) M.Kumar [3] elaborates the logging plays a crucial role in keeping the institution safe and secure. Network security devices provide the best security to prevent the network from the attackers but the attackers keep on trying to attack by using different strategies to attack so the Network security devices should be alert and updated accordingly. Log report gives detailed information to discover many loop-holes in networks because of misconfiguration, bugs in the Software, Hardware, and Firmware. Thus the proposed paper gives a safe log storage platform for a long time and better understands complex attack patterns.
4) Z.Guan [4] discusses the rapid growth in the field of computer and information technology creates various types of data striking speed by people or smart devices. According to survey 29 petabytes of user-generated data are stored, accessed on Facebook and 302 billion emails are sent every day in 2018 and also millions of sensors are generating real-time data. The proposed paper shows that the data is tending at a very high speed 100,000 data records within 1.51s, and statistic over a data set of 1,024 records within 1.99s. Ethereum blockchain with a smart contract provides an efficient output.
5) M.Chowdhury [5] estimates that in recent years there has been vast growth in the field of data evaluation of the organization such as hospitals, universities, banks are growing day by day due to web-enabled systems. Online social networks such as Facebook, Twitter, and LinkedIn, etc have transformed people's attitudes towards data sharing. The privacy risk is minimized by using consortium based blockchain network and by using valid data creators can be connected to blockchain. The proposed work decreases overall cost and upgrades the decision-making process.
6) M.Shen [6] concludes that the smart cities are including more and more advanced Internet-of-Things (IoT) configuration such as transportation, manufactory, energy transmission, and agriculture resulting in a huge amount of data gathered from various IoT devices deployed in many city sectors. Blockchain is mainly a distributed filing system developed to allow the sharing of forgery-proof records among different parties. Blockchain methods enable secure and genuine IoT data sharing. IoT data supplier can encode the data instances by its own private key, and then keep track of the encrypted data on blockchain via particular formatted transactions.
7) U.Uchibeke [7] explains certain steps should be taken to provide secure data management by solving problems such as data security and access control issues. There has

been rapid growth in the field of information technology heterogeneous data is generated ranging from agriculture, business, finance/banking education, medicine, and healthcare.It is very necessary to provide security and privacy for such huge data. The technology called blockchain presents a solution to the challenges linked with standard and centralized access control and ensures data clarity and imputable for secure data sharing, examine user and data self-sovereignty for the owner.

8) S. Ramamoorthy [8] discusses data security and Storage Management becomes the most favorable work for any IT operation in the coming years. In the proposed paper author impose the secure cloud data access and modification framework using blockchain approaches. The same method also can be imposed on the IoT platforms by replacing the cloud nodes with IoT sensors. The user applies for data access estimates before getting the service entry from cloud data center.In proposed methodology hybrid approach to merge the BlockChain(BC) Technology with cloud computing to ensure data security among the cloud users.
9) W. Liang [9] establishing the industrial blockchain network system many types of research had tried for security issues of data transmission in industrial IoT. A time-stamp and a block mark attached with block and data block orderly chain composed in Blockchain technology. Various threats of malicious attacks should be improved through blockchain data transmission. If there exists a mugger in the blockchain network, he can build multiple identities attacker may leak the important data or may cheat the security authentication of the blockchain network. The secure data transmission technique can be applied to industrial IoT, such as power, energy.
10) P.Urien [10] estimates the rapid growth in the industrial Internet of Things (IoT), productivity efficiency has been widely improved. The previous methodology of blockchain data transmission methodology in industrial Internet of Things (IoT) have low security, high management cost of the trading center, and big difficulty in supervision. First deals with deployment of the CCSC application in low-cost low power object powered by an open hardware (Arduino) platform, and integrates sensor data in ethereum transactions. Thus plan to develop BIoT platforms in order to evaluate the benefits of blockchain technology for IoT services.
11) C.Cai [11] introduces blockchain-based hybrid cloud data management model will provide secure data and service access. This model can be used in both public and Private cloud configurations to ensure data privacy. There are many security level management features established using this cloudBC structure which is Community cloud data management Ensure data privacy and security,Open Decentralized Ledgers, Monitoring and Tracking the service history, Group approval for data access and Modification, Updates reflects in Global Table, Digital dispute ensures data authenticity, Native platform management. Thus proposed methodology assures the development of secure and trustworthy blockchain applications and systems.
12) S.Sharma [12] explains how Blockchain is evolving and how blockchain is one of the mostpromising and resourceful technologies of cloud base security. In a distributed database system blockchain is meant to store, read and validate transactions. Due to the use of cryptographic data structures Tampering of Blockchain is enormously thought-provoking. In proposed methodology authors compare the blockchain with the various platforms on which blockchain can be implemented. Blockchain applications are used for building a secure infrastructure of cloud computing.
13) X. Zheng [13] discusses high-speed development of mobile computing, wearable technology, and wireless sensing, mankind has been using different types of mobile and wearable devices, such as smartphone, smartwatch, smart band, and smart glasses, etc due to this large amount of data has been produced. Properly splitting personal health data will help all related stakeholders including the patients, researchers, companies and even the whole public healthcare system. Thus it is necessary to bring fence for the data sharing and puts data security and authority providers are attractive bags for cyber-attacks.
14) C. Liu[14] narrates fast growth in field of smart portable mobile terminals MTs, e.g., smartphones, UAVs, which are fitted out with rich set of sensors e.g., camera, gyroscope, and GPS, has enables a new type of data collection method for industrial IoT (IIoT), namely Mobile Crowdsensing (MCS).A classic MCS system contains a cloud-based server and a collection of MTs.Ethereum blockchain and deep reinforcement learning (DRL) are proposed by author for efficient data collection and secure sharing.It can provide a web-enabled higher security levels and stronger opposition to attack than a traditional database-based data sharing scheme for different types of attacks.
15) M.Singh [15] estimates that the Internet of Things (IoT) is now at its starting level but very soon, it is going to impact almost every day-to-day items we use. Homes and businesses by transforming objects that are used to be offline into online systems.can are attacked as it is growing day by day. To secure IoT systems in the time blockchain has emerged as a possible solution.In proposed paper infrastructure of IoT which is form on Blockchain network and at end model is kept for the security of internet of things using blockchain.

III. CONCLUSION

This paper studied the most of the past existed techniques in data vending process which involves the mediators along with the access control mechanism for the stored data. The deep analysis reveals some facts like many methodologies are suffering from the problem of automatization of the access control mechanism. However, this paper decided to use the furnished algorithm of machine learning for the automatic data access control mechanism using linear regression and Hidden Markov model which are catalyzed by the Fuzzy c means clustering technique. By using this technique this paper can provide the automatic access control mechanism

for the asked data along with the high protection of the data using blockchain techniques.

REFERENCES

[1] Weiqi Dai, Chunkai Dai, Kim-Kwang Raymond Choo, Changze Cui, Deiqing Zou, and Hai Jin," SDTE: A Secure Blockchain-based Data Trading Ecosystem" IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information 2019.

[2] Gábor Magyar," Blockchain: solving the privacy and research availability tradeoff for EHR data" IEEE 30th Jubilee Neumann Colloquium • November 24-25, • Budapest, Hungary 2017

[3] Manish Kumar, Ashish Kumar Singh, Dr. T V Suresh Kumar," Secure Log Storage Using Blockchain and Cloud Infrastructure" IISC, Bengaluru July 10-12, 2018.

[4] Zhangshuang Guan, Xiaobei Shao, and Zhiguo Wan," Secure, Fair and Efficient Data Trading without Third Party Using Blockchain" IEEE Confs on Internet of Things, Green Computing and Communications, Cyber, Physical and Social Computing, Smart Data, Blockchain, Computer and Information Technology, Congress on Cybermatics 2018

[5] Mohammad JabedMorshed Chowdhury∗, Alan Colman∗, Muhammad AshadKabir, Jun Han∗ and Paul Sarda∗," Blockchain as a Notarization Service for Data Sharing with Personal Data Store" 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering 2018

[6] Meng Shen, Xiangyun Tang, Liehuang Zhu, Xiaojiang Du, and Mohsen Guizani, "Privacy-Preserving Support Vector Machine Training over Blockchain-Based Encrypted IoT Data in Smart Cities" This article has been accepted for publication in a future issue of this journal but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2901840, IEEE Internet of Things Journal 2019.

[7] UchiUgobameUchibeke, Sara HosseinzadehKassani, Kevin A. Schneider, Ralph Deters," Blockchain access control Ecosystem for Big Data security " IEEE Confs on Internet of Things, Green Computing and Communications, Cyber, Physical and Social Computing, Smart Data, Blockchain, Computer and Information Technology, Congress on Cybernetics 2018.

[8] S. Ramamoorthy, B.Baranidharan," CloudBC-A Secure Cloud Data access Management system" 978-1-5386-9371-1/19/.00c IEEE 2017.

[9] W. Liang, M. Tang, J. Long, X. Peng, J. Xu, and K. C. Li, "A Secure Fabric Blockchain-based Data Transmission Technique for Industrial Internet-of-Things" 1551-3203 (c) IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information 2018.

[10] Pascal Urien," Towards Secure Elements For Trusted Transactions in Blockchain and Blockchain IoT (BIoT) Platforms. " Certicom Research, January 27, Version 2.0 2010

[11] ChengjunCai, HuayiDuan, and Cong Wang," Tutorial: Building Secure and Trustworthy Blockchain Applications" 2018 IEEE Secure Development Conference 2018.

[12] Shweta Gaur Sharma Dr. Laxmi Ahuja," Building Secure Infrastructure for Cloud Computing using Blockchain " Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS) IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3 2018.

[13] Xiaochen Zheng, Raghava Rao Mukkamala, Ravi Vatrapu, JoaqunOrdieres-Mer´e," Blockchain-based Personal Health Data Sharing System Using Cloud Storage" IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom) 2018.

[14] Chi Harold Liu, Senior Member, IEEE, Qiuxia Lin, Shilin Wen," Blockchain-enabled Data Collection and Sharing for Industrial IoT with Deep Reinforcement Learning" Citation information: DOI 10.1109/JIOT.2019.2901840 2015.

[15] Madhusudan Singh, Abhiraj Singh, Shiho Kim" Blockchain: A Game Changer for Securing IoT Data" IEEE SECURITY AND PRIVACY, http://www.jbonneau.com/doc/BMCNKF15-IEEESP bitcoin.pdf. 2015

