International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Systematic Review and Future Research Directions on Privacy Preservation Techniques Using Blockchain

Sandip S. Chauhan[1*], Dr. Dinesh Vaghela[2*]

Submitted: 07/02/2024 **Revised**: 13/03/2024 **Accepted**: 20/03/2024

Abstract: Big data is an extremely large data set with a diverse and complicated production of organized and unstructured data. The management of storing, analyzing, and finding meaningful outcomes from processing data is usually aided by the characteristics of big data Through its decentralized system architecture, block chain technology ensures data protection and privacy. Due to the block chain consensus process and encryption, data recorded in the block chain is tempered proof and cannot be readily changed, and a massive amount of computing power will be necessary. Through its connected chain, block chain's transaction ledger offers data auditing. Data auditing is possible thanks to block chain's transaction ledger. All valid data transfers are recorded and put into the block chain ledger, which assures data quality through a variety of verification processes. Block chain is also used in the big data mining process, where data is collected from many networks and organizations with the purpose of demonstrating data with an exponential growth in risk factors. A block chain platform can also assist data scientists in monetizing their labour by allowing them to trade analytic results stored on the network.

Keywords*: ChainPPDM, Block-chain, Privacy Preservation, Big Data*

Introduction

The big-data strategy has grown in significance across a range of corporate processes and sales evaluation strategies. There are a lot of privacy issues, however, which slow down the development of their analytic tools. Reports from big-data have yielded several useful findings. But most of them aren't open to the public and are utilised only inside companies due to privacy concerns. Prior to moving forward with big-data analysis, it is critical to address privacy concerns[1][2]. Data security refers to the practice of keeping data safe against unauthorised access, use, or disclosure by maintaining data's availability, integrity, and confidentiality. A data security strategy should consist of three parts: gathering necessary data, storing it securely, and erasing unnecessary data. Appropriate data usage is the definition of data privacy. Businesses and vendors should only use customer information for the specified reasons, and not reveal any personal details, when they get data or information. Consequently, the ideal approach is one that permits data exchange in a protected setting, maintaining the need for individual privacy while making the data realistically applicable for analysis. The PII, which includes details like name, address, age, etc., is grouped or hidden. The problem is that these categories aren't always "undistinguishable" all the time. Reason being, when users establish connections, server-side authentication of users is always required, and users' contact lists are kept on the server side. Therefore, the service provider has full visibility into the social network's architecture. With the rise of new distributed settings like clouds and social networks[3], there is a growing need for privacy-preserving big data analytics and management solutions. Several ongoing projects are attempting to address this issue by developing methods and algorithms that can be applied to large, distributed big data repositories[4]. A number of methods exist for protecting personal information, but three of the most important ones are:

- data anonymization.

- privacy-preserving data publishing.

- privacy that differs. To make data anonymous, we must remove or obscure key stream properties that might compromise the confidentiality of the intended streams. Kanonymization, l-diversity, and t-closeness are well-known methods for the static situation.

Gathering all of the data into a single location and storing it there before processing it is the simplest way to do machine learning. Nonetheless, large-scale data is often created by several sources and kept in a dispersed way across different locations for various purposes. Significant communication overhead and serious data privacy and security problems would result from transmitting such large volumes of data from several data holders to a single server. These pervasive datasets may be retrieved from a wide range of domains and applications, including

[1]SandipKumar S. Chauhan, PhD research scholar, IT/Computer engineering at Gujarat Technology University, E-mail Id: sandymba2006@gmail.com

[2]Dr. DineshKumar B. Vaghela, Assistant Professor, Information Technology

Department, Shantilal Shah engineering College, Bhavnagar - 364060, Gujarat, India, E-mail Id: dineshvaghela28@gmail.com

**Corresponding Author E-mail Id: sandymba2006@gmail.com*

bioinformatics, medical analysis, clickstream analysis, retail, market basket analysis, and clickstream mining and analytics. However, there are privacy issues about the study of big data sets that include sensitive personal information[5]. The primary concerns with data analytics and management are to data privacy and security. Data transformations may lessen the data's usefulness, making data mining knowledge extraction less precise or even impossible. The Privacy-Preserving Data Mining paradigm is this. While data mining on the modified data may still be done effectively, PPDM approaches are intended to maximise data value while guaranteeing a certain amount of privacy. Under the condition of data privacy preservation, PPDM performs data mining activities[6]. Research in the fields of data mining and knowledge discovery from databases involves the automated finding of relationships from massive volumes of data that were previously unknown. Recent developments in data gathering, data dissemination, and associated technologies prompted a fresh look at the state of the art in data mining techniques[7].

We are all too aware that the flood of data also brings up new privacy problems, even as big data generates tremendous benefits for technological advancement and economic development. In order to strike a balance between the advantages of big data and the protection of individual privacy, it is necessary to thoroughly identify the privacy needs in big data architecture[8]. Needs for data privacy while collecting huge data: The widespread nature of big data collecting raises concerns about eavesdropping and accidental data leaks. Consequently, we need to use physical protection measures and information security procedures to guarantee data privacy before safely storing it if the data is personal and sensitive[9]. Concerns about personal data privacy in large data storage; getting into someone's system when they're storing huge data is much worse than eavesdropping on their data while they're collecting it. Once it's working, it may reveal additional personal information about individuals. That is why there must be physical and cyber safeguards in place to protect the privacy of all data[10][11]. Needs for data privacy while processing large datasets: Big data processing is an essential part of big data analytics as it finds new ways to use data to advance technology and the economy. With big data processing efficiency becoming a key performance indicator, meeting privacy standards in big data processing is becoming more difficult. We should never compromise privacy for efficiency, and we should always do our best to safeguard individuals' privacy while also maximising their productivity[12]. Furthermore, since big data processing encompasses data from several organisations, sharing big data is crucial. However, guaranteeing privacy while sharing large data becomes a formidable challenge in the field of big data processing. Consequently, it is preferable to develop algorithms for large data processing and sharing that are both efficient and protect users' privacy. The main method for finding new information is data mining. Data mining could lead to the extraction of sensitive information, but it also allows for the efficient discovery of important, non-obvious information from massive amounts of data[13][14]. Methods for randomization, anonymization, and distributed alteration of original data to maintain privacy are categorised as follows. Also included are analytics and comparisons of privacy-preserving clustering and association rule mining. While this method preserves some characteristics of the original data while distorting it, it has the drawback of making it difficult to query the data after a perturbation and a high level of protection for random numbers[15].

While data encryption based solutions may accomplish privacy preservation and guarantee data confidentiality and integrity, they aren't well-suited for protecting large datasets because of the complexity associated with encryption and decryption. One drawback of data anonymity-based solutions is that they cannot protect against background knowledge attacks; these methods aim to preserve privacy by hiding or deleting the identity of the data owner and sensitive properties[16]. To strike a balance between privacy preservation and data usefulness, the ideal approach would maximise the preservation of sensitive data while minimising any impact on the other. Among data anonymization methods, the k-anonymity model is the most well-known. There must be at least K records in each equivalence class, and the records within each class may share values with the privacy information's characteristics. The data collection must be partitioned into numerous equivalence classes. Links can only be attacked with a probability of 1/k or below. Researchers have spent years refining the k-anonymity approach, adding features like (a, k)-anonymity and (p, a)-sensitive k-anonymity, which were suggested by researchers from other countries. A k-anonymity technique that has been suggested by academics in China that can withstand various restrictions. Nevertheless, the k-anonymity paradigm is not without its flaws. Any time a set of records share a sensitive attribute value, it makes it easy for an attacker to access private data. As a solution to the k-anonymity mode issue, the l-diversity model was developed in 2006 . It addresses the issues with the k-anonymity paradigm. However, similarity assaults are too much for the model to handle. This means that the percentage of the value of a sensitive property is too high. An attacker has a good chance of obtaining personal information in this scenario. Scholar subsequently put out the t-closeness model. It stipulates that there can't be a greater disparity than t between the distribution of attributes throughout the whole data table and the distribution of sensitive attributes in equivalence classes.

In addition to fixing the issues with the l-diversity model, this may also stop similarity assaults[18].

There has been a meteoric rise in data collecting over the last decade from households, companies, and even government entities. Though data sharing helps with mining and analysis, it might put people's privacy at danger if shared too widely. In order to publish helpful information while protecting data privacy, privacy reserving data publication [19] offers methods that leverage several privacy models. Privacy models like k-anonymity l-diversity [14] and t-closeness [15] do not provide complete protection against assaults if the adversary has previous information about persons in the data, in contrast to differential privacy. assaults like this may take several forms, including table-linkage, attribute-linkage, and probabilistic assaults. However, differential privacy protects against these kinds of assaults without revealing an individual's involvement in the released data and without assuming that an adversary has any prior information. Our aim in this project is to secure differential privacy for publicly available data. The proliferation of mobile computing and improvements in location-acquisition methods have resulted in a deluge of data pertaining to the motion of items in motion. item movement data may comprise either the precise places an item visited or the exact times at which those locations were visited. The movement of humans, cars, animals, or even natural disasters may provide the basis for movement data.[19] Data on motion may also be categorised into two main types: sequential and trajectory. Each sequence in sequential data records the places visited by a moving entity in the specified order. Contrarily, a series of doublets (l; t) denoting the location l that was visited at timestamp t is used to indicate an object's movement in trajectory data. Our focus here is on the Inter Planetary File System (IPFS) and how it relates to the challenge of protecting the privacy of large datasets.

Businesses, organisations, and governments are racing to create electronic systems for managing people information because to the lightning-fast progress in information technology. Hundreds of thousands—if not millions—of people's data is collected and stored by the system. Due to the massive amount of data needed for big data growth, any interference or leakage might have catastrophic consequences. A relatively recent area of study within data mining is privacy preservation in data mining, or PPDM. Its end goal is to make it possible to safely mine massive volumes of data for insights without exposing or inferring any personally identifiable information. Three methods exist in data mining that aim to protect individuals' anonymity: Anonymization, Differential Privacy, Data and Consent

Objective of work

1. Proposes a block chain-based solution for preserving the privacy of large data for data mining operations.
2. Examined the block chain's flaws and offered a better solution.
3. To transport enormous amounts of data in Big Data, we developed an IPFS network. As a result, before transferring data, encrypt large data and store it on the IPFS network.

Privacy and Security

Warning and Authorization Notice and consent is the gold standard for protecting user privacy on the web. There is a notification about privacy risks that appears every time someone uses a new service or application. Before utilising the service, the customer must agree to the notice. An person may protect his right to privacy using this strategy. The onus for protecting one's privacy is therefore placed on the person. This approach presents a plethora of difficulties when used with massive data [6]. The majority of the time, when notice and permission are granted, the applications of large data are either unanticipated or unknown. This necessitates that the warning be updated whenever big data is used for an alternative objective. Additionally, customers are burdened with the need to agree to notices due to the quick processing and collection of large data. Using third parties who provide a variety of privacy profiles is one way to change big data's notice and consent processes.

Despite the proliferation of privacy-protecting methods, data anonymization remains the foundation for the vast majority of them[7].

Randomization happens when random variables are included into a dataset, often via the use of probability distribution. Polls, sentiment analysis, etc., all make use of randomization. It is not necessary to be aware of other entries in the data for randomization to work. It finds use while preparing data or when collecting data. Randomization eliminates the need to optimise for anonymization. The time difficulty and data usefulness of implementing randomization on huge datasets make it impossible, as shown in the experiment detailed below[8][9]. We ran a Map Reduce task on 10,000 entries that were imported into the Hadoop Distributed File System from an employee database. Using wage and age as variables, we have conducted experiments to categorise the staff. We created a database of 15,000 records by randomly adding 5,000 records to it in order to implement randomization[11].

Data distribution technique

This strategy involves dispersing data across several places. Both horizontal and vertical distributions of data exist. When data is scattered over many places with shared characteristics, we say that the distribution is horizontal.

Horizontal data distribution is the best option when you need to do aggregate operations or functions on data without actually distributing the data. A store, for instance, may track sales at each of its locations using analytics software. One potential downside is that sharing personal information with a third-party analyst for data analysis might put the data owner at danger of having it exposed. There is no assurance of privacy when using clustering and classification algorithms on distributed data. If data is spread out over several sites controlled by various companies, the results of aggregate functions might help one party find data held by another. All participating sites are expected to be honest in these types of situations[12][13]. When many entities are in charge of various storage areas for people' personal data, a phenomenon known as vertical distribution occurs, as seen in Figure 1.For instance, knowing a suspect's health, financial, and vocational details might help law enforcement during a criminal investigation. Some or all of this information may be unavailable at the moment. This kind of data structure is called a vertical distribution, and it keeps a tiny portion of a person's qualities in each area. Since data consolidation from all these sites is necessary for analytics, there is a chance of privacy infringement.

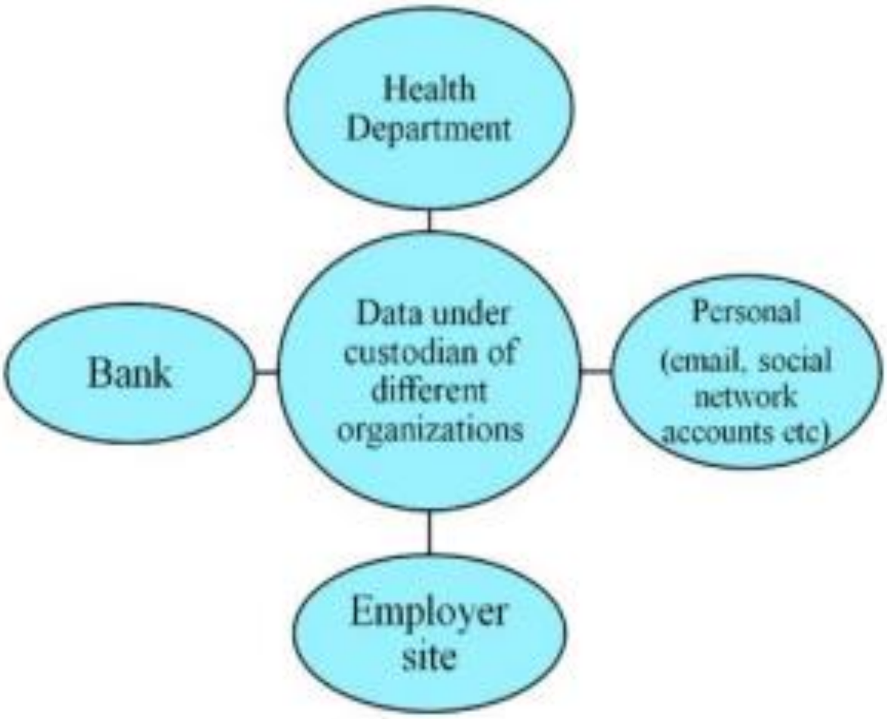
Health Department
Bank
Data under custodian of different organizations
Personal (email, social network accounts etc)
Employer site

Fig 1. Vertical distribution of personal identifying data

A big obstacle to sharing datasets for analytics is ensuring privacy when data is vertically distributed, with attributes located in different places and different people accountable for them. As an example, the investigating officer may feel the need to get specific data from the accused's employer, healthcare provider, or financial institution in order to have a deeper grasp of the accused's character. If the investigating officer finds out any personal information about the accused throughout this process, the accused may experience humiliation or abuse. Anonymization is not applicable when analytics do not need full records. While data dispersion isn't a foolproof means of privacy protection, it is quite comparable to cryptographic techniques[15].

Systematic Review

Author(s)	Year	Title	Techniques	Summary	Pros	Cons
Ali Mohammadi Ruzbahani	2024	AI-Protected Blockchain-based IoT environments: Harnessing the Future of Network Security and Privacy	Homomorphic Encryption, Access Control	Proposes a blockchain-based framework for secure data sharing in IoT using homomorphic encryption and fine-grained access control policies.	High security for IoT data, fine-grained access control	High computational overhead, complex implementation
Benyu L et al.	2023	A Blockchain-Based Privacy-Preserving Data Sharing Scheme with Security-Enhanced Access Control	Attribute-Based Encryption, Smart Contracts	Introduces a data sharing scheme that combines attribute-based encryption and smart contracts to ensure secure and private data transactions on blockchain.	Flexible access control, enhanced data security	Potential latency in smart contract execution, key management complexity
Lin Chen et al	2022	Overview of Medical Data Privacy Protection based on Blockchain Technology	Differential Privacy, Smart Contracts	Explores the use of differential privacy and smart contracts to protect sensitive healthcare data stored and shared on blockchain.	Protects sensitive healthcare data, regulatory compliance	Differential privacy may reduce data utility, smart contract vulnerabilities
Haiping Huang et.al	2021	A blockchain-based scheme for privacy-preserving and secure sharing of medical data	Zero-Knowledge Proofs, Encryption	Proposes a privacy-preserving framework for supply chain networks using zero-knowledge proofs and encryption to ensure data confidentiality and integrity.	Ensures data integrity and confidentiality in supply chains	Complex cryptographic protocols, performance overhead

Faiza Loukil et al.	2020	Privacy-Preserving IoT Data Aggregation Based on Blockchain and Homomorphic Encryption	Homomorphic Encryption, Secure Aggregation	Presents a secure data aggregation scheme using homomorphic encryption for IoT systems to enhance privacy and data security in blockchain-based environments.	Enhances data privacy in IoT systems, supports secure data aggregation	High computational costs, scalability issues in large-scale IoT networks
Xuhui Chen et.al	2019	Asynchronous Blockchain-based Privacy-preserving Training Framework for Disease Diagnosis	Homomorphic Encryption, Ring Signatures	Introduces a decentralized framework for healthcare systems that uses homomorphic encryption and ring signatures to ensure privacy and security of patient data.	Enhances patient data privacy, decentralized control	Homomorphic encryption is computationally intensive, ring signature management complexity
Shunrong Jiang et.al	2019	Secure and Privacy-preserving Energy Trading Scheme based on Blockchain	Differential Privacy, Secure Multi-Party Computation	Proposes a blockchain-based data trading scheme that employs differential privacy and secure multi-party computation to protect data privacy during transactions.	Preserves data privacy during trading, ensures data security	Differential privacy may reduce data accuracy, complex computation for multi-party protocols
Saide Zhuet al	2019	Hybrid Blockchain Design for Privacy Preserving Crowdsourcing Platform	Zero-Knowledge Proofs, Identity-Based Encryption	Introduces a privacy-preserving scheme for vehicle data sharing in intelligent transportation systems using zero-knowledge proofs and identity-based encryption.	Ensures data privacy in ITS, enhances data security	Zero-knowledge proofs are computationally expensive, identity management complexity

Cryptographic techniques

Before making the data available for analytics, the data holder has the option to encrypt it. However, traditional encryption methods are very difficult to implement on a broad scale and should only be used when data is being collected. In cases when it is necessary to do aggregate calculations on the data without divulging the inputs, differential privacy approaches have previously been implemented. To get some aggregate information from both x and y without actually sharing x and y, a function F (x, y) is constructed, for instance, if x and y are two data items. For situations like vertical distribution, when x and y are owned by separate parties, this might be useful[16]. However, differential privacy won't work if all of the data is stored in one central place by only one company. Similarly, safe multiparty computing has been tried, but it falls short when it comes to protecting users' privacy. Encryption used in data analytics reduces data usefulness. As a result, encryption is not only a pain to set up, but it also makes data less useful[17].

Multidimensional sensitivity-based anonymization (MDSBA)

Traditional approaches to anonymization for structured data records with good representation include bottom-up and top-down generalisation. But scalability and data loss become serious problems when the same methods are used to massive data collections. An enhanced variant of anonymization, multidimensional sensitivity-based anonymization outperformed traditional methods. An enhanced anonymization approach, multidimensional sensitivity-based anonymization allows for the use on huge data sets with less information loss and preset quasiidentifiers. This method makes advantage of the Apache Map Reduce technology to manage massive datasets[18].

This approach uses the scripting language Apache Pig for implementation, which reduces development work. Since every Apache Pig script must be transformed into a MapReduce job in the end, the coding efficiency of Apache Pig is lower than that of a MapReduce job. While MDSBA works well with huge datasets when the data is not in motion, it is not applicable to streaming data[19].

Proposed ChainPPDM Method

An innovative new on-chain and off-chain approach for protecting large amounts of data privately is ChainPPDM. Increasing storage capacity is an issue in the blockchain business. Blockchain is completely different from traditional database systems; it is not similar in any manner. Big data analysis, on the other hand, is, in our opinion, quicker than competing database systems. Data mining and pattern recognition are areas where bid data is not overlooked. Personal information security is paramount[20].

ChainPPDM, whereas HDFS is used for less sensitive data. The ChainPPDM system makes use of IPFS to transfer data between the data collector database and the data miner warehouse. The data collector encrypts sensitive information before providing it to the data miner. A private data hash is now essential for blockchain-stored private information as well as non-sensitive data kept in HDFS massive data warehouses. The data collector and data miner construct a smart contract to use private data. Fig. 2 shows the ChainPPDM data mining process.

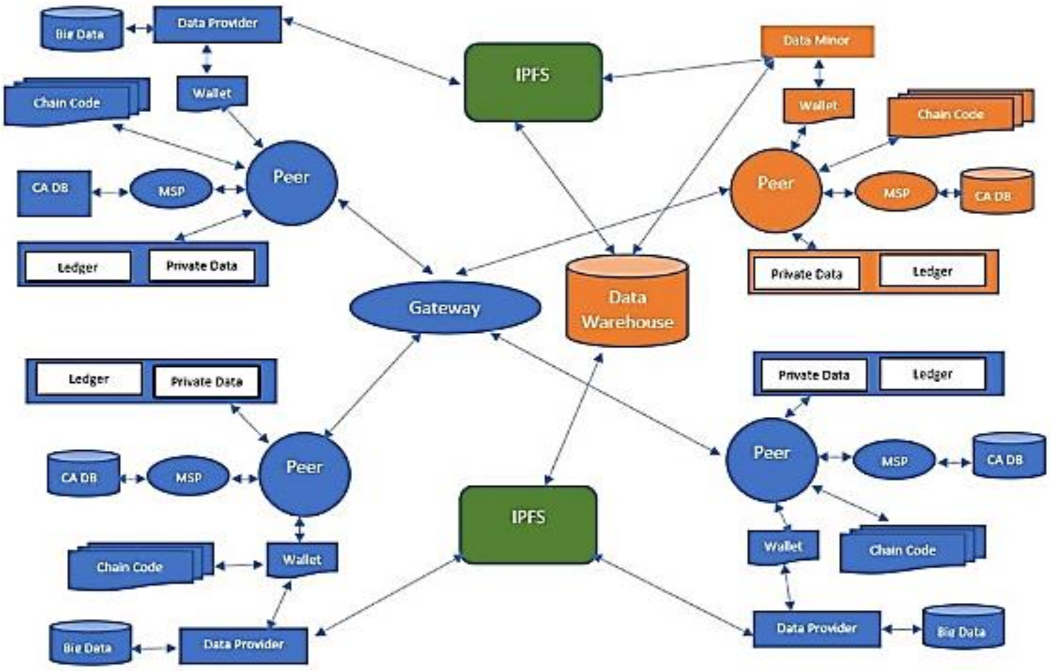

Fig 2. ChainPPDM – Proposed Method

One of the four core technologies used in the implementation is ChainPPDM. Licence for Hyperledger Fabric Another option is the blockchain-based HDFS system. It will be used by IPFS (Internet File System), a cryptographic hashing mechanism. Comparatively, SHA-256 produces smaller outputs than SHA-512. Therefore, less bandwidth is often required for data transmission and storage. In most cases, SHA-256 will reduce the amount of memory and processing power required. In this comparison, SHA-256 outperforms SHA-512 by 31% while processing shorter inputs. Even with longer inputs, SHA-512 is just 2.9% faster while hashing. Satoshi Nakamoto likely opted for SHA-256 instead of SHA-512 because the Bitcoin network hashes many smaller inputs. These inputs include 33-byte public keys and 80-byte block headers. Furthermore, SHA-512 outperforms SHA-256 on 64-bit CPUs compared to 32-bit ones. When the Bitcoin network was launched in 2009, 32-bit processors were far more common than 64-bit ones. The validation of transactions is further slowed down by the rise in computing cost caused by the need for larger SHA-512 blocks and more blockchain consensus rounds. Five tuples make up the ChainPPDM mechanism.

$chainPPDM = (D, O, DM, IN, mn)$

were

(1) $D = \{d1, d2, d3 \dots dn\}$ are data sets for data mining operation.

(2) $O = \{o1, o2, o3 \dots om\}$ is set of organization which provide data set for data mining operation

(3) $DM = \{dm1, dm2, dm3 \dots dm\ \}$ is set of data mining organization which give output of pattern or for decision making system.

(4) $IN = \{in1, in2, in3 \dots in\ \}$ is set of IPFS node which used by data collector and data miner for transfer data from collector to miner off-chain server (

(5) $mn \subseteq M$ is node of blockchain. Every organization is had at least one node which used for private data transfer and received output from datamining.

$om = \{mn\ \}\ where\ 0 < i \leq k$; $om\ is$ set of organization

Where, k is total number of node(s) of organization. Hyperledger Fabric is store private data in collection which allow organization to endorse and query on private data. Hyperledger fabric have 3000 transaction throughput capacity.

Challenges and Future Directions

While blockchain presents significant opportunities for enhancing privacy preservation, several challenges remain. These challenges must be addressed to fully realize the potential of blockchain-based privacy techniques:

1. **Regulatory Compliance**: The decentralized and immutable nature of blockchain can conflict with existing data protection regulations that require the ability to alter or delete data. Future research should focus on developing compliant blockchain solutions, such as permissioned blockchains that allow controlled access and modification of data.
2. **Usability**: Blockchain-based privacy techniques must be user-friendly to ensure adoption. Simplifying the user experience and providing clear guidelines for managing privacy settings can enhance the usability of blockchain solutions.
3. **Security**: While blockchain offers enhanced security, it is not immune to attacks. Ensuring the robustness of cryptographic techniques and addressing vulnerabilities in smart contracts and consensus mechanisms are critical for maintaining privacy and security.

Future research directions should focus on:

- Ensuring compliance with data protection regulations while leveraging blockchain's benefits.
- Improving the usability of blockchain-based privacy techniques to encourage broader adoption.
- Strengthening the security of blockchain systems to protect against emerging threats.

By addressing these challenges and pursuing these research directions, the potential of blockchain technology to revolutionize privacy preservation can be fully realized, fostering a more secure and trustworthy digital landscape.

Conclusion

Blockchain technology presents an exciting opportunity to improve privacy protection in many different areas. It solves numerous problems with existing privacy protection systems because to its decentralised, transparent, and irreversible nature. To be sure that blockchain-based solutions are scalable, compatible with other systems, and in line with regulatory requirements, more research and development are necessary to address the present obstacles and reach their full potential. The inventive use of blockchain technology will propel digital transformation ahead while protecting individual and organisational privacy; this is the future of privacy preservation.

References

[1] Ali Mohammadi Ruzbahani, "AI-Protected Blockchain-based IoT environments: Harnessing the Future of Network Security and Privacy," in 2024 Cryptography and Security https://arxiv.org/abs/2405.

[2] B. Li, J. Yang, Y. Wang, X. Huang, J. Ren and L. Wang, "A Blockchain-Based Privacy-Preserving Data Sharing Scheme with Security-Enhanced Access Control," *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, Rio de Janeiro, Brazil, 2023, pp. 825-830, doi: 10.1109/CSCWD57460.2023.10152751.

[3] L. Chen, Q. Yu, W. Liang, J. Cai, H. Zhu and S. Xie, "Overview of Medical Data Privacy Protection based on Blockchain Technology," *2022 IEEE 7th International Conference on Smart Cloud (SmartCloud)*, Shanghai, China, 2022, pp. 200-205, doi: 10.1109/SmartCloud55982.2022.00039.

[4] Haiping Huang, Peng Zhu, Fu Xiao, Xiang Sun, Qinglong Huang, A blockchain-based scheme for privacy-preserving and secure sharing of medical data,Computers & Security,Volume 99,2020,102010,

[5] Loukil F, Ghedira-Guegan C, Boukadi K, Benharkat A-N. Privacy-Preserving IoT Data Aggregation Based on Blockchain and Homomorphic Encryption. *Sensors*. 2021; 21(7):2452. https://doi.org/10.3390/s21072452

[6] X. Chen, X. Wang and K. Yang, "Asynchronous Blockchain-based Privacy-preserving Training Framework for Disease Diagnosis," *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, 2019, pp. 5469-5473, doi: 10.1109/BigData47090.2019.9006173.

[7] S. Jiang, X. Zhang, J. Li, H. Yue and Y. Zhou, "Secure and Privacy-preserving Energy Trading Scheme based on Blockchain," *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Taipei, Taiwan, 2020, pp. 1-6, doi: 10.1109/GLOBECOM42002.2020.9348246.

[8] S. Zhu, H. Hu, Y. Li and W. Li, "Hybrid Blockchain Design for Privacy Preserving Crowdsourcing Platform," *2019 IEEE International Conference on Blockchain (Blockchain)*, Atlanta, GA, USA, 2019, pp. 26-33, doi: 10.1109/Blockchain.2019.00013.

[9] S. Rahmadika, P. V. Astillo, G. Choudhary, D. G. Duguma, V. Sharma and I. You, "Blockchain-Based Privacy Preservation Scheme for Misbehavior Detection in Lightweight IoMT Devices," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 710-721, Feb. 2023, doi: 10.1109/JBHI.2022.3187037

[10] Q. N. Tran, B. P. Turnbull, H. -T. Wu, A. J. S. de Silva, K. Kormusheva and J. Hu, "A Survey on Privacy-Preserving Blockchain Systems (PPBS) and a Novel PPBS-Based Framework for Smart Agriculture," in *IEEE Open Journal of the Computer Society*, vol. 2, pp. 72-84, 2021, doi: 10.1109/OJCS.2021.3053032.

[11] D. Naidu, B. Wanjari, R. Bhojwani, S. Suchak, R. Baser and N. K. Ray, "Efficient Smart contract for Privacy Preserving Authentication in Blockchain using Zero Knowledge Proof," *2023 OITS International Conference on Information Technology (OCIT)*, Raipur, India, 2023, pp. 969-974, doi: 10.1109/OCIT59427.2023.10430710.

[12] N. Khan, A. Lahmadi, Z. Kräussl and R. State, "Management Plane for Differential Privacy Preservation Through Smart Contracts," *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)*, Antalya, Turkey, 2020, pp. 1-8, doi: 10.1109/AICCSA50499.2020.9316507.

[13] P. Anbumani and R. Dhanapal, "Review on Privacy Preservation Methods in Data Mining Based on Fuzzy Based Techniques," *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Greater Noida, India, 2020, pp. 689-694, doi: 10.1109/ICACCCN51052.2020.9362801.

[14] Y. Wang, A. Zhang, P. Zhang and H. Wang, "Cloud-Assisted EHR Sharing With Security and Privacy Preservation via Consortium Blockchain," in *IEEE Access*, vol. 7, pp. 136704-136719, 2019, doi: 10.1109/ACCESS.2019.2943153

[15] R. N. Nortey, L. Yue, P. R. Agdedanu and M. Adjeisah, "Privacy Module for Distributed Electronic Health Records(EHRs) Using the Blockchain," *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, Suzhou, China, 2019, pp. 369-374, doi: 10.1109/ICBDA.2019.8713188.

[16] M. Mito, K. Murata, D. Eguchi, Y. Mori and M. Toyonaga, “A Data Reconstruction Method for The Big-Data Analysis,” in 2018 9th International Conference on Awareness Science and Technology (iCAST), 2018.

[17] S. H. Begum and F. Nausheen, “A comparative analysis of differential privacy vs other privacy mechanisms for Big Data,” in 2018 2nd International Conference on Inventive Systems and Control (ICISC), 2018.

[18] F. Khodaparast, M. Sheikhalishahi, H. Haghighi and F. Martinelli, “Privacy Preserving Random Decision Tree Classification Over Horizontally and Vertically Partitioned Data,” in 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology

Congress(DASC/PiCom/DataCom/CyberSciTech), 2018.

[19] T. Guo, K. Dong, L. Wang, M. Yang and J. Luo, “Privacy Preserving Profile Matching for Social Networks,” in 2018 Sixth International Conference on Advanced Cloud and Big Data (CBD), 2018.