A scheme for sharing data that is both traceable and secure, utilizing blockchain technology.

JOANA NUSHI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF BACHELOR
IN
COMPUTER SCIENCE

University of New York Tirana

JULY 2023

APPROVAL PAGE

This is to certify that I have read this project and that, in my opinion, it is fully adequate, in scope and quality, as a thesis for the degree of Bachelor of Computer Science

A scheme for sharing data that is both traceable and secure, utilizing blockchain technology.

Joana Nushi

Miralda Cuka _____ _____

This is to confirm that this thesis complies with all the standards set by the Department of Engineering of University of New York Tirana.

Date Seal/Signature

PLAGIARISM CLEARANCE PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

First Name, Last name: Joana Nushi

Signature:

Contents

Table of Figures:

Table of Tables:

ABSTRACT

In the era of digitalization, secure and efficient data sharing is crucial for collaborative environments. However, traditional data-sharing strategies often rely on centralized platforms, leading to concerns regarding transparency and data security. The proposed method introduces a secure and traceable data-sharing system that utilizes encryption and blockchain-based storage mechanisms. Firstly, an encryption-based method is employed to protect shared data and enable precise control over access. Only authorized parties can access the data while maintaining data integrity and confidentiality. It combines on-chain and off-chain technology by encrypting data of the hash value and IPFS. The blockchain stores the hash value of the encrypted data, creating an unchangeable and transparent record of the data's integrity. Simultaneously, genuine data is securely saved off-chain via the InterPlanetary File System (IPFS). To further enhance security, elliptic curve cryptography (ECC) encryption is applied before storing the hash value.

Furthermore, a log tracking system based on smart contracts is implemented, enabling real-time tracking of data sharing activities. This system provides the identity tracking requirements of both data sharing parties, ensuring accountability and transparency. Data sharing records are maintained on the blockchain and visualized for easy monitoring and auditing. The effectiveness of the suggested approach in protecting information, facilitating immediate identification monitoring, and guaranteeing efficient data transmission is supported by empirical findings. By reducing the risks associated with centralized platforms while boosting data security and privacy, this collaborative data-sharing system offers a secure, open, and effective substitute for centralized platforms for data sharing in collaborative situations.

Keywords: Blockchain, Data sharing, Data tracking, Smart contracts, IPFS, ECC, Off-chain storage, On-chain storage

ABSTRAKT

Ne epoken e digitalizimit, shperndarja e te dhenave ne menyre te sigurt dhe efikase eshte shume e rendesishme. Megjithate, strategjite tradicionale per shperndarjen e te dhenave shpesh bazohen ne platform ate centrializuara, duke sjelle shqetesime ne lidhje me transparencen dhe sigurine e te dhenave. Metoda e propozuar paraqet nje system te sigurt per ndarjen e te dhenave duke perdor mekanizmin e kodimit dhe depozitimit me ane te teknologjise se Blockchain. Se pari, perdoret nje metode bazuar ne kodim per te mbrojtur te dhenat e shperndara dhe per te lejuar nje kontroll te sakte mbi qasjen. Vetem palet e autorizuara mund te kene qasje tek te dhenat duke ruajtur integritetin dhe konfidencialitetin e te dhenave. Kjo metode kombinon teknologjine e on-chain dhe off-chai duke koduar te dhenat e vleres hash dhe IPFS. Blockchain ruan vleren e hash te te dhenave te koduara, duke krijuar nje rregjistrim te pashmangshem dhe transparent te te dhenave. Ne te njente kohe, te dhenat e verteta ruhen me siguri jasht blockchain permes IPFS. Per te permiresuar me tej sigurine, perdoret kodimi me kodin eliptik ECC pare se te ruhet vlera e hash.

Ky sistem ofron kerkesat per gjurmimin e identitetit te dy paleve te ndarjes se te dhenave, duke siguruar pergjegjshmeri dhe transparence. Rregjistrot e ndarjes se te dhenave ruhen ne blockchain dhe vizualizohen per monitorim dhe auditim te lehte. Efikasiteti i kesaj metode kundrejt mbrojtjes se informacionit dhe garantimit te transmetimit efikas mbeshtetet nga zbulimet empirike. Duke reduktuar rreziqet qe lidhen me platformat qendrore ndersa rrit sigurine dhe privatesine e te dhenave.

Fjalet kyce: Blockchain, Ndarja e te dhenave, Gjurmimi I te dhenave, IPFS, ECC, Depozitimi Off-chain, Depozitimi On-chain

CHAPTER 1

Introduction

In the contemporary world, data symbolizes a wide range of symbols that capture factual events in diverse formats like text, numbers, and images. The rapid progress of information technology has led to a significant upsurge in the amount of data generated by individuals from different backgrounds, turning it into a priceless and desirable resource for society. However, due to varying business requirements, data often becomes fragmented and gets distributed among different departments and individuals, leading to the presence of numerous segregated datasets. Regrettably, this fragmentation gives rise to "data silos" and severely restricts the effective utilization of data, thus impeding its true value. Unlocking the complete capacity of data resources, it becomes imperative to establish mechanisms for facilitating data sharing, enabling individuals to distribute their data for the benefit and use of others.

Traditionally, data sharing relies on third-party platforms as intermediaries. In this process, data demanders pay the platform to access the data uploaded by data owners. After the data exchange concludes, the platform compensates the data owner with a fee. However, keeping data on a third-party platform exposes the data owner to potential loss of control, as it becomes more susceptible to alterations and unauthorized resale. Additionally, the responsibility for maintaining data-sharing records lies solely with the third-party platform administrators, preventing data owners from tracing their own data and leading to a lack of transaction transparency and security.

Blockchain technology offers an alternative solution to address the shortcomings of conventional data sharing. Blockchain ensures data integrity and transparency by employing an immutable and transparent ledger system. If the data recorded gets tempered, it would disrupt the entire chain due to its chronological structure. Moreover, all participants in the blockchain collaborate to safeguard the data stored within it. This decentralized and tamper-proof nature of blockchain presents new possibilities for overcoming the limitations that we have nowdays.

Blockchain-based data-sharing solutions have witnessed widespread adoption across diverse domains, including supply chains, smart grids, and electronic medical records. However, while these solutions offer advantages such as immutability and transparency, they also present challenges in ensuring robust data security. The inherent openness of data stored on the blockchain raises concerns regarding unauthorized access and privacy breaches. To overcome these limitations, innovative approaches and enhancements are required to strike a balance between data security and efficient data management within blockchain-based data-sharing systems.

In summary, current data-sharing protocols using blockchain technology ensure data integrity, transparency, and trackability. However, transparency poses privacy risks, and the increasing data volume strains blockchain storage.

1. A solution is an attribute-based data security, It allows for "one-to-many" encryption and decryption of shared data. With this technology, individuals who own the data have the ability to adjust the decryption control over encrypted data.

2. We have devised a collaborative approach for data security storage that combines on-chain and off-chain elements. By encrypting and storing shared data off-chain through IPFS, while simultaneously storing the encrypted hash value on the blockchain, we ensure robust data security while alleviating the storage burden on the blockchain.

3. Through the utilization of smart contracts, we have devised a cutting-edge system for recording visual data sharing. This innovative system automatically stores comprehensive data-sharing records while simultaneously enabling robust identity monitoring of all entities engaged in the data-sharing process.

Related Work

Variuos industries have had the need for data-sharing solutions. In the healthcare sector, pioneering experiments like the Medical Data Sharing System have demonstrated the potential of blockchain technology in improving patient access to medical information among different hospitals. However, MedRec's reliance on the PoW consensus mechanism has led to computational burdens and delays, prompting further exploration of more efficient alternatives.

To establish a transparent and patient-centric data-sharing environment, ongoing healthcare initiatives aim to develop secure and auditable systems for sharing electronic health records. Drawing inspiration from MedRec and MedicalChain, researchers have proposed a easy access control strategy for medical records that leverages blockchain technology. This strategy empowers patients to upload their medical records easily, facilitating seamless and secure data sharing.

Beyond healthcare, the potential applications of blockchain-based data sharing have extended to diverse domains. For instance, in the food supply chain, Liu, Sun, and Song (2020) introduced a blockchain-based framework that enhances data sharing and traceability. Similarly, Majdalawieh, El-Hajj, Eid, and Ali (2021) proposed a blockchain-based scheme to ensure data security and traceability in supply chains. In the context of industrial Internet of Things (IoT), Yang, Wu, and Li (2022) presented a blockchain-based architecture that records all data sharing operations for enhanced security and auditing. Furthermore, Chenli, Li, and Wang (2022) developed a blockchain-based data-sharing platform utilizing smart contracts to regulate access, providing an additional layer of control and security. (Wang & Guan, 2023)

While blockchain technology offers inherent advantages in terms of data integrity and immutability, the transparency of the blockchain can also introduce risks of data leakage. To address these concerns, researchers have explored various encryption methods to safeguard data stored on the blockchain. Studies by Zhaoliang, Huang, and Wang (2021) and Zhong et al. (2022) highlight the effectiveness of encryption techniques in securing blockchain data. For instance, Dong et al. (2020) proposed encrypting shared food information using the Advanced Encryption Standard (AES), mitigating potential vulnerabilities. However, symmetric encryption

techniques, like AES, raise concerns regarding key leakage since they use one key for everything.

To address key management challenges, Zheng et al. (2018b) proposed the use of key keeper apps in conjunction with blockchain technology for secure data exchange in cloud environments. By storing encryption keys in dedicated key keeper programs and validating access rights through the blockchain, this approach enhances security and control. Alternatively, Baralla et al. (2021) suggested employing the RSA. Rivest-Shamir-Adleman is an asymmetric encryption algorithm that reduces the risk of key leakage. However, the RSA encryption technique may introduce efficiency concerns, impacting overall system performance.

In the pursuit of privacy preservation and enhanced security, researchers have explored innovative techniques such as proxy reencryption and searchable encryptionproposed. The Medchain architecture, which utilizes proxy reencryption to protect people privacy. However, this approach raises concerns regarding vulnerability to attacks.Park et al. (2021)

Homomorphic encryption-based strategies are innovative approaches that allow for computations to be performed on encrypted data without decrypting it. This advanced encryption technique enables secure processing of sensitive information while preserving privacy. With homomorphic encryption that offers privacy-preserving solutions for various usages, including cloud computing, data analytics, and machine learning.

To further optimize data query efficiency, the adoption of storage systems like the InterPlanetary File System (IPFS) has been proposed. IPFS utilizes multiple nodes to store data copies, allowing for simultaneous file reading by multiple nodes, thus significantly improving query performance.

Blockchain-based data-sharing solutions continue to have new development, with new advancements and applications emerging across industries. In the financial sector, blockchain technology is being explored for secure and transparent transactional data sharing. Researchers are investigating the potential of blockchain in streamlining cross-border payments, reducing transaction costs, and enhancing financial inclusion (Rosenfeld, 2022).

Moreover, the use of blockchain in government and public administration is gaining traction. Governments around the world are exploring blockchain-based systems for secure and efficient

data sharing in areas. For instance, the United Arab Emirates has implemented a blockchain-based platform called "UAE Pass" to streamline government services and facilitate secure data exchange between government entities and citizens (Alkhateri et al., 2021).

In the context of intellectual property rights management, blockchain is being considered as a solution to combat piracy and ensure proper attribution of digital assets. Blockchain-based platforms can create transparent and tamper-proof records of ownership, allowing artists, musicians, and creators to protect their intellectual property rights and receive fair compensation for their work (Iansiti & Lakhani, 2017), (Wang & Guan, 2023).

In summary, the current advancements in blockchain-based data sharing hold immense potential for enhancing security, transparency, and traceability across various industries. However, ongoing research endeavors aim to address the challenges posed by data leakage, encryption efficiency, storage strain, and secure access to shared data. The exploration of attribute-based encryption algorithms, hybrid storage solutions, and the integration of emerging technologies like IPFS which offer promising avenues.

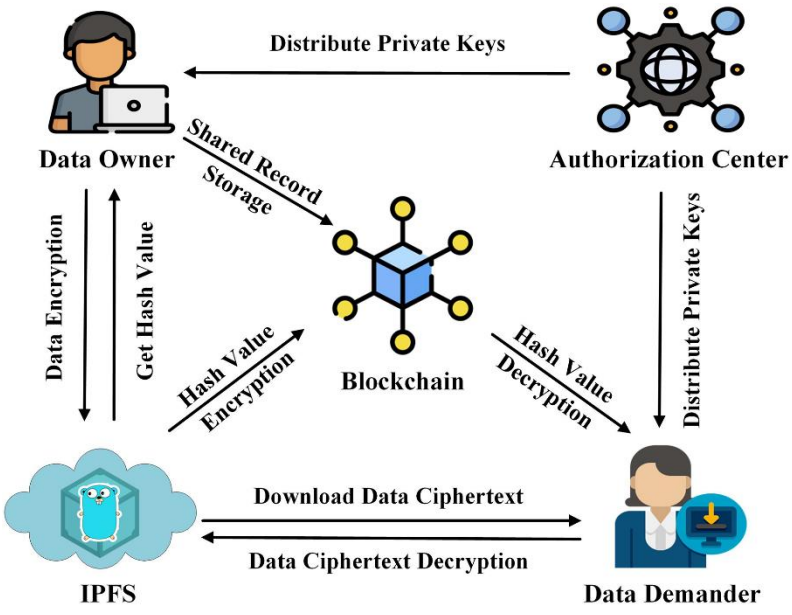


Figure 1: System architecture

In order to address the challenges at hand, we introduce a groundbreaking solution that incorporates attribute-based encryption to ensure the security of shared data while satisfying the demand for efficient

"one-to-many" encryption and decryption capabilities. By allowing for single-user encryption and multiple-user Using IPFS as a storage mechanism providing a scalable and reliable storage solution while making sure not to increase storage. Increasing security by combining IPFS and ECC algorithms. This doubly encrypted hash value is then stored on the blockchain, preventing unauthorized access and unauthorized decryption solely based on the stored hash value.

Our proposed solution harnesses the capabilities of attribute-based encryption to provide robust protection for shared data while allowing for efficient and customizable access control. By connecting certain traits to users, our technique ensures that only those with the required attribute set may decode the private key and extract the plaintext from the ciphertext. This attribute-based encryption technique not only significantly increases the process' overall security but also reduces the computational complexity of encryption and decryption. Wang and Guan (2023) further support the efficacy and benefits of this approach.

Moreover, the utilization of IPFS as a storage solution in our proposed framework brings numerous benefits. IPFS stores redundant copies of data across multiple nodes, ensuring availability and security. By using IPFS, we add security and reliability to applications. Furthermore, IPFS facilitates efficient data query operations by allowing multiple nodes to simultaneously read and retrieve the files, significantly improving data query performance (Nizamuddin et al., 2019).

We add an additional layer of protection: before putting the hash value in IPFS, encrypt it with elliptic curve cryptography (ECC). This extra encryption step adds an additional level of safeguarding to the encrypted data stored within the IPFS system. ECC offers strong encryption capabilities with relatively shorter key lengths, making it an efficient choice for securing the doubly encrypted hash value on the blockchain. By employing ECC encryption, we ensure that unauthorized users cannot retrieve the ciphertext by solely accessing the hash value stored on the blockchain, thus adding an additional layer of protection to the data sharing process.

By combining attribute-based encryption, IPFS storage, and doubly encrypted hash values on the blockchain, our proposed solution offers a robust and secure data-sharing framework. Unauthorized access, storage constraints, data security, and access control issues are all resolved by it. By carefully integrating these technologies, we want to increase data sharing's overall efficacy and security and set the framework for more dependable and trustworthy blockchain-based data-sharing ecosystems.

CHAPTER 2

Designing the structure:

A data-sharing strategy that increases data security and minimizies blockchain storage work. As indicated in the diagram, the system architecture is made up of four components.

- IPFS (InterPlanetary File System): IPFS is a decentralized protocol that offers off-chain storage services to data owners. It provides a distributed and secure solution for storing and retrieving files, offering efficiency advantages compared to storing data directly on the main blockchain.
- Authorization Center: The authorization center is responsible for generating the necessary cryptographic keys. These keys are essential for secure data encryption and decryption.
- Data Owner: The data owner applies attribute encryption techniques to protect the data. This process involves encrypting the data using attribute-based keys, enabling "one-to-many" encryption and decryption.
- Blockchain: The blockchain plays a crucial role in establishing a data-sharing log record. It employs ECC (Elliptic Curve Cryptography) encryption on the hash value of the off-chain data ciphertext.

Attribute Encryption: A Data Protection Method for Secure and Controlled Access

Looking at asymmetric encryption we see that the time required for encryption and decryption is higher compared to symmetric encryption. To address these constraints, attribute encryption appears to be a potential approach. Attribute encryption extends asymmetric encryption techniques and tackles the key leaking in symmetric encryption systems.

Bethencourt, Sahai, and Waters (2007) propose attribute encryption as a concept that encompasses various strategies, which can be summarized as follows:

1. Setup: The authorization center determines the security parameter λ and the global attribute set U. Using the Setup function, the center generates the public parameter GP and the master key MSK required for encryption.

2. Key Generation: Given a user's attribute set Au and the master key MSK, the KeyGen function generates the corresponding attribute private key SK specific to that user.
3. Encryption: A user encrypts data by selecting the access structure T, providing the public parameter GP, and inputting the data M. The Encrypt function processes this information to produce the ciphertext CH representing the encrypted data.
4. Decryption: To decrypt data, a user employs their attribute private key SK, their attribute set Au, and the ciphertext CH. The Decrypt function reverses the encryption process, retrieving the original data M from the ciphertext CH.

The access structure employed in attribute encryption determines which users within the attribute set possess the necessary authority to decrypt. By matching a user's attributes with the specified access structure, their private key becomes capable of decrypting the ciphertext effectively. This attribute-based encryption method ensures the security of data stored in the blockchain while fulfilling the requirements for scenarios where multiple users need to encrypt or decrypt the data.

A visual representation, in the form of a flowchart (Figure 2), illustrates the step-by-step process of the off-chain data security technique based on attribute encryption. The process begins with the authorization center initiating the essential parameters during the system startup phase. Following this, the center generates a unique attribute private key for the data demander based on their specific attribute set. Meanwhile, the data owner customizes the access structure by specifying the required attributes for decrypting the ciphertext.

In an effort to overcome the prevailing challenges, we introduce a revolutionary approach that encompasses attribute-based encryption to ensure the protection of shared data.

When the access structure is established and the data owner has succecfully uploads the ciphertext to IPFS, an extra layer of enxryption is added by using ECC. Furthermore, as participants in the data exchange process, the blockchain scrupulously registers their identifying details, generating a comprehensive and transparent record of their actions for future reference and auditing requirements.

Upon decryption of the hash value, the data demander gains access to the IPFS data ciphertext. Armed with their attribute private key, we can unlock the ciphertext, thereby revealing the

original data in its unencrypted form. This streamlined two-step process ensures that the data demander can securely and confidently retrieve the desired information stored on IPFS, enabling them to perform their authorized tasks effectively.

By leveraging the attribute encryption system and employing this seamless process, we establish a robust and secure framework for off-chain data security. This approach not only provides authorized access to data stored on IPFS but also ensures the integrity and confidentiality of the shared information throughout the entire data-sharing process.

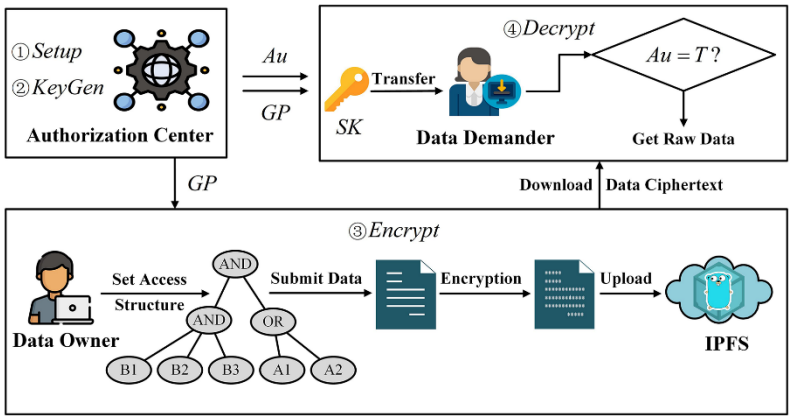


Figure 2: Development of a data safeguarding technique utilizing attribute-based encryption.

An innovative solution for securely storing data in collaborative environments, catering to both on-chain and off-chain settings.

To meet the need for "one-to-many" data encryption and decryption, attribute encryption technology proves highly efficient. The possibility of unauthorized access to the encrypted data exists, however, when several data requesters have the same features. Additional steps must be taken to resolve this issue, including strict access control regulations, reliable authentication procedures, and ongoing monitoring of data access activities. By including these safeguards, the attribute-based encryption system's overall security is improved, guaranteeing that only authorized users with the right qualities may access and decode the data. This all-encompassing strategy provides protection against data abuse or unauthorized disclosure.

Introducing elliptic curve cryptography, or ECC, a very popular method used to increase security. It is based on elliptic curves over finite fields and their mathematical features. One of the primary benefits of ECC is its ability to provide solid security while using lower key lengths than classic encryption algorithms such as RSA.

Because CPU power and memory are limited, using ECC is the solution. ECC uses a public key and a private key provided from the equation. Public keys can be shared, but only specify person can have access to the private key. Data can be encrypted into ciphertext with mathematical operations or with the private key. ECC is frequently used in applications needing secure communication and data protection because it provides security, effective key management, and saves data storage. Its efficiency and lower processing needs have made it popular.

$E: y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_5$

where a_i (i=1, 2, 3, 4, 5) ∈ K and $\Delta \neq 0$, K is the specified rational number field, and D is the elliptic curve equation's discriminant:

$$\begin{cases} \Delta = -d_2^2 d_8 - 8d_4^3 - 27d_6^2 + 9d_2d_6 \\ d_2 = a_1^2 + 4a_2 \\ d_4 = 4a_2 + a_1a_3 \\ d_6 = a_3^2 + 4a_5 \end{cases}$$

$d_8 = a_1a_5 + 4a_2a_3 - a_1a_3a_4 + a_2a_3^2 - a_4^2$

Table 1 : Algorithm for On-chain and Off-chain .

1.Obtain the hash value HCH of the data from IPFS.

2.Generate elliptic curves using the parameters Ep (a, b).

3.Determine the coordinates (x, y) of the elliptic curve group.

4.If the data owner has encrypted the ciphertext's hash value HCH:

a. Calculate the elliptic curve equation using Equation (3) to find the value of y.

b. Retrieve the points that satisfy the elliptic curve equation E (a, b) and locate the base point G (x0, y0).

c. Generate a private key r for the data demander and compute their public key R = rG using the base point G (x0, y0).

d. Share the public key R with the data owner.

e. The data owner encrypts the hash value HCH using the data demander's public key R.

f. Output the encrypted hash value ciphertext C and securely store it in the blockchain.

g. The data demander decrypts the ciphertext C using their private key r to obtain the decrypted hash value.

5:End the algorithm.

Eq. (1) is known as the Weierstrass equation when Eq. (2) is fulfilled by the elliptic curve E (Falco et al., 2018). We may get the generic equation for the elliptic curve by reducing Eq. (2) further:

$E : y2 = x3 + ax + b$

Where $(a, b, x, y) \in Ep$, Ep represents a finite field consisting of elements, and p a significantly large prime number.

Multiple processes are involved in the ECC encryption process to ensure data security. It starts by choosing an elliptic curve, denoted by Ep (a, b), and a base point G on that curve. The associated public key, R = rG, is then calculated after selecting a private key, indicated as r. R and G are points on the elliptic curve Ep (a, b).The plaintext is encrypted by embedding it into a point on the specified elliptic curve, and the encryption procedure employs the public key R. Only the intended receiver with the accompanying private key may successfully decrypt the ciphertext and get the original plaintext. The private key plays a crucial role in the decryption process as it allows the recipient to perform the necessary mathematical operations to reverse the encryption and obtain the original data.

The described technique follows a specific process to ensure secure data access:

1. Acquisition of Hash Value: We get the hash value from the InterPlanetary File System (IPFS). The hash value serves as a unique identifier for the data.
2. Generating Elliptic Curves: Using the provided parameters, the technique generates elliptic curves, which are mathematical curves used in cryptography. These curves have specific properties that make them suitable for encryption.
3. Calculating Curve Coordinates: The technique calculates the coordinates of the elliptic curve group based on the generated curves. These coordinates help define the mathematical structure necessary for encryption.
4. Solving the Elliptic Curve Equation: If the data owner has encrypted the hash value of the ciphertext (encrypted data), the technique solves the elliptic curve equation to determine the value of 'y'. This step is important for further calculations.

5. Finding a Base Point: Using the determined 'y' value, the technique finds a base point on the elliptic curve. This base point acts as a reference for further encryption and decryption operations.
6. Generating Private and Public Keys: Getting the private key that only specified persons can access. And the public key that is sharable with others.
7. Encryption by the Data Owner: By using the public key of the data provided earlier, the data owner encrypts the hash value of the data securely.
8. Storing Encrypted Value on the Blockchain: The encrypted value is securely stored on the blockchain. The blockchain provides an extra layer of security.
9. Decryption by the Data Requester: People who want to access the date use the public key to decrypt the encrypted data. This allows them to access the desired data securely.

By combining attribute encryptiofn with ECC encryption, a comprehensive data security mechanism is established. The flowchart for off-chain data security based on attribute encryption involves activities such as parameter generation, attribute key creation, access structure customization, data encryption, storage on IPFS, encryption of the hash value, and storage on the blockchain. This two-step process ensures that authorized data demanders can securely access the desired information while unauthorized users are prevented from obtaining the ciphertext and performing decryption operations.

Table 2: Smart Contract Interface Design.

Contract Function	Contract Logic	Contract Method	Contract Description
User Registration	Identity Registration	userRegist()	Registers the data owner or demander in the system.
Data Write	Hash Value Sent	hashSent()	The data owner sends the encrypted hash value of the off-chain data to the data demander.
Hash Value Receive		hashReceived()	The data demander receives the encrypted hash value of the data owner's off-chain data.
Data Query	Data Owner Query	querySent()	Data owners query data-sharing records on-chain.
Data Demander Query		queryReceived()	Data demanders query data-sharing records on-chain.

The presented table displays the interface design of smart contracts, encompassing various contract functions, their corresponding contract logic, contract methods, and a concise explanation of each function. The functionalities include user registration, data writing, hash value exchange, and data querying.

Mechanism for tracking logs using smart contracts.

Within blockchain networks, event logs play a crucial role in documenting and recording various activities. These operational logs, continuously generated within the blockchain network, offer valuable insights into the sequence of events and enable efficient identity monitoring in the context of data-sharing processes. Smart contracts, acting as essential interfaces for interacting with the blockchain, possess inherent capabilities for automated execution and immutability, ensuring seamless storage and retrieval of data-sharing records. Recognizing the significance of this aspect, we have developed a sophisticated log tracking system that utilizes smart contracts as its foundation. For a comprehensive understanding of the system's functionality, refer to Table 1, which outlines the well-designed interface of the smart contract.

Algorithm 2 provides a detailed depiction of the step-by-step process involved in creating and querying these vital data-sharing records. By seamlessly integrating these functionalities into the log tracking system, we aim to streamline and optimize the overall data-sharing experience while ensuring the utmost security and reliability.

The log tracking system based on smart contracts offers numerous benefits in terms of data governance and accountability. By automatically storing data-sharing records on the blockchain, the system ensures the integrity and immutability of the logged information. This provides a transparent audit trail that can be accessed and verified by authorized parties, eliminating any concerns of data tampering or unauthorized modifications.

The utilization of smart contracts as the underlying technology for the log tracking system also brings increased efficiency and automation. Through the automated execution of smart contracts, the system eliminates manual intervention and reduces the risk of human error in recording and managing data-sharing events. This streamlines the process and enables real-time tracking of data exchanges, empowering stakeholders with up-to-date and accurate information.

The visual presentation of logs adds a new dimension to the understanding and analysis of data-sharing processes. By transforming the raw data-sharing records into visually appealing and intuitive logs, users can easily comprehend the flow of events, identify patterns, and gain valuable insights. This visualization aspect enhances decision-making capabilities and facilitates effective monitoring of data exchanges. The smart contract interface enables data owners to securely upload their data-sharing records to the blockchain. Data demanders can easily access and verify the authenticity of the records, ensuring efficient and transparent data exchange.

Overall, the log tracking system based on smart contracts revolutionizes the way data-sharing processes are monitored and governed. By leveraging the power of blockchain technology and smart contracts, it enhances transparency, accountability, and efficiency in data exchanges. The intuitive interface, automated execution, and visual representation of logs contribute to a robust and trustworthy data governance framework, empowering organizations and individuals to securely share and access data while maintaining a comprehensive record of all transactions.

CHAPTER 3: Experiments and analysis

The experimental environment

During the experimental phase, a blockchain network was built using Hyperledger Fabric, as described by Kumar et al. in their 2021 research. IPFS (InterPlanetary File System) was used as the underlying infrastructure to meet the blockchain's storage requirements. Table 2 contains particular setup data for the software used in the experimental environment. Building on this foundation, a user-friendly solution for data exchange and transfer was created. Through a web interface, this solution interfaces easily with both Hyperledger Fabric and IPFS, offering an easy and efficient method of communication between data owners and data demanders inside the system. By utilizing this solution, the exchange of data becomes effortless and enables smooth collaboration between all parties involved.

Table 3: Algorithm for creating and retrieving data-sharing records.

1. If the user registration is successful (based on the "userRegistration" function), the user is registered in the system and assigned an identity ID.
2. Once registered, the data owner encrypts the data and obtains the ciphertext CH.
3. The algorithm computes the hash value HCH of the encrypted data.
4. Algorithm 1 is called to obtain the hash of the ciphertext C.
5. If the condition is met (userRegistration is successful), the algorithm proceeds to the next steps.
6. If the "sendHash" function is successful, the data owner sends the encrypted hash value (C) to the data demander, along with the owner's ID (OwnerID) and the demander's ID (DemanderID).
7. If the data demander successfully receives the encrypted hash value and other information, the algorithm proceeds.
8. The algorithm returns the transaction ID and the block height (blockNum) associated with this data-sharing event.
9. If the data owner wants to query the data-sharing record using the transaction ID, they can use the "sendQuery" function.
10. If the data demander successfully receives the query and retrieves the query result, the algorithm proceeds.
11. The algorithm returns the query result to the data demander.

Table 4: Data space occupancy

System Component	Description
Virtual machine version	VMware15 pro
Hyperledger Fabric version	Hyperledger Fabric 2.4
IPFS version	IPFS 0.14.9
Smart contract language	Go 15.7
Computer CPU	Intel Core i5-10300H @ 2.50 GHz
Memory	16 GB

Off-chain data security experiment

Data owners keep their data off-chain in a variety of forms, including video, audio, images, and text, during the data exchange process. Figure 3 shows an example of off-chain data encryption utilizing a text file.

Figure 3A shows the data encryption testing interface, whereas Figure 3B shows the encrypted text file itself. The shared data is turned into ciphertext using attribute-based encryption as a way of data security, guaranteeing that the data stays safely hidden outside the blockchain. This method also allows several data demanders to decode shared data encrypted by a single data owner, allowing for a "one-to-many" encryption and decryption procedure. As a result, this reduces the system overhead associated with repetitive data encryption. The encrypted data is then uploaded, and Figure 4 depicts the hash value of the encrypted data retrieved from IPFS.

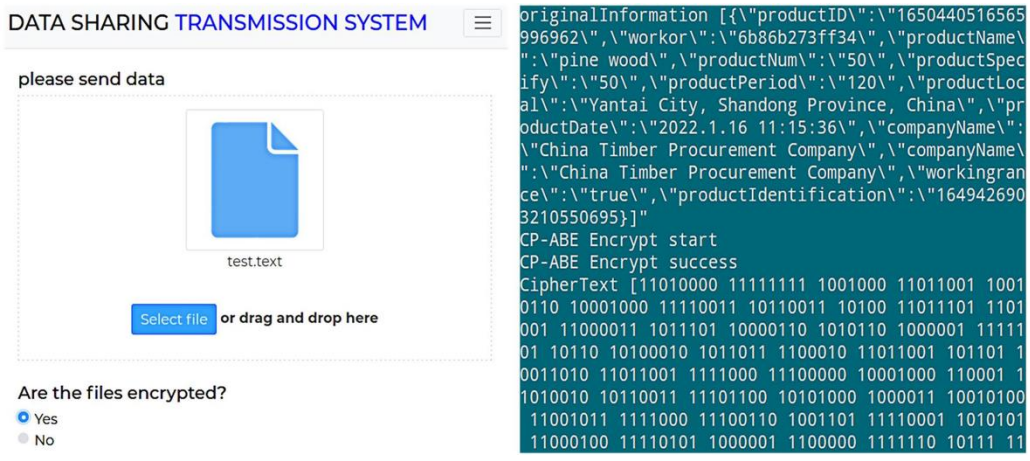


Figure 3: The encrypted data

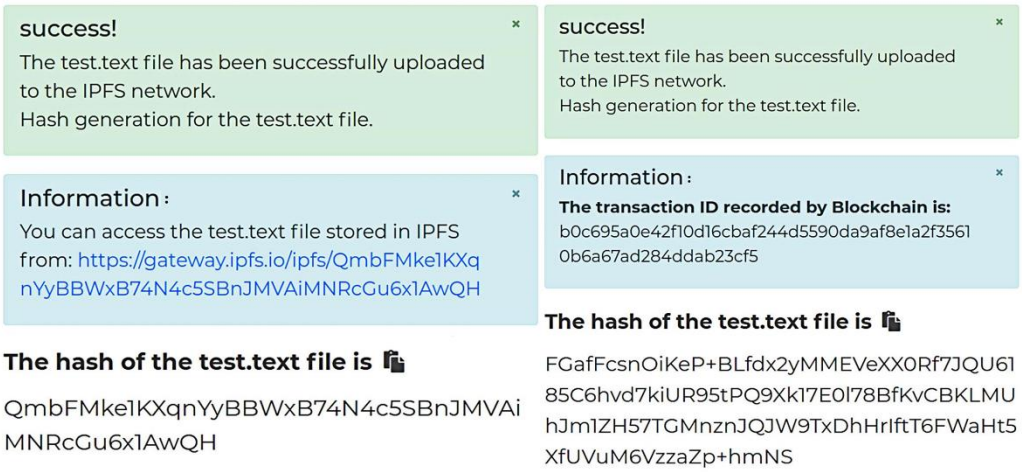


Figure 4: The cryptographic hash value of the plaintext before and after the encryption process.

Figure 4, tells us that the data ciphertext can be encrypted or left unencrypted, leading to different outcomes. The result of an unencrypted hash of the data ciphertext(Figure 4A) and encryption (Figure 4B).

Figure 4A highlights that if the hash value is not encrypted, unauthorized data demanders can utilize it to directly extract the data ciphertext from IPFS. However, in contrast, if the characteristics of a data demander meet the necessary attributes for decrypting the ciphertext, they can successfully decrypt the data ciphertext.

Figure 4B gives us a ciphertext given by the ECC encryption and the data ciphertext stored in IPFS. In the IPFS system, only the data demander possessing the corresponding private key can decrypt the hash value and gain access to the ciphertext.

Figure 5 presents an example of a visual log, which demonstrates how smart contracts are employed to automatically record and track the data-sharing process. This process is visualized through visual logs, enabling the tracking of the identities of both parties involved in the shared data. The use of smart contracts ensures the seamless recording and monitoring of the data-sharing process, while the visual logs provide a clear visual representation of the process and enhance transparency.

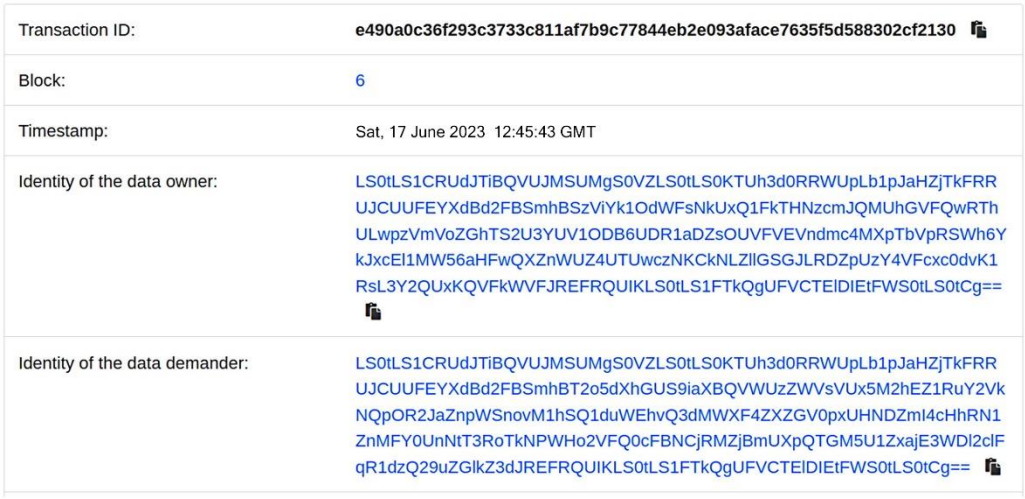


Figure 5: Shared records

According to the explanation, Figure 5 is a graphical depiction of a chronological record recording each data sharing event between the data owner and the data demander. This visual log provides critical information such as the block number, the names of the individuals involved, and timestamps showing the moment of sharing. By providing this visual record, participants in the data exchange process can be identified in real time, increasing openness and accountability throughout the data sharing process. (Zhenqi & Shaopeng, 2023)

To minimize the storage burden on the blockchain, our approach adopts an efficient strategy by offloading the storage of shared data to off-chain platforms. This allows for the preservation of storage capacity within the blockchain for essential blockchain-related data. Table 3 shows us the importance of reducing storage space. It tells us the needed required space and the space used from the blockchain.

Table 5: Data space

Raw Data Size	Off-chain Storage Solution	Proposed Scheme
1 MB	1 MB	0.1 MB
5 MB	5 MB	0.1 MB
10 MB	10 MB	0.1 MB
15 MB	15 MB	0.1 MB

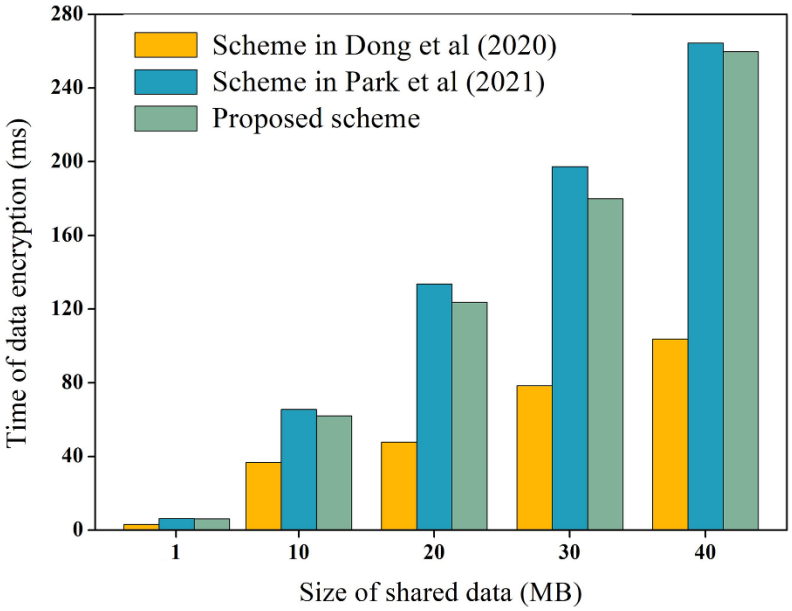
By utilizing off-chain platforms for shared data storage and efficiently managing storage resources within the blockchain, our approach optimizes storage usage while maintaining the integrity and accessibility of the shared data in a scalable and sustainable manner.

Performance analysis of the system

Analyzing the cost of encrypting and decrypting data: by utilizing off-chain platforms for shared data storage and effectively managing storage resources within the blockchain. Shown in Table 3.

We performed the experiment 100 times and calculated the averages to confirm the dependability of our experimental data. Figure 6 depicts the original data's encryption overhead, whereas Figure 7 depicts the decryption overhead. We conducted a performance evaluation of our proposed technique in comparison to the methods presented by Dong et al. (2020) and Park et al. (2021) for original data encryption and decryption. Dong et al. (2020) proposed the utilization of the AES symmetric encryption technique, while Park et al. (2021) suggested employing the proxy re-encryption strategy for data privacy protection. These recently introduced approaches are technologically similar to our proposed strategy and served as the basis for comparison.

Figure 6: Comparation of encryption



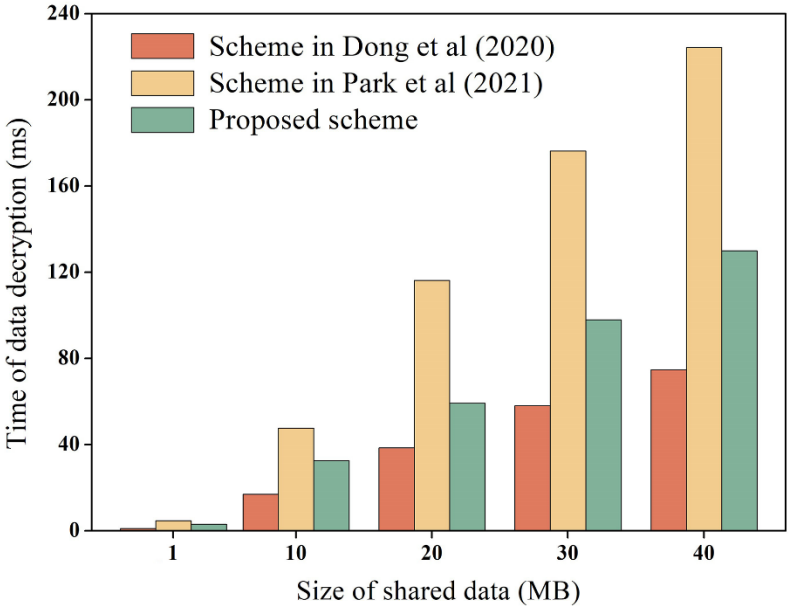


Figure 7: Comparation of decryption

Upon analyzing Figures 6 and 7, it becomes evident that AES encryption and decryption have the lowest execution cost, while proxy re-encryption incurs the highest execution cost. The primary factor contributing to this divergence is the inherent characteristics of the AES encryption and decryption process. AES employs a byte round-robin conversion operation, which allows for relatively swift encryption and decryption of original data using same key. Consequently, data owners must share the secret key with data requesters in order to enable the decryption of the encrypted data. This key exchange process raises concerns regarding the potential leakage of the secret key, posing a risk to the overall security of the data. In contrast, the suggested approach eliminates the necessity for key exchange during the initial encryption and decryption of data, hence increasing security. Furthermore, the proposed technique surpasses the proxy re-encryption method in terms of efficiency for encrypting and decrypting the original material. The proxy re-encryption method includes transforming the ciphertext into several ciphertexts via a proxy server, allowing various data requesters to decode them. However, this multi-step procedure adds complexity and escalates the system's implementation cost.

We founded the average value which is shown in Table 4.

Table 6: Overhead of Encryption and Decryption on Hash Values.	
Encryption time	156.24 µs
Decryption time	86.54 µs

The data presented in Table tells the efficiency of the system impacted from the ECC algorithm. It not only added a extra layer of protection but it also caused a significant decrease in the system's overall efficiency.

In the context of this research, the measurement of the number of transactions completed per second on the blockchain network is data throughput. It serves as a crucial indicator to evaluate the process cost. We specifically focused on assessing the throughput performance for shared data storage and query operations.

In order to ensure the accuracy and reliability of our experimental results, we conducted multiple iterations of the experiments, specifically 100 times, and derived the average outcomes. This comprehensive approach allowed us to obtain robust and representative results for performance evaluation. The findings of these experiments are visually presented in Figure 8, enabling a clear and intuitive understanding of the comparative analysis between the different schemes being examined.

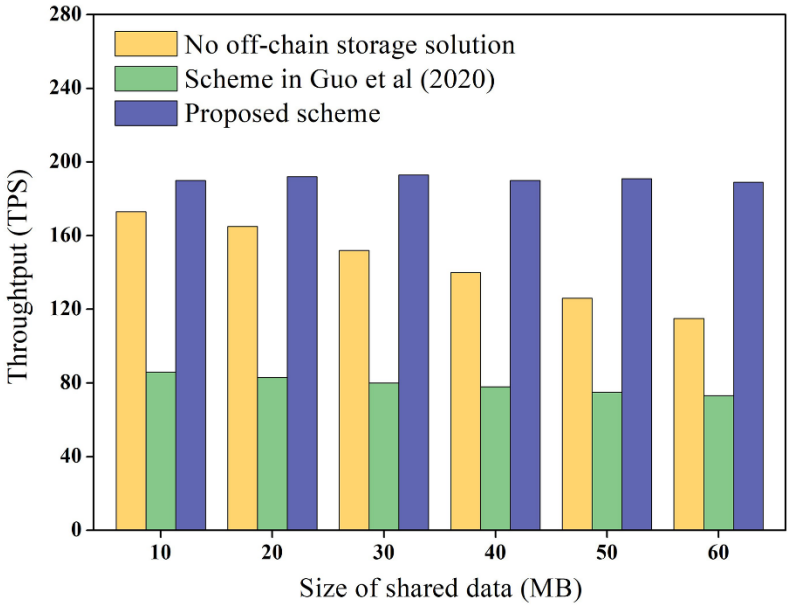


Figure 8: Experimental Results

The throughput patterns that were found in various methods are shown visually in Figure 8. It indicates that as data size rises, the throughput of the method without under-chain storage gradually lowers and that the throughput of the Guo et al. (2020) strategy is lowerer. In contrast, the suggested approach achieves a greater consistent throughput than the chainless under-storage scheme. The operational discrepancies between the systems can be related to the throughput differential. During the data storage and query procedures, the Guo et al. (2020) method requires cross-chain activities between the master and slave chains. These cross-chain activities invariably generate inefficiencies in blockchain storage and query functions, resulting in lower

performance. In contrast, the suggested technique that stores the encrypted hash value.This economical storage strategy contributes to the suggested scheme's greater throughput. The smaller data size reduces computational overhead, resulting in better speed and throughput as compared to competing techniques.

Data security of the proposed scheme

In our scheme, we have thoroughly evaluated various. We hope to evaluate the effectiveness and performance of our scheme in regard to these essential data security measures.

1. Protection of Off-chain Data Security

This research contributes to the field of off-chain data security by offering a comprehensive solution to protect sensitive data stored outside the blockchain. The proposed approach successfully addresses the challenges associated with off-chain data security, ensuring confidentiality, integrity, and availability. By implementing the suggested measures, decentralized systems can strengthen their overall security posture and foster greater trust among users.

2. Protection of On-chain Data Security

One important topic discussed in this research was the protection of on-chain data security. By using the technology of blockchain, the data stored had a great importance. By implementing different algorithms, we managed to get through the challenges by protecting on-chain data.

3. Data Integrity.

The integrity of the data was achieved by encrypting and storing the data using IPFS algorithm. The data stored in the blockchain was secure and the data integrity level was met.

4. Facilitating Data Tracking

This research has focused on the important task of facilitating data tracking within data management systems. With the exponential growth of data and the increasing complexity of data flows, effective tracking mechanisms have become essential for organizations to maintain data integrity, enhance data governance, and ensure regulatory compliance.

5. Enhancing Trustworthiness

This research has focused on enhancing the trustworthiness of data within data management systems. In today's digital landscape, trust is a crucial factor in establishing successful relationships and transactions. By addressing the challenges related to data trustworthiness, this study has contributed to strengthening the confidence and reliability of data in various domains.

6. Enforcing Authorization

This research has focused on enforcing authorization within data management systems to ensure data security and protect against unauthorized access. With the increasing volume and sensitivity of data, robust authorization mechanisms are critical for organizations to control data access and maintain confidentiality.

7. Ensuring Authentication

This research has emphasized the importance of ensuring authentication within data management systems to enhance data security and verify user identities. Future work may involve exploring emerging authentication technologies and integrating decentralized identity systems to further strengthen data security and user privacy. Overall, this research provides valuable insights and recommendations for organizations aiming to enforce authentication and bolster data protection.

8. Promoting Reliability

This research emphasizes the importance of promoting reliability within data management systems. By implementing robust data validation, error detection, and redundancy strategies, organizations can enhance data accuracy and consistency. Establishing data governance frameworks and monitoring systems further ensures ongoing reliability.

9. Enabling Validation

This research has focused on enabling validation within data management systems to ensure data integrity and accuracy. Validation plays a crucial role in verifying the correctness and reliability of data, enhancing overall data quality.

Chapter 4

Conclusion

In today's digital landscape, the secure and efficient sharing of data has become a critical challenge across various industries. The emergence of blockchain technology offers promising solutions to address data security, access control, and storage efficiency concerns. This paper presents an innovative and resilient method for data sharing that harnesses the capabilities of blockchain technology. Our proposed system introduces a traceable and secure approach to data sharing, providing a comprehensive solution to overcome the inherent limitations of traditional data sharing methods.

Attribute Encryption for Enhanced Data Security:

One of the fundamental pillars of our approach is the use of attribute encryption to protect shared data. By implementing this encryption technique, we ensure fine-grained access control, empowering data owners with the ability to define customized access hierarchies and limitations. This advanced encryption method guarantees that only authorized parties can access and decrypt shared data, mitigating the risks associated with unauthorized data access and leakage.

Smart Contract-Based Log System:

To facilitate the identification and monitoring of data sharers, we develop a sophisticated log system using smart contracts. This log mechanism serves as an immutable record of each data-sharing event, capturing critical information such as the names of the participating parties, sharing timestamps, and the specific data blocks involved. By leveraging the transparency and tamper-proof nature of smart contracts, this log system promotes openness, accountability, and trust among the data-sharing participants. Real-time identification tracking adds an additional layer of security, enabling the detection and mitigation of any unauthorized access or manipulation attempts.

Performance and Security Assessments:

To evaluate the effectiveness and efficiency of our proposed data-sharing strategy, we conducted comprehensive performance and security assessments. Through rigorous testing and comparative analysis, we compared our approach with alternative methodologies. The results demonstrated

significant improvements in data throughput and security. Our technology effectively lowers the overhead costs associated with data encryption and decryption, enhancing the overall efficiency of the system. The real-time identity monitoring feature enhances data sharing security, ensuring that any unauthorized access or manipulation attempts are promptly identified and addressed.

Finally, this study presents a traceable and secure data-sharing method that makes use of blockchain technology. We provide a comprehensive solution to the difficulties of data security, access control, and accountability by merging attribute encryption, collaborative on-chain and off-chain data storage, and a smart contract-based log system. Our suggested method improves data security, enables fine-grained access control, and optimizes storage use. The conducted performance and security assessments validate the superiority of our approach, positioning it as a promising solution for secure and efficient data exchange in various industries. By leveraging the benefits of blockchain technology, our approach establishes a robust and trustworthy framework for data sharing in the digital era.

References

1. Al Omar A, Bhuiyan MZA, Basu A, Kiyomoto S, Rahman MS. 2019. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Future Generation Computer Systems 95(15):511–521 DOI 10.1016/j.future.2018.12.044.
2. Albeyatti A. 2018. White paper: medicalchain. Chiasso: MedicalChain Self-Publication. Available at https://medicalchain.com/Medicalchain-Whitepaper-EN.pdf.
3. Azaria A, Ekblaw A, Vieira T, Lippman A. 2016. Medrec: using blockchain for medical data access and permission management. In: 2016 2nd International Conference on Open and Big Data (OBD). 25–30.
4. Baralla G, Pinna A, Tonelli R, Marchesi M, Ibba S. 2021. Ensuring transparency and traceability of food local products: a blockchain application to a smart tourism region. Concurrency and Computation: Practice and Experience 33(1):5857–5875 DOI 10.1002/cpe.5857.
5. Bethencourt J, Sahai A, Waters B. 2007. Ciphertext-policy attribute-based encryption. In: 2007 IEEE Symposium on Security and Privacy (SP'07). Piscataway: IEEE, 321–334.
6. Chen Y, Ding S, Xu Z, Zheng H, Yang S. 2019. Blockchain-based medical records secure storage and medical service framework. Journal of Medical Systems 43(1):1–9 DOI 10.1007/s10916-018-1121-4.
7. Chen J, Yin X, Ning J. 2022. A fine-grained and secure health data sharing scheme based on blockchain. Transactions on Emerging Telecommunications Technologies 113(4):1–15 DOI 10.1002/ett.4510.
8. Chenli C, Tang W, Gomulka F, Jung T. 2022. Provnet: networked bi-directional blockchain for data sharing with verifiable provenance. Journal of Parallel and Distributed Computing 166(9):32–44 DOI 10.1016/j.jpdc.2022.04.003.
9. Deepa N, Pham QV, Nguyen DC, Bhattacharya S, Prabadevi B, Gadekallu TR, Maddikunta PKR, Fang F, Pathirana PN. 2022. A survey on blockchain for big data: approaches, opportunities, and future directions. Future Generation Computer Systems 113(10):1–20 DOI 10.1016/j.future.2022.01.017.
10. Dong Z, Chen J, Chen Y, Shao R. 2020. Food traceability system based on blockchain. In: Proceedings of the 2020 International Conference on Aviation Safety and Information Technology. 571–576.

11. Falcão MI, Miranda F, Severino R, Soares MJ. 2018. Weierstrass method for quaternionic polynomial root-finding. Mathematical Methods in the Applied Sciences 41(1):423–437 DOI 10.1002/mma.4623.

12. Guo S, Wang F, Zhang N, Qi F, Qiu X. 2020. Master-slave chain based trusted cross-domain authentication mechanism in IoT. Journal of Network and Computer Applications 172(7):102812–102823 DOI 10.1016/j.jnca.2020.102812.

13. Guo L, Yang X, Yau WC. 2021. Tabe-dac: efficient traceable attribute-based encryption scheme with dynamic access control based on blockchain. IEEE Access 9(12):8479–8490 DOI 10.1109/ACCESS.2021.3049549.

14. Kaur H, Alam MA, Jameel R, Mourya AK, Chang V. 2018. A proposed solution and future direction for blockchain-based heterogeneous medicare data in cloud environment. Journal of Medical Systems 42(8):1–11 DOI 10.1007/s10916-018-1007-5.

15. Koblitz N, Menezes A, Vanstone S. 2000. The state of elliptic curve cryptography. Designs, Codes and Cryptography 19(2):173–193 DOI 10.1023/A:1008354106356.

16. Kumar R, Marchang N, Tripathi R. 2022. Smdsb: efficient off-chain storage model for data sharing in blockchain environment. In: Machine Learning and Information Processing. Berlin: Springer, 225–240.

17. Kumar R, Tripathi R, Marchang N, Srivastava G, Gadekallu TR, Xiong NN. 2021. A secured distributed detection system based on IPFS and blockchain for industrial image and video data security. Journal of Parallel and Distributed Computing 152(6):128–143 DOI 10.1016/j.jpdc.2021.02.022.

18. Li W, Zhou Z, Fan W, Gao J. 2022. Design of data sharing platform based on blockchain and IPFS technology. Wireless Communications and Mobile Computing 2022(6):1–7 DOI 10.1155/2022/3937725.

19. Liu J, Sun X, Song K. 2020. A food traceability framework based on permissioned blockchain. Journal of Cybersecurity 2(2):107–115 DOI 10.32604/jcs.2020.011222.

20. Lu Y, Li P, Xu H. 2022. A food anti-counterfeiting traceability system based on blockchain and internet of things. Procedia Computer Science 199(3):629–636 DOI 10.1016/j.procs.2022.01.077.

21. Majdalawieh M, Nizamuddin N, Alaraj M, Khan S, Bani-Hani A. 2021. Blockchain-based solution for secure and transparent food supply chain network. Peer-to-Peer Networking and Applications 14(6):3831–3850 DOI 10.1007/s12083-021-01196-1.

22. Manzoor A, Braeken A, Kanhere SS, Ylianttila M, Liyanage M. 2021. Proxy re-encryption enabled secure and anonymous iot data sharing platform based on blockchain. Journal of Network and Computer Applications 176(4):102917–102972 DOI 10.1016/j.jnca.2020.102917.

23. Nizamuddin N, Salah K, Azad MA, Arshad J, Rehman M. 2019. Decentralized document version control using ethereum blockchain and IPFS. Computers & Electrical Engineering 76(9):183–197 DOI 10.1016/j.compeleceng.2019.03.014.

24. Park YH, Kim Y, Lee SO, Ko K. 2021. Secure outsourced blockchain-based medical data sharing system using proxy re-encryption. Applied Sciences 11(20):9422 DOI 10.3390/app11209422.

25. Singh AP, Pradhan NR, Luhach AK, Agnihotri S, Jhanjhi NZ, Verma S, Ghosh U, Roy DS. 2020.

26. A novel patient-centric architectural framework for blockchain-enabled healthcare applications. IEEE Transactions on Industrial Informatics 17(8):5779–5789 DOI 10.1109/TII.2020.3037889.

27. Sun J, Yao X, Wang S, Wu Y. 2020. Blockchain-based secure storage and access scheme for electronic medical records in IPFS. IEEE Access 8(4):59389–59401 DOI 10.1109/ACCESS.2020.2982964.

28. Wan PK, Huang L, Holtskog H. 2020. Blockchain-enabled information sharing within a supply chain: a systematic literature review. IEEE Access 8(4):49645–49656 DOI 10.1109/ACCESS.2020.2980142.

29. Wang B, Li Z. 2021. Healthchain: a privacy protection system for medical data based on blockchain. Future Internet 13(10):247 DOI 10.3390/fi13100247.

30. Wang Y, Su Z, Zhang N, Chen J, Sun X, Ye Z, Zhou Z. 2020. Spds: a secure and auditable private data sharing scheme for smart grid based on blockchain. IEEE Transactions on Industrial Informatics 17(11):7688–7699 DOI 10.1109/TII.2020.3040171.

31. Wang Z, Tian Y, Zhu J. 2018. Data sharing and tracing scheme based on blockchain. In: 2018 8th International Conference on Logistics, Informatics and Service Sciences (LISS). 1–6. Xia Q, Sifah EB, Asamoah KO, Gao J, Du X, Guizani M. 2017a. Medshare: trust-less medical data

sharing among cloud service providers via blockchain. IEEE Access 5(12):14757–14767 DOI 10.1109/ACCESS.2017.2730843.

32. Xia Q, Sifah EB, Smahi A, Amofa S, Zhang X. 2017b. Bbds: blockchain-based data sharing for electronic medical records in cloud environments. Information-an International Interdisciplinary Journal 8(2):44–60 DOI 10.3390/info8020044.

33. Yang X, Li M, Yu H, Wang M, Xu D, Sun C. 2021. A trusted blockchain-based traceability system for fruit and vegetable agricultural products. IEEE Access 9(5):36282–36293 DOI 10.1109/ACCESS.2021.3062845.

34. Yang L, Zou W, Wang J, Tang Z. 2022. Edgeshare: a blockchain-based edge data-sharing framework for industrial internet of things. Neurocomputing 485(4):219–232 DOI 10.1016/j.neucom.2021.01.147.

35. Ye H, Park S. 2021. Reliable vehicle data storage using blockchain and IPFS. Electronics 10(10):1130–1145 DOI 10.3390/electronics10101130.

36. Zhang A, Lin X. 2018. Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain. Journal of Medical Systems 42(8):1–18 DOI 10.1007/s10916-018-0995-5.

37. Zhaoliang L, Huang W, Wang D. 2021. Functional agricultural monitoring data storage based on sustainable block chain technology. Journal of Cleaner Production 281(5):124078–124087 DOI 10.1016/j.jclepro.2020.124078.

38. Wang, Z., & Guan, S. (2023). A blockchain-based traceable and secure data-sharing scheme. School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai, Shandong, China. DOI: 10.7717/peerj-cs.1337

39. Zheng X, Mukkamala RR, Vatrapu R, Ordieres-Mere J. 2018b. Blockchain-based personal health data sharing system using cloud storage. In: 2018 IEEE 20th International Conference on eHealth Networking, Applications and Services (Healthcom). Piscataway: IEEE, 1–6.