A Dual-Chain and Differential Privacy-Based Solution for Medical Data Privacy Protection and Access Control

Cen Gu[1], Luping Wang[2]*, Hongjie Wu[3]*
Suzhou University of Science and Technology, 99 Xuefu Road, SuZhou, 215004, China[1,2,3]
Jiangsu Province Key Laboratory of Intelligent Building, Energy Efficiency, 99 Xuefu Road, SuZhou, 215004, China[3]

***Abstract*—As living standards rise, people are paying increasing attention to health. Vast quantities of medical data are generated daily, yet each piece contains sensitive information such as patients' names, mobile numbers, email addresses, and places of employment. Should this information be compromised, the consequences would be irreversible, causing severe damage. Traditional solutions merely implement access control policies, permitting data access only to authorised personnel. While this approach offers some protection, even compliant users cannot be entirely trusted and may engage in malicious activities. Once data is accessed, patients' sensitive information becomes fully exposed to the user, posing a significant data security risk. Therefore, this study proposes a medical data sharing scheme based on Dual-Chain and differential privacy. It employs a hybrid approach combining private chains, consortium chains, and IPFS. Internal hospital personnel can access data after de-identification, while external parties can only access data that has been de-identified and subsequently augmented with noise. This significantly enhances security. The experimental section of this study also demonstrates that the proposed scheme effectively protects data, while the data shared with external users enables them to successfully complete downstream tasks.**

***Keywords*—Blockchain; differential privacy; IPFS; medical data**

I. Introduction

People's living standards are rising as a result of the economy's quick development. Health has become an increasingly prominent concern [1][2]. Consequently, in this era of information and data, vast amounts of medical data are generated daily [3]. Sharing and analyzing this data helps doctors make better decisions and provide effective patient care. However, this data often contains sensitive patient information, such as names, employers, phone numbers, and email addresses [4][5]. Moreover, during the sharing process, data users are not entirely trustworthy and may engage in malicious activities [6]. This significantly increases the risk of information leaks, resulting in extremely low security and reliability [7].

II. Related Works

Over time, with continuous technological advancement, significant achievements have been made in the privacy protection and access control of medical data. Akinyele et al. [8] introduced an attribute-based encryption access control scheme for personal health data on mobile devices, enabling data owners to exercise granular control over data access and sharing. Zhou et al. [9] designed a CP-ABE-based data exchange mechanism that allocates access permissions based on requester attributes and defines specific access policies, incorporating attribute timestamps to support attribute revocation. However, a common limitation of these ABE-based access control systems is their reliance on a single attribute authority to manage attribute permissions and key distribution. As a result, there is a major single point of breakdown: should the authority be compromised, it jeopardises system stability and overall security, undermining reliability.

To address potential single points of failure, Chase et al. [10] proposed a solution based on multi-attribute authorization authorities. However, it still relies on a fully trusted central authority, creating a significant vulnerability: should the central authority cease to be trustworthy, the entire system's security would be compromised, and the integrity of all encrypted data would be undermined, failing to achieve true decentralization. Liu et al. [11] subsequently suggested adopting a distributed multi-authority access control scheme, yet regulatory bodies continued to play a dominant role. The system remains only partially decentralized and remains susceptible to centralized power, which undermines its security and efficiency.

Subsequently, with the advent of blockchain technology [12][13], its inherent transparency and decentralization characteristics prompted integration efforts with existing systems. In 2016, Azaria et al. [14] proposed MedRec, an electronic health record management system. By employing distributed ledger technology to store patient records on a blockchain, it created a decentralized record management framework. This approach preserves data integrity and prevents tampering, while facilitating seamless sharing of medical records between distinct healthcare providers. Ownership, access rights, and data integrity of medical records are governed through smart contracts [15][16]. B. Chen et al. [17] and B.B. Gupta et al. [18] proposed integrating blockchain with CP-ABE to enhance security. In order to overcome centralized healthcare systems' shortcomings, Wang et al. [19] developed a trusted healthcare data management platform leveraging blockchain smart contracts and attribute-based encryption for flexible access control. Jin et al. [20] combined blockchain, smart contracts, and proxy re-encryption for secure medical information sharing. However, it lacks attribute integration, limiting fine-grained access control, and depends heavily on user activity, which may compromise security. Chen et al. [21] designed a blockchain-based access control scheme with hybrid on-chain/off-chain

*Corresponding authors.

storage to reduce expense of storage. However, the absence of hashing and signing the IPFS-generated content identifier (CID) on upload enables potential injection of malicious files, lowering system security.

However, the aforementioned approaches merely protect data at the access control layer by establishing an access policy, allowing only those who meet the policy to access the data. Yet in reality, even some visitors who comply with the access policy may not be entirely trustworthy. Over time, they could potentially engage in malicious activities, leading to data leaks and misuse. Therefore, Shao et al. [22] proposed a solution combining dynamic reputation with token-based allocation. Blockchain nodes distribute varying quantities of tokens to each visitor based on their reputation. After each access event, the reputation value is updated, and the token allocation is reassigned. While this approach offers some protection for data privacy, it relies on the assumption that visitors remain trustworthy throughout. If a visitor betrays trust during the period when they possess tokens for access, it could still result in data leakage. Cai et al. [23] advised a federated blockchain-based healthcare data sharing system with enhanced proxy re-encryption for improved data sharing security [24][25]. However, data privacy protection remains inadequate. Therefore, this study addresses the shortcomings of all previous approaches by proposing a novel solution with the following key contributions:

- This work introduces a role-oriented data classification framework that structurally separates internal raw-data access from external privacy-preserving data release. Instead of applying de-identification and differential privacy as isolated preprocessing steps, the proposed scheme embeds them into distinct operational domains governed by different access policies. Internal medical personnel access encrypted, but structurally intact datasets under strict attribute-based control, while external researchers interact only with differentially private representations. This separation reduces cross-domain inference risk and limits unnecessary exposure of raw medical records.
- A functionally decoupled dual-chain architecture is designed to enforce domain-specific governance. The private hospital chain manages identity authentication and fine-grained CP-ABE-based access control for raw data, whereas the consortium chain supervises differentially private data publication and external access requests. This architectural separation is not merely a deployment choice but a security-driven design that reduces trust concentration, isolates attack surfaces, and mitigates risks associated with single-chain or centralized systems.
- The scheme integrates blockchain integrity guarantees with IPFS-based encrypted storage in a policy-enforced workflow. By storing only metadata and cryptographic hashes on-chain, while maintaining encrypted bulk data off-chain, the framework ensures tamper-evident verification without compromising scalability. Unlike conventional blockchain-IPFS integrations that focus solely on storage optimization, this design tightly couples data integrity validation with access policy enforcement and privacy quantification, forming a unified secure data-sharing model.

III. Organization

This study's general organization is separated into four major sections: Section IV displays the preliminary knowledge used in the work. Section V analyzes the overall scheme, mainly from the overall model of the scheme, the description of the scheme and the specific flow of the scheme. Section VI analyzes the security of the entire solution. The performance analysis of the suggested plan is provided in Section VII. Lastly, we conclude the study in Section VIII.

IV. Background Knowledge

This section primarily introduces the technologies used in the study: blockchain technology, differential privacy technology, bilinear mapping, and linear secret sharing.

A. Blockchain

Blockchain is a fault-tolerant, append-only replicated state machine maintained by a distributed set of nodes over an unreliable network [26][27]. Transactions are batched into blocks that commit cryptographically to prior state, yielding tamper-evident immutability under standard hash-collision assumptions. A consensus layer establishes a globally consistent total order and finality; permissionless systems achieve Sybil resistance through resource-based mechanisms, whereas permissioned deployments typically rely on Byzantine fault-tolerant agreement. Public-key cryptography underpins identity, authentication, and authorization, and light-client protocols enable inclusion verification without full replication. Deterministic smart contracts specify application logic as state-transition functions executed by validators, enabling verifiable automation, while raising concerns about correctness, composability, and adversarial execution. Current research targets scalability, latency and throughput, privacy via advanced cryptography, and cryptoeconomic robustness. Despite open challenges, blockchains provide auditability and credible neutrality for applications in digital assets, decentralized finance, and provenance tracking [28][29].

B. Differential Privacy

Differential privacy [30][31], is a provable privacy framework for statistical release and machine learning that injects noise calibrated to the query's sensitivity, ensuring that the inclusion or exclusion of any single individual has only a controlled, quantifiable effect on the output distribution; formally, for neighboring datasets $D \sim D'$ that differ in one record and a randomized mechanism M, if the inequality below holds for every measurable event S, then M satisfies (ε, δ)-differential privacy [32][33], where ε bounds multiplicative leakage and δ allows a negligible failure probability, as shown in Eq. (1):

$$\forall D \sim D', \forall S : \ \Pr\big[M(D) \in S\big] \ \leq \ e^{\varepsilon} \Pr\big[M(D') \in S\big] + \delta. \tag{1}$$

C. Bilinear Mapping

In elliptic curve cryptography, a bilinear mapping is a mapping e, in which elements of two elliptic curve groups are mapped to another group [34][35]. Suppose we have two groups G_1 and G_2, and a target group G_T [36]. We define a bilinear mapping as shown in Eq. (2):

$$e : G_1 \times G_2 \to G_T \tag{2}$$

This mapping has several important properties:

- Bilinear: For any $p \in G_1$, $q \in G_2$ and scalars a and b, the bilinear mapping is satisfied in Eq. (3):

$$e(p^a, q^b) = e(p, q)^{ab} \tag{3}$$

- Non-degradation: The bilinear mapping is not degenerate, there exists at least one $p \in G_1$ and $q \in G_2$, such that $e(p, q) \neq 1$ (unit element).
- Computability: For most implementations, computing bilinear mappings is efficient in practice.

D. Linear Secret Sharing Scheme

Linear Secret Sharing Scheme (LSSS) [37][38] first splits the secret into n distinct components and then sends these n parts to n distinct participants, and a set of authorized participants can recover the secret by merging their shares [39][40]. It is defined as follows:

The description of the access framework is (M, ρ), P stands for the group of characteristics, containing the attribute names and attribute values. M is a linear secret sharing matrix of size $l \times n$. ρ is a mapping function, and $\rho(i)$ refers to the mapping of the ith row of the matrix M to an attribute name. Map the secret s to a value on $\mathbb{Z}_p^*$ with p is a prime number. Construct a vector $\mathbf{v} = (s, r_2, r_3, \ldots, r_n)^T$ by randomly choosing $\{r_2, r_3, \ldots, r_n\} \in \mathbb{Z}_p^*$, and compute $\gamma_i = M_j \cdot \mathbf{v}$ where γ_i refers to the attribute named $\rho(i)$ — the secret share of all uses of the attribute [41][42].

Secret reconstruction: Let $B \in M$ be the collection of attributes that comply with the authorizations of the access structure, where $I = \{i \mid \rho(i) \in B\} \subseteq \{1, 2, \ldots, l\}$. Being able to find the coefficients $\{x_i \in \mathbb{Z}_p^*\}_{i \in I}$ in polynomial time such that $\sum_{i \in I} x_i M_i = (1, 0, \ldots, 0)$, and then compute $s = \sum_{i \in I} x_i \gamma_i$.

V. Scheme Structure

This proposal primarily consists of six entities: data owners, data users, attribute authorization agencies, authoritative certification agencies, IPFS, and Blockchain. Fig. 1 displays the complete model diagram.

1) Data Owner (DO): DO is primarily responsible for establishing access policies, authorizing visitors who meet these policies, and encrypting and uploading both original de-identified medical data and data with privacy-preserving noise added to IPFS. They then upload metadata to the blockchain, comprising the CID returned by IPFS, the CID's hash value, and the hash value of the encrypted ciphertext.

2) Data User (DU): DU here likely refers to internal medical personnel and external research scientists. Before accessing data, DU must submit an authorization request to DO. Only DU meeting the access policy are authorized to access the data.

3) Attribute Authorities (AA): Multiple Attribute Authorities work together to manage all attributes, authenticate the identity of the visitor and generate the visitor's attribute key.

4) Certification Authority (CA): All users must register with the CA, which issues certificates to users, generates their public and private keys, and assigns them a unique *id*.

5) Interplanetary File System (IPFS): IPFS primarily stores large-volume raw medical data to alleviate storage tension on the chain and returns the CID to the DO.

6) Blockchain (BC): This is mostly made up of a consortium blockchain created by outside research institutes and a private blockchain inside the hospital. The blockchain stores metadata and records all activities.

The following steps make up the overall scheme's basic process:

- Step 1: Each entity in the system first registers with CA and obtains its own public and private key, which it uploads to the BC.
- Step 2: DO encrypts both types of data using symmetric encryption algorithms and upload them to IPFS. Metadata composed of the CID of data that has undergone only de-identification processing is uploaded to the hospital's internal private chain. Metadata composed of the CID of data that has been added noise via DP is uploaded to the consortium chain. Separate access policies are established for internal and external personnel.
- Step 3: DU requests authorization from DO to access data. DO verifies compliance with access policies. If compliant, AA will assign DU an attribute private key. Blockchain nodes will send metadata to DU.
- Step 4: Upon receiving the metadata, DU first performs a hash operation on the CID within the metadata. If the hash matches, it uses the CID to request access to the original data from IPFS.
- Step 5: After obtaining the encrypted raw data from IPFS, DU first performs a hash operation on the ciphertext and verifies its consistency with the hash value stored in the metadata. If consistent, it uses its own attribute private key to decrypt the symmetric encryption key, thereby decrypting the raw data for access.

A. Scheme Definition

1) Initialization phase: During the initialization phase, all entities register with the CA and generate global parameters.

a) CA.setup: CA runs *CA.Setup* to initialize the system. Given the security parameter λ, it selects a prime p and instantiates multiplicative cyclic groups G and G_T of order p with generator g, together with the bilinear map $e : G \times G \to G_T$.

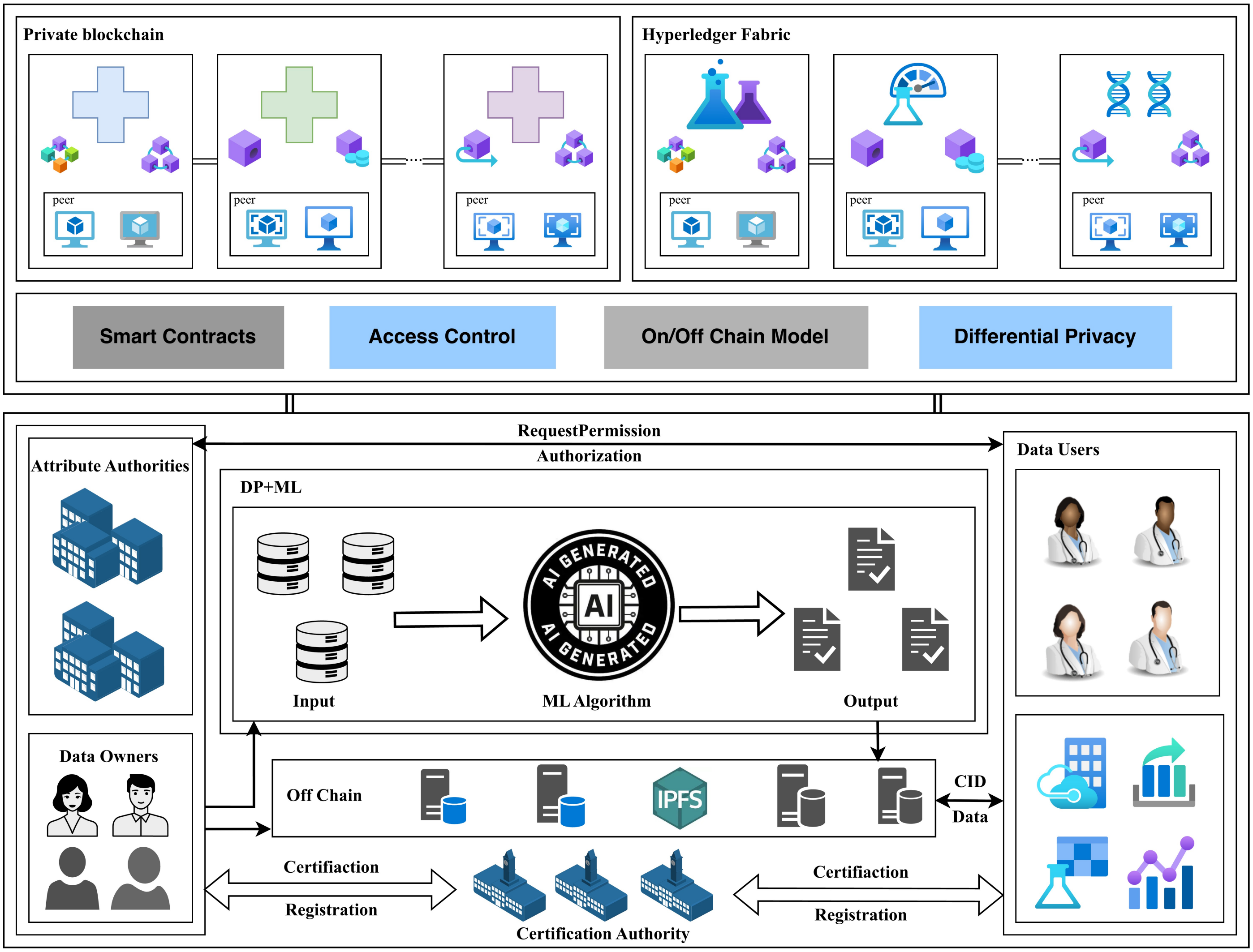

Fig. 1. System model.

It also chooses hash functions $H : \{0,1\}^* \to G$. The resulting global public parameters GPP are given in Eq. (4):

$$GPP = \{G, G_T, g, p, H, e\} \tag{4}$$

b) AA.register: Every AA submits a request together with its identifying information $info_a$, to register with the CA. The CA derives a unique identifier $aid \leftarrow H(info_a)$. The AA then samples $\alpha_{aid}, \beta_{aid} \in \mathbb{Z}_p$ to form its private key SK_{aid}, computes the public key $PK_{aid} = \left(g^{\beta_{aid}}, e(g,g)^{\alpha_{aid}}\right)$, and registers this public key with the CA.

c) DO.register: Each DO registers with the CA by submitting a request along with its identity information $info_o$. The CA computes a unique identifier $oid \leftarrow H(info_o)$. The DO then samples $\gamma \in \mathbb{Z}_p$ as its private key SK_{oid}, derives the corresponding public key $PK_{oid} = g^{\gamma}$, and registers this public key with the CA.

d) DU.register: By submitting a request and supplying its identifying information $info_u$, each DU registers with the CA. The CA derives a unique identifier $uid \leftarrow H(info_u)$. The DU samples $\mu \in \mathbb{Z}_p$ as its private key SK_{uid}, computes the corresponding public key $PK_{uid} = g^{\mu}$, and registers this public key with the CA.

2) Key generation phase:

a) KeyGen: The user requests a private key from each attribute authority AA_{aid} that governs their attributes. Let S be the user's attribute set, and assume $att \in S$ is administered by AA_{aid}. Before issuance, the authority verifies the user's identity uid and the validity of $Cert_{uid}$. If verification succeeds, AA_{aid} samples $r_{aid} \in \mathbb{Z}_p$ and constructs the user's attribute private-key component, as in Eq. (5):

$$SK_{att} = \left(g^{\alpha_{aid}} \cdot H(att)^{r_{aid}},\ g^{r_{aid}}\right) \tag{5}$$

Then the final private key of the user is the set of private keys produced by all relevant attribute authorizations, as shown in Eq. (6):

$$SK_{uid,att} = \bigcup_{aid} \{SK_{att} \mid att \in S \cap AA_{aid}\} \tag{6}$$

3) De-identification and DP phase: We utilized 500,000 outpatient medical records, first de-identifying sensitive fields such as patient names, workplaces, mobile numbers, and email addresses to protect privacy. This anonymized data was provided to internal personnel for viewing and use. Subsequently, we employed a rule-driven weakly supervised learning approach to match the de-identified data with a set of positive-class terms (infection-related) from the "initial diagnosis" field, generating binary labels as shown in Eq. (7). During feature construction, we exclusively utilized non-diagnostic text categories ("chief complaint + medical history"), while excluding diagnostic keywords used for annotation. This yields a bag-of-words count vector, as shown in Eq. (8). For each sample vector, we perform per-sample processing, as described in Eq. (9) and apply ℓ_1 regularization to limit its contribution. During the differential privacy phase, independent Laplace noise is added to each dimension of the pruned count vector, as shown in Eq. (10):

$$y = 1\{\, \exists\, k \in k_{\text{pos}} \mid k \subset \text{diagnosis} \,\} \tag{7}$$

$$x = \big[\, c(w_1), \ldots, c\big(w_{|v|}\big) \,\big] \tag{8}$$

$$x^{\text{clip}} = x \cdot \min\!\left(1, \frac{C}{\|x\|_1}\right) \tag{9}$$

$$\tilde{x} = x^{\text{clip}} + \eta,\; \eta_j \overset{\text{i.i.d.}}{\sim} \text{Laplace}(0, b),\; b = \frac{C}{\epsilon}. \tag{10}$$

4) Data encryption phase: DO chooses to employ symmetric encryption techniques to encrypt both the de-identified data and the data with additional noise via DP due to the effectiveness of encrypting massive amounts of raw data. DO selects different symmetric encryption key $k \in \mathbb{G}_T$ to encrypt each type of data separately to acquire the ciphertext C. Separate access policies (M, ρ) used to encrypt different symmetric key k. Here, M is an $l \times n$ matrix and ρ maps each row index i to an attribute. First sample a random vector $\mathbf{v} = (k, v_2, \ldots, v_n) \in \mathbb{Z}_p^n$. Compute the secret shares, as in Eq. (11); each x_i is associated with the attribute $\rho(i)$. For each row i, assume $\rho(i)$ is administered by authority AA_{aid}; choose $t_i \in \mathbb{Z}_p$ and derive the ciphertext component, as in Eq. (12). Next, compute the pairing value, as in Eq. (13). The final metadata is given in Eq. (14):

$$x_i = M_i \cdot \mathbf{v},\; \forall i \in \{1, \ldots, l\} \tag{11}$$

$$C_{i,1} = g^{\beta_{aid} x_i},\; C_{i,2} = H(\rho(i))^{x_i},\; C_{i,3} = g^{t_i} \tag{12}$$

$$D_i = e(g,g)^{\alpha_{aid} x_i} \cdot e(g,g)^{\beta_{aid} t_i} \tag{13}$$

$$\begin{aligned} metadata = \Big(&C, hash(C),\; CID,\; hash(CID), \\ &\{C_{i,1},\, C_{i,2},\, C_{i,3},\, D_i\}_{i=1}^{\ell},\; k \cdot \prod_{aid} e(g,g)^{\alpha_{aid} k}\Big) \end{aligned} \tag{14}$$

5) Data decryption phase: If DU meets the access policy, in addition to a reconstruction factor $\{w_i\}$ such that $\sum_{i \in I} w_i M_i = (1, 0, \ldots, 0)$. The pairwise value of each share is recovered through Eq. (15) and use the reconstruction coefficients w_i to decrypt the symmetric key k, according to Eq. (16) and Eq. (17). Eventually the DU decrypts the C via the obtained symmetric key k to get the data it wants to access.

$$\frac{e(C_{i,1}, g^{\alpha_{aid}} H(\rho(i))^{r_{aid}})}{e(C_{i,3}^{\beta_{aid}}, g^{r_{aid}})} = \frac{e(g^{x_i}, g^{\alpha_{aid}} H(\rho(i))^{r_{aid}})}{e(g^{\beta_{aid} t_i}, g^{r_{aid}})} \tag{15}$$

$$\prod_{i \in I} (e(g,g)^{\alpha_{aid} \lambda_i})^{w_i} = e(g,g)^{\sum_{i \in I} \alpha_{aid} \lambda_i w_i} = e(g,g)^{\sum_{aid} \alpha_{aid} k} \tag{16}$$

$$k = \frac{k \cdot \prod_{aid} e(g,g)^{\alpha_{aid} k}}{e(g,g)^{\sum_{aid} \alpha_{aid} k}} \tag{17}$$

VI. Security Analysis

This section primarily analyzes the proposed scheme from three perspectives: threat model, correctness, and resistance to attacks.

A. Threat Model

In this work, we consider a probabilistic polynomial-time adversary operating in a realistic medical data-sharing environment. The adversary is assumed to have full visibility of all public system parameters, blockchain transactions, on-chain metadata, content identifiers CID, and encrypted data stored on IPFS. In particular, the adversary may obtain complete read access to ciphertext corresponding to both internally shared de-identified datasets and externally shared differentially private datasets. The adversary may attempt active manipulation, including CID substitution, ciphertext tampering, replay of outdated metadata, or injection of malicious files into the IPFS storage layer. Multiple malicious data users may also collude by combining their attribute private keys in an attempt to satisfy an access policy not satisfied individually. Furthermore, the adversary may possess auxiliary background knowledge, such as demographic attributes or partial medical information, and attempt linkage attacks, membership inference, or statistical re-identification against externally released data. At the system level, metadata analysis and traffic pattern observation, including access frequency and timing information, are also considered possible attack vectors.

We assume that standard cryptographic hardness assumptions hold, including the discrete logarithm and bilinear Diffie–Hellman assumptions in pairing-based groups, as well as the semantic security of symmetric encryption. The Certification Authority is trusted to correctly initialize the system and protect master secrets, while Attribute Authorities follow the protocol for attribute key issuance and are not assumed to be fully colluding with the adversary. Blockchain nodes operate under Byzantine fault tolerance assumptions, and the adversary controls fewer nodes than the consensus threshold required to rewrite confirmed ledger history. IPFS is considered an untrusted storage layer and provides no confidentiality guarantee. Under these assumptions, we define the security objectives of confidentiality, fine-grained access control, collusion resistance, data integrity, and differential privacy protection for external data release.

B. Proof of Correctness

First, consider the reconstruction properties of LSSS. Let B denote the set of attributes satisfying the access policy, and let the corresponding row index set be:

$$I = \{\, i \mid \rho(i) \in B \,\} \subseteq \{1, \ldots, l\}.$$

According to the LSSS definition, there exists a set of reconstruction coefficients $\{w_i\}_{i\in I} \subset \mathbb{Z}_p$, such that:

$$\textstyle\sum_{i\in I} w_i M_i = (1, 0, \ldots, 0).$$

From Eq. (11), it can be seen that the secret share for each row is:

$$x_i = M_i \cdot v \quad (v = (k, v_2, \ldots, v_n)^T).$$

By taking a linear combination of these shares, k is obtained by Eq. (18):

$$\begin{aligned} \sum_{i\in I} w_i x_i &= \sum_{i\in I} w_i (M_i \cdot v) \\ &= \left(\sum_{i\in I} w_i M_i\right) \cdot v \\ &= (1, 0, \ldots, 0) \cdot v \\ &= k \end{aligned} \tag{18}$$

In the decryption process description, the text uses λ_i to denote the secret share associated with the i-th row, equivalent to the aforementioned x_i. The above derivation demonstrates that provided the access structure is satisfied, the secret k can be linearly recovered using an appropriate reconstruction coefficient w_i. For each row $i \in I$, administered by AA_{aid}, the DU holds the attribute private-key component:

$$SK_{\rho(i)} = \big(g^{\alpha_{aid}} H(\rho(i))^{r_{aid}},\, g^{r_{aid}}\big).$$

Using the ciphertext components and the private key above, DU computes the pairing ratio in Eq. (19):

$$\begin{aligned} \frac{e(C_{i,1},\, g^{\alpha_{aid}} H(\rho(i))^{r_{aid}})}{e\Big(C_{i,3}^{\beta_{aid}},\, g^{r_{aid}}\Big)} &= \frac{e\big(g^{\beta_{aid} x_i},\, g^{\alpha_{aid}} H(\rho(i))^{r_{aid}}\big)}{e(g^{\beta_{aid} t_i},\, g^{r_{aid}})} \\ &= e(g,g)^{\alpha_{aid}\lambda_i} \end{aligned} \tag{19}$$

By bilinearity, the masking randomness t_i, r_{aid} cancels out, leaving a factor depending only on $\alpha_{aid}\lambda_i$. DU then uses the reconstruction coefficients w_i over all $i \in I$ to obtain $e(g,g)^{\sum_{aid} \alpha_{aid} k}$, where the last step uses $\sum_{i\in I} w_i \lambda_i = k$. This yields the masked exponent $\sum_{aid} \alpha_{aid} k$. Finally, combining this with the masked key term from the metadata, DU acquires k by Eq. (20):

$$k = \frac{k \cdot \prod_{aid} e(g,g)^{\alpha_{aid} k}}{e(g,g)^{\sum_{aid} \alpha_{aid} k}} \tag{20}$$

C. Resistance to Attacks

The proposed scheme leverages multi-authority CP-ABE, blockchain with IPFS, and differential privacy to jointly provide confidentiality, integrity, and privacy, thereby resisting several typical classes of attacks. For passive eavesdropping and key-guessing attacks, even a powerful adversary with full access to blockchain transactions and IPFS storage only observes the ciphertext C, the CID and their hashes, the ABE ciphertext components $\{c_{i,1}, c_{i,2}, c_{i,3}, D_i\}$, and the masked key term $k \cdot \prod_{aid} e(g,g)^{\alpha_{aid} k}$. Without any legitimate attribute private key SK_{att}, the adversary cannot form the pairing ratios, nor can they derive $e(g,g)^{\sum_{aid} \alpha_{aid} k}$. Attempting to guess k from $k \cdot \prod_{aid} e(g,g)^{\alpha_{aid} k}$ requires solving discrete logarithm or DBDH-type problems in bilinear groups, which is infeasible under appropriate parameters. The scheme is thus robust against eavesdropping and brute-force attacks.

For unauthorized access and collusion, access control is enforced by a multi-authority CP-ABE with an LSSS access structure. If a user's attribute set S does not satisfy the policy, there exist no coefficients w_i such that $\sum_{i\in I} w_i M_i = (1, 0, \ldots, 0)$, and hence $\sum_{i\in I} w_i \lambda_i = k$ cannot hold. Even when multiple unauthorized users collude and pool their attribute keys into a union set S_U, if this union still fails to satisfy the access matrix, LSSS theory guarantees that no valid reconstruction coefficients can be found, so the colluding users cannot elevate their collective access privileges or derive k beyond what their authorized attributes permit.

For data tampering and replay attacks, the scheme relies on blockchain-stored hashes and timestamps to guarantee integrity and auditability. Upon decryption, DU recomputes the hashes of the received CID and the ciphertext C retrieved from IPFS and compares them to on-chain values. Any modification of CID or ciphertext causes hash mismatches and will be detected and rejected. Replay attacks that attempt to resubmit outdated metadata or access requests are mitigated by the append-only, timestamped blockchain ledger, which records the latest state and access logs. The system can detect and prevent abnormal replays by checking against the most recent blockchain state, thus resisting tampering and replay attacks on IPFS and the blockchain.

Regarding inference attacks and privacy leakage, external researchers only see de-identified and differentially private feature vectors. De-identification removes explicit identifiers such as names and contact information from raw records, while the differential privacy mechanism bounds the impact of any single record on the output distribution within a factor of e^{ε}. This formally limits the advantage of any adversary, regardless of side information, in inferring sensitive attributes of any individual patient from $\tilde{x}$. Taken together, the cryptographic confidentiality of k, fine-grained access control, integrity protection, and rigorous differential privacy guarantees show that the proposed scheme can effectively withstand eavesdropping, tampering, collusion, replay, and inference attacks, while still supporting practical and privacy-preserving medical data sharing.

TABLE I. Comparison of Existing Schemes

Aspect	**Scheme**					
	[17]	**[18]**	**[19]**	**[20]**	**[23]**	**Ours**
Patient-centered	✓	✓	✓	✗	✗	✓
Data Confidentiality	✓	✓	✓	✓	✓	✓
Identity Privacy Protection	✗	✗	✓	✗	✓	✓
Secure Distributed Storage	✗	✗	✓	✓	✓	✓
Fine-grained Access Control	✓	✗	✓	✗	✗	✓
Data Privacy Protection	✗	✗	✗	✗	✗	✓
Dual-Chain	✗	✗	✗	✗	✗	✓

TABLE II. Grid of C/ϵ Values

	ϵ					
$C\backslash\epsilon$	6	8	12	20	30	40
3	0.500	0.375	0.250	0.150	0.100	0.075
4	0.667	0.500	0.333	0.200	0.133	0.100
5	0.833	0.625	0.417	0.250	0.167	0.125
6	1.000	0.750	0.500	0.300	0.200	0.150

VII. Performance Analysis

Every experiment in this work was carried out on a device equipped with an Apple M4 processor, 16GB RAM, and macOS Sequoia 15.5 operating system.

A. Functional Analysis

According to Table I, the proposed scheme in this study outperforms existing approaches in terms of patient-centricity, identity information protection, data confidentiality, secure distributed storage, granular access control, and data privacy protection. In [18] and [23], the authors neglect identity privacy protection in their approaches, while [17] relies on cloud storage for medical data, introducing potential data security risks. In [19], the authors employ trust-based organizational access control, limiting patients' ownership control over their medical data. In [20], the authors fail to achieve granular access control with its identity-based approach. All these solutions neglect privacy-preserving data processing. Consequently, the proposed solution in this study demonstrates significant advantages across these dimensions.

B. Performance Evaluation of DP

During the differential privacy phase, we add independent Laplace noise to each dimension of the truncated counts and employ a dimensionless noise intensity *scale*$= C/\varepsilon$ to uniformly compare the privacy–utility tradeoff under different truncation thresholds and privacy budgets. This study systematically scans combinations of $C \in \{3,4,5,6\}$ and $\varepsilon \in \{6,8,12,20,30,40\}$, as shown in Table II. We split the 500,000 data points into 95% for training, 5% for validation, and 5% for testing. During training, we feed the de-identified features—perturbed with differential privacy—into a continuous scoring function $s(x) = \sigma(w^\top x + b_0)$. On the validation set, we perform a grid search over the threshold $t \in [0,1]$ and select the value that maximizes the F_1 score; this threshold is then fixed for the testing phase. Finally, we visualize the evaluation metrics of the experimental results, as shown in Fig. 2.

First, in Fig. 2(a), Fig. 2(b), and Fig. 2(d), the three metrics AUROC, AUPRC, and F_1 increase monotonically as ε grows, from a moderate regime at $\varepsilon = 6$ the performance is in a moderate regime with AUROC in [0.78,0.84], AUPRC in[0.45,0.62], and F_1 in [0.48,0.59]. At $\varepsilon = 12$ the metrics rise to AUROC in [0.88,0.91], AUPRC in [0.68,0.79], and F_1 in [0.64,0.71]. Near $\varepsilon = 20$, the method enters a high-performance regime with AUROC in [0.96,0.97], AUPRC in [0.91,0.94], and F_1 in [0.83,0.86]. Further increasing to $\varepsilon = 30$ and 40 approaches the no-DP ceiling both AUROC and AUPRC near 0.99, $F_1 > 0.95$. This pattern rapidly gains with larger ε followed by diminishing returns beyond $\varepsilon \approx 20$ indicates that our method preserves strong predictive signal under moderate privacy, while additional relaxation yields limited benefit.

Second, the effect of the clipping threshold C appears as a horizontal shift at fixed ε: smaller C (more aggressive clipping) lowers the curves, with larger gaps at small ε; when ε is sufficiently large, the curves converge, suggesting that the impact of clipping on utility is attenuated at high ε.

Third, and crucially, [see Fig. 2(c)] in the summary plot of *scale* $= C/\varepsilon$ versus AUPRC, the points from different C values almost lie on a single master curve, showing that the noise *scale* $= C/\varepsilon$ is the key dimensionless variable governing utility: as *scale* increases from 0.1 to 1.0, AUPRC decreases smoothly and monotonically, and results for different C align at the same *scale*. This "curve collapse" not only confirms the theoretical expectation of our clip-then-noise design, but also provides a practical tuning rule for deployment.

Overall, under the Laplace mechanism with rigorous differential privacy guarantees, high utility is achieved already at moderate ε. The metrics improve monotonically with ε and exhibit a clear "sweet spot" around $\varepsilon = 12 - 20$, and all three evaluation dimensions consistently support these conclusions, indicating that DP-processed shared data are useful and controllable for downstream binary classification.

C. Time Consumption

In this step, we performed encryption and decryption operations on 24 datasets of approximately 500MB each after adding noise via DP, recording the time consumed, as shown in Fig. 3. This figure illustrates the distribution of time overhead across 24 data files during the symmetric encryption

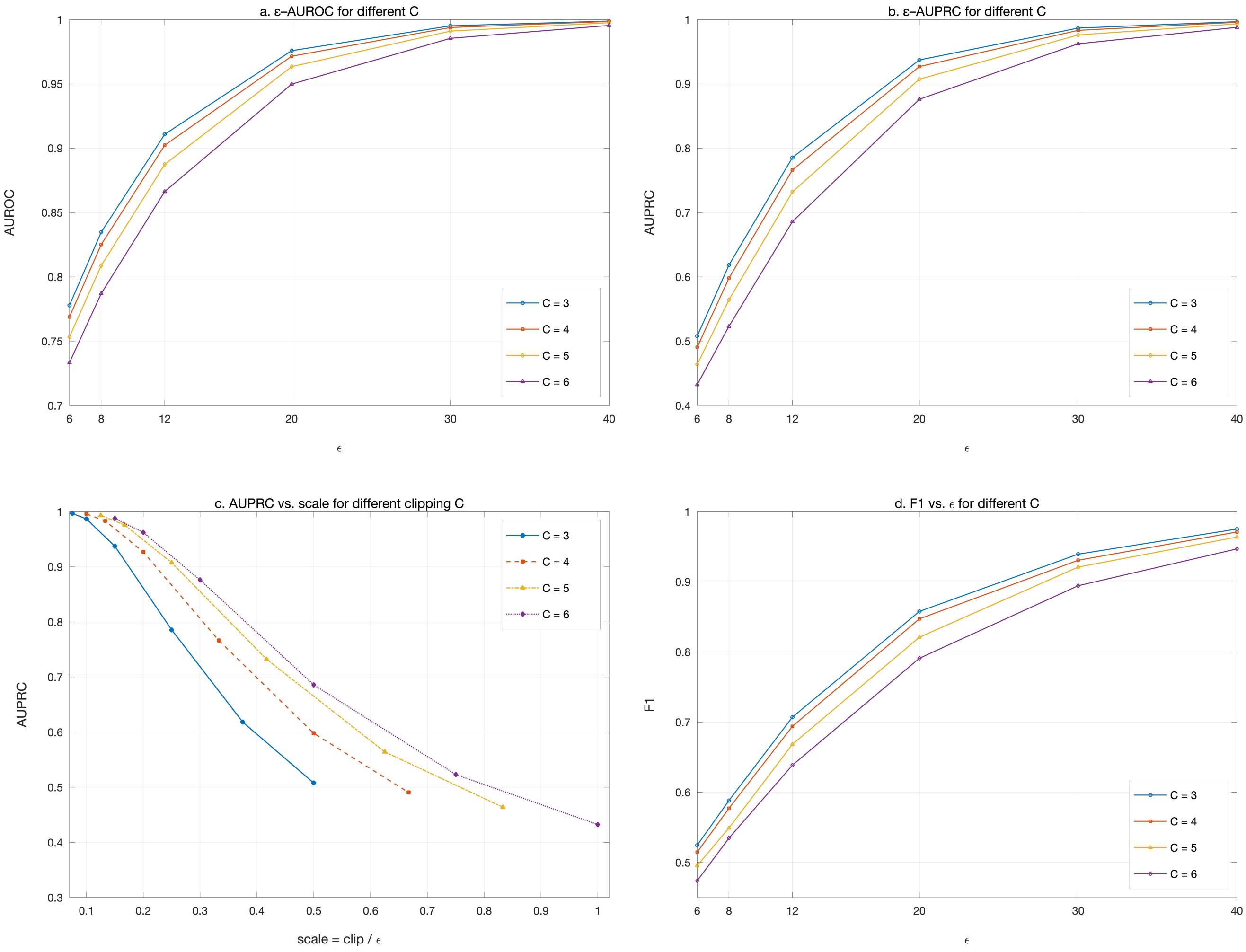

Fig. 2. Experimental results display.

TABLE III. Hash Operation Time

Aspects	Time					
	dp_C3_e6	dp_C3_e8	dp_C3_e12	dp_C3_e20	dp_C3_e30	dp_C3_e40
enc	0.215 s	0.205 s	0.224 s	0.203 s	0.231 s	0.216 s
CID	0.008 ms	0.008 ms	0.005 ms	0.008 ms	0.007 ms	0.010 ms

and decryption phases. Overall, single-pass processing latency remains sub-second: the average encryption time is $0.294\,\mathrm{s}$ (range 0.255–$0.372\,\mathrm{s}$), while the average decryption time is $0.263\,\mathrm{s}$ (range 0.243–$0.292\,\mathrm{s}$). Compared with encryption, decryption achieves an average reduction of approximately $0.031\,\mathrm{s}$, a 10.3% decrease in latency, reflecting the lighter key-recovery path. Both phases exhibit low variability: the standard deviation of encryption time is $0.029\,\mathrm{s}$ and that of decryption time is $0.014\,\mathrm{s}$. This indicates stable and predictable operation across file/parameter combinations ($C = 3$–6, $\varepsilon = 6$–40). Notably, the most time-consuming samples occur under the C=3 configuration with smaller ε values, whereas other settings remain low and comparable. These results demonstrate the symmetric scheme's strong scalability and engineering viability—delivering consistent latency without sacrificing throughput, while further reducing processing overhead on the decryption side. After uploading the encrypted files to IPFS, the system obtain the CID and compile the metadata for upload to BC. We recorded the time consumed by the hash operations on the CID and ciphertext within the metadata, as shown in Table III. The hash operations for the ciphertext consistently took around 0.2s, while the CID hash times were in the millisecond range—demonstrating highly efficient processing. Finally, we recorded the time taken to upload each of these 24 files to IPFS and download them from IPFS. As shown in Fig. 4, both processes were completed within one second, demonstrating high efficiency.

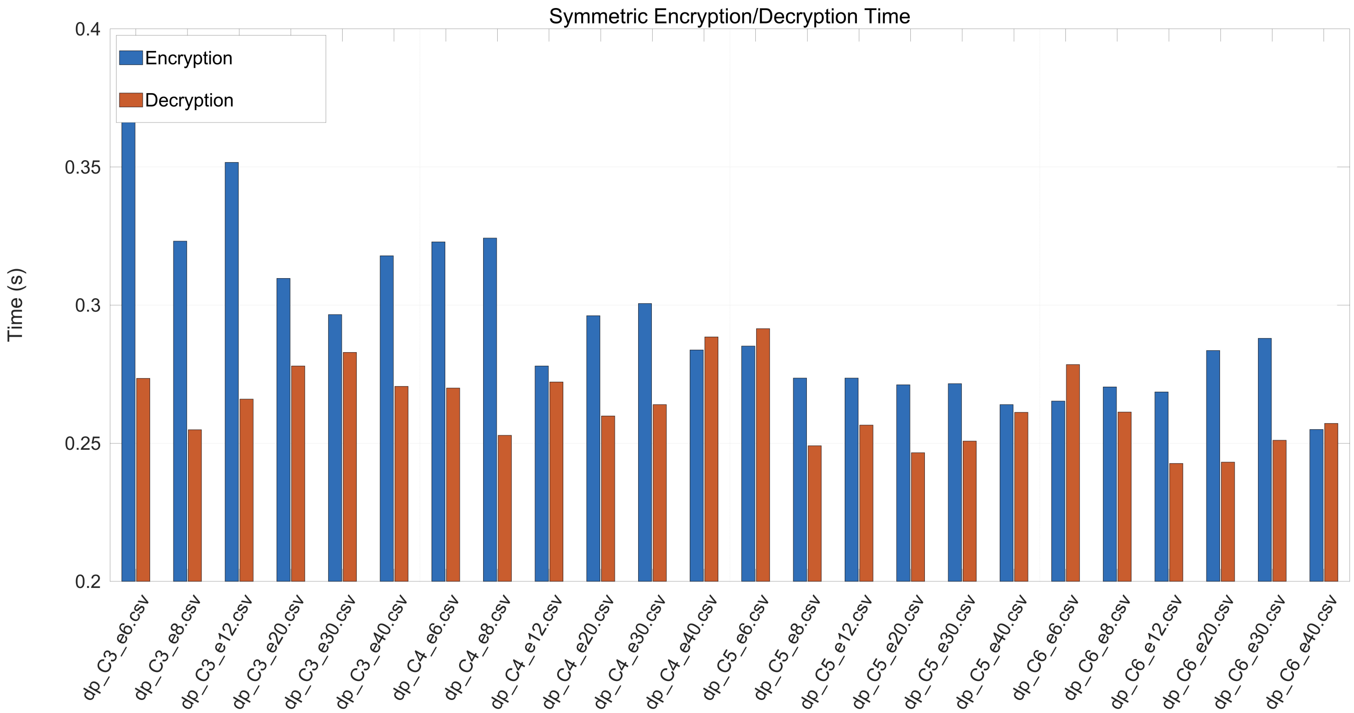

Fig. 3. Symmetric encryption and decryption time.

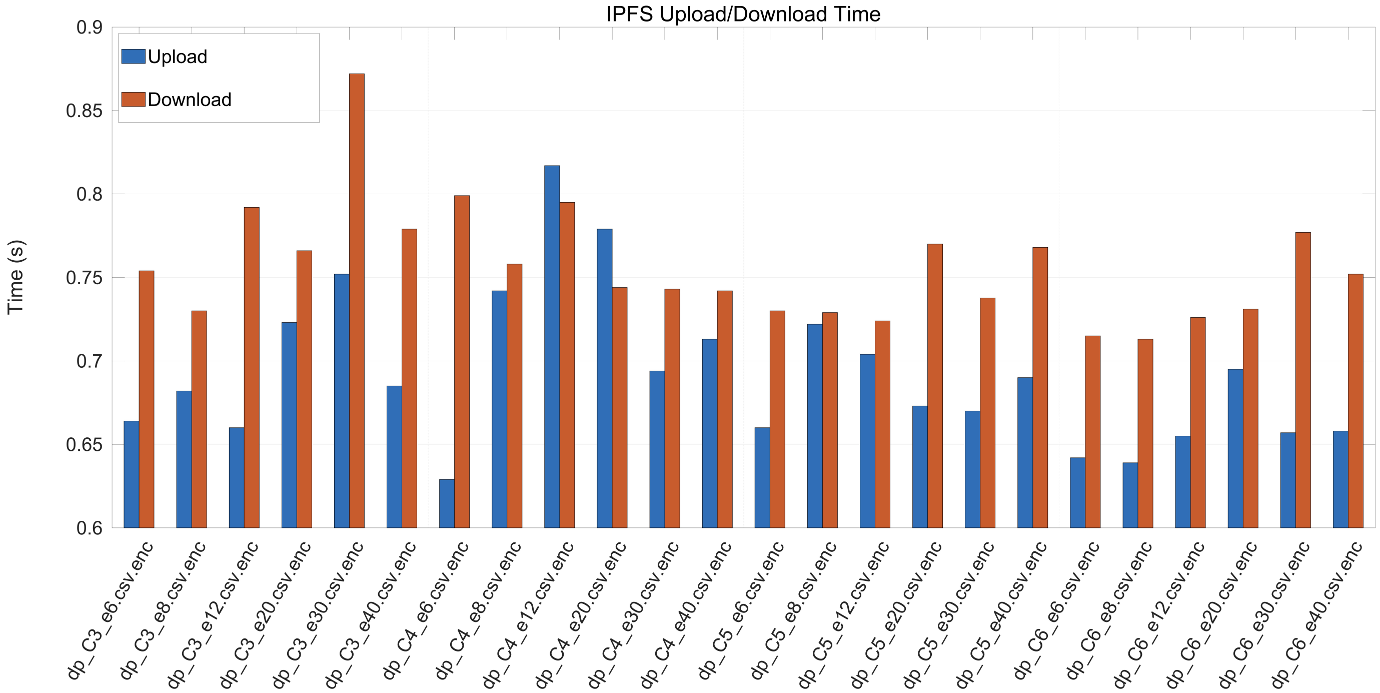

Fig. 4. IPFS upload and download time.

VIII. Conclusion

This study proposes a security-driven dual-chain framework for medical data sharing that structurally separates internal raw-data access from external privacy-preserving data release. By integrating multi-authority CP-ABE, blockchain-based governance, IPFS off-chain storage, and differential privacy under a unified security model, the proposed scheme achieves layered protection across confidentiality, integrity, access control, and statistical privacy dimensions. Unlike conventional blockchain-based storage solutions that primarily focus on data immutability, the proposed architecture emphasizes role-oriented domain separation and trust decoupling. The private hospital chain enforces fine-grained access control over encrypted medical records, while the consortium chain supervises privacy-budgeted data publication for external research institutions. This structural separation reduces trust concentration, isolates attack surfaces, and mitigates cross-domain inference risks. Experimental results demonstrate that even after differential privacy perturbation, the released data maintain sufficient utility for downstream analytical tasks, thereby achieving a practical balance between privacy protection and data usability.

Overall, the proposed framework provides a unified, auditable, and quantifiably private data-sharing model suitable for heterogeneous medical environments. Future work will further explore adaptive privacy budget allocation, traffic-obfuscation mechanisms, and comprehensive evaluation of blockchain consensus latency and scalability under large-scale deployment.

Acknowledgment

This work was supported by the National Natural Science Foundation of China [62502334, 62506255, 62472301, 62372318], National Research Project [2020YFC2006602], Opening Topic Fund of Big Data Intelligent Engineering Laboratory of Jiangsu Province [SDGC2157], Suzhou Blockchain Data Privacy Protection Innovation Application Laboratory, Provincial Key Laboratory for Computer Information Processing Technology, Soochow University [KJS2166].

References

[1] D. Kalita and K. B. Mirza, "Multivariate glucose forecasting using deep multihead attention layers inside neural basis expansion networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 5, pp. 3654–3663, 2025.

[2] L. Ge, A. N. McInnes, A. S. Widge, and K. K. Parhi, "Prediction of clinical response of transcranial magnetic stimulation treatment for major depressive disorder using hyperdimensional computing," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 5, pp. 3678–3686, 2025.

[3] Y. Shi, M. Wang, H. Liu, F. Zhao, A. Li, and X. Chen, "Mif: Multi-shot interactive fusion model for cancer survival prediction using pathological image and genomic data," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 5, pp. 3247–3258, 2025.

[4] H. Wu, J. Liu, T. Jiang, Q. Zou, S. Qi, Z. Cui, P. Tiwari, and Y. Ding, "Attentionmgt-dta: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism," *Neural Netw.*, vol. 169, no. C, p. 623–636, Jan. 2024. [Online]. Available: https://doi.org/10.1016/j.neunet.2023.11.018

[5] J. Liu, S. Guan, Q. Zou, H. Wu, P. Tiwari, and Y. Ding, "Amdgt: Attention aware multi-modal fusion using a dual graph transformer for drug–disease associations prediction," *Know.-Based Syst.*, vol. 284, no. C, Jan. 2024. [Online]. Available: https://doi.org/10.1016/j.knosys.2023.111329

[6] J. Liu, F. Hu, Q. Zou, P. Tiwari, H. Wu, and Y. Ding, "Drug repositioning by multi-aspect heterogeneous graph contrastive learning and positive-fusion negative sampling strategy," *Inf. Fusion*, vol. 112, no. C, Dec. 2024. [Online]. Available: https://doi.org/10.1016/j.inffus.2024.102563

[7] H. Wu, J. Liu, R. Zhang, Y. Lu, G. Cui, Z. Cui, and Y. Ding, "A review of deep learning methods for ligand based drug virtual screening," *Fundamental Research*, vol. 4, no. 4, pp. 715–737, mar 2024.

[8] J. A. Akinyele *et al.*, "Securing electronic medical records using attribute-based encryption on mobile devices," in *Proceedings of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*. ACM, 2011.

[9] Y. Zhou *et al.*, "Sistema de telemedicina con protección de privacidad basada en cp-abe," *Investigación Clínica*, vol. 60, no. 6, pp. 1615–1626, 2019.

[10] M. Chase, "Multi-authority attribute based encryption," in *Theory of Cryptography: 4th Theory of Cryptography Conference (TCC 2007), Amsterdam, the Netherlands, February 21–24, 2007, Proceedings 4*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[11] Z. Liu *et al.*, "Fully secure multi-authority ciphertext-policy attribute-based encryption without random oracles," in *Proceedings of the European Symposium on Research in Computer Security*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[12] Y. Jalghoum, A. Tahtamouni, and S. Mohammad, "An overview of blockchain technology in finance: A jordanian exploratory study," *Int. J. Crit. Infrastructures*, vol. 22, no. 6, 2026. [Online]. Available: https://doi.org/10.1504/ijcis.2026.10068995

[13] P. Jiang and L. Zhu, *Blockchain Technology - Cross-Chain Regulation and Privacy*. Springer, 2025. [Online]. Available: https://doi.org/10.1007/978-981-96-4395-0

[14] A. Ekblaw *et al.*, "A case study for blockchain in healthcare: "medrec" prototype for electronic health records and medical research data," in *Proceedings of the IEEE Open & Big Data Conference*, vol. 13, 2016.

[15] Z. Zheng *et al.*, "An overview on smart contracts: Challenges, advances and platforms," *Future Generation Computer Systems*, vol. 105, pp. 475–491, 2020.

[16] M. Bartoletti, R. Marchesin, and R. Zunino, "Scalable UTXO smart contracts via fine-grained distributed state," *Future Gener. Comput. Syst.*, vol. 175, p. 108023, 2026. [Online]. Available: https://doi.org/10.1016/j.future.2025.108023

[17] B. Chen *et al.*, "Bpvse: Publicly verifiable searchable encryption for cloud-assisted electronic health records," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 3171–3184, 2023.

[18] B. B. Gupta *et al.*, "Blockchain-assisted secure fine-grained searchable encryption for a cloud-based healthcare cyber-physical system," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 12, pp. 1877–1890, 2021.

[19] F. Wang *et al.*, "Lightweight and secure data sharing based on proxy re-encryption for blockchain-enabled industrial internet of things," *IEEE Internet of Things Journal*, vol. 11, no. 8, pp. 14 115–14 126, 2023.

[20] B. Jin *et al.*, "Bcas: Blockchain-based secure access and sharing scheme for ehr data," *Digital Communications and Networks*, 2024.

[21] Y. Chen *et al.*, "A blockchain-based medical data sharing mechanism with attribute-based access control and privacy protection," *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, p. 6685762, 2021.

[22] I. Sukhodolsky and S. Zapechnikov, "A blockchain-based access control system for cloud storage," in *Proceedings of the 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*. IEEE, 2018.

[23] C. Cai and Y. Liu, "Blockchain-based solution for secure sharing of medical data," *Computer Applications and Software*, vol. 41, no. 12, pp. 360–366, 2024.

[24] D. Zhu, Z. Zhou, Y. Li, H. Zhang, Y. Chen, Z. Zhao, and J. Zheng, "A survey of data security sharing," *Symmetry*, vol. 17, no. 8, p. 1259, 2025. [Online]. Available: https://doi.org/10.3390/sym17081259

[25] R. Rao, "Financial data sharing based on cloud computing security: the application and effectiveness of homomorphic encryption technology," in *11th IEEE Conference on Big Data Security on Cloud, BigDataSecurity 2025, New York City, NY, USA, May 9-11, 2025*. IEEE, 2025, pp. 161–166. [Online]. Available: https://doi.org/10.1109/BigDataSecurity66063.2025.00029

[26] D. B. Gayathri and D. Sangeetha, "An efficient blockchain-based incremental provable data possession for a secure electronic healthcare system," *Biomed. Signal Process. Control.*, vol. 112, p. 108919, 2026. [Online]. Available: https://doi.org/10.1016/j.bspc.2025.108919

[27] L. Liao, J. Zhao, Q. Zhang, and H. Fang, "A blockchain-enhanced trust-driven batch authentication scheme for secure vanets," *Ad Hoc Networks*, vol. 180, p. 104045, 2026. [Online]. Available: https://doi.org/10.1016/j.adhoc.2025.104045

[28] Q. Chen, C. Peng, and D. Xu, “Fuzzy password authentication key exchange protocol in universal composable framework for blockchain privacy protection,” *Comput. Stand. Interfaces*, vol. 95, p. 104032, 2026. [Online]. Available: https://doi.org/10.1016/j.csi.2025.104032

[29] M. Xie, Y. Yu, R. Chen, Y. Zhao, J. Ning, X. Yang, and Z. L. Jiang, “Multiuser data integrity auditing atop blockchain with secure user revocation for cognitive iot networks,” *Comput. Stand. Interfaces*, vol. 95, p. 104042, 2026. [Online]. Available: https://doi.org/10.1016/j.csi.2025.104042

[30] Y. Zhou, Y. Li, A. Sheng, and G. Qi, “Multi-pursuer single-evader privacy-preserving differential games,” *Appl. Math. Comput.*, vol. 508, p. 129612, 2026. [Online]. Available: https://doi.org/10.1016/j.amc.2025.129612

[31] T. Wang, T. Lin, Z. Liu, X. Xie, and S. Yao, “ADPGAN: adaptive differential privacy-preserving GAN for image privacy,” *Inf. Fusion*, vol. 125, p. 103515, 2026. [Online]. Available: https://doi.org/10.1016/j.inffus.2025.103515

[32] A. Zarei, “Differential privacy in secure multiparty computation and deep neural networks,” Ph.D. dissertation, Norwegian University of Science and Technology, Trondheim, Norway, 2025. [Online]. Available: https://hdl.handle.net/11250/3172984

[33] M. Adnan, M. H. Syed, A. Anjum, and S. Rehman, “A framework for privacy-preserving in iov using federated learning with differential privacy,” *IEEE Access*, vol. 13, pp. 13 507–13 521, 2025. [Online]. Available: https://doi.org/10.1109/ACCESS.2025.3526934

[34] I. Sahu, “Bilinear-inverse-mapper: Analytical solution and algorithm for inverse mapping of bilinear interpolation of quadrilaterals,” *Adv. Eng. Softw.*, vol. 208, p. 103975, 2025. [Online]. Available: https://doi.org/10.1016/j.advengsoft.2025.103975

[35] B. Huang, P. Huang, H. Yuan, and S. Liang, “A verifiable ranked ciphertext retrieval scheme based on bilinear mapping,” *Concurr. Comput. Pract. Exp.*, vol. 34, no. 12, 2022. [Online]. Available: https://doi.org/10.1002/cpe.5829

[36] Y. Chen *et al.*, “A bilinear map pairing based authentication scheme for smart grid communications: Pauth,” *IEEE Access*, vol. 7, pp. 22 633–22 643, 2019.

[37] A. Beimel, “Secret-sharing schemes: A survey,” in *Coding and Cryptology, IWCC 2011*, ser. Lecture Notes in Computer Science, Y. M. Chee, Z. Guo, S. Ling, F. Shao, Y. Tang, and C. Xing, Eds., vol. 6639. Berlin, Heidelberg: Springer, 2011, pp. 11–46.

[38] U. Gupta and H. Mahdavifar, “Security of linear secret sharing schemes with noisy side-channel leakage,” *IACR Cryptol. ePrint Arch.*, p. 987, 2025. [Online]. Available: https://eprint.iacr.org/2025/987

[39] A. Beimel, O. Farràs, Y. Mintz, and N. Peter, “Linear secret-sharing schemes for forbidden graph access structures,” *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 2083–2100, 2022. [Online]. Available: https://doi.org/10.1109/TIT.2021.3132917

[40] K. Blackwell and M. Wootters, “A characterization of optimal-rate linear homomorphic secret sharing schemes, and applications,” in *15th Innovations in Theoretical Computer Science Conference, ITCS 2024, January 30 to February 2, 2024, Berkeley, CA, USA*, ser. LIPIcs, V. Guruswami, Ed., vol. 287. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024, pp. 16:1–16:20. [Online]. Available: https://doi.org/10.4230/LIPIcs.ITCS.2024.16

[41] A. Beimel, A. Ben-Efraim, C. Padró, and I. Tyomkin, “Multi-linear secret-sharing schemes,” in *Theory of Cryptography*, ser. Lecture Notes in Computer Science, Y. Lindell, Ed., vol. 8349. Berlin, Heidelberg: Springer, 2014, pp. 394–418.

[42] T. Feneuil and M. Rivain, “Threshold linear secret sharing to the rescue of MPC-in-the-head,” Cryptology ePrint Archive, Paper 2022/1407, 2022. [Online]. Available: https://eprint.iacr.org/2022/1407