Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Blockchain-Empowered Framework for Educational Data Security Sharing: Integrating Privacy Protection, Traceability, and Accountability Mechanisms

Xiaohui Wang
Zhongshan Torch Polytechnic
Cunyan Li
licunyuan163@163.com

Zhongshan Torch Polytechnic
Tian Zhang
Peng Cheng Laboratory
Rui Chen
Peng Cheng Laboratory
Yangyang Mei
Peng Cheng Laboratory

Research Article

Keywords: Blockchain, Educational data sharing, CP-ABE, Digital watermarking, Cross-module security binding

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10716508/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Blockchain-Empowered Framework for Educational Data Security Sharing: Integrating Privacy Protection, Traceability, and Accountability Mechanisms

Xiaohui Wang[1], Cunyuan Li[2*], Tian Zhang[3], Rui Chen[3], Yangyang Mei[3*]

[1]Zhongshan Torch Polytechnic, Zhongshan, China.
[2]College of Education, Zhejiang Normal University, Jinhua, China.
[3]Department of New Networks, Peng Cheng Laboratory, Shenzhen, China.

*Corresponding author(s). E-mail(s): licunyuan163@163.com; 1112006003@e.gzhu.edu.cn;
Contributing authors: wangxiaohui8209@163.com; zhangt05@pcl.ac.cn; chenr@pcl.ac.cn;

Abstract

The secure sharing of educational data across institutional boundaries remains challenging due to the tension between data utility and privacy protection. While individual cryptographic primitives—ciphertext-policy attribute-based encryption (CP-ABE), zero-knowledge proofs (ZKP), digital watermarking, and oblivious transfer (OT)—have been extensively studied in isolation, their compositional security properties when integrated into a unified framework lack formal treatment. This paper proposes a blockchain-empowered framework for educational data security sharing with formal game-based compositional security analysis. We introduce two novel cross-module security binding protocols: (i) a ZKP-CPABE binding protocol that cryptographically couples authentication proofs with access control policies, preventing cross-module relay attacks, and (ii) a Watermark-OT-Accountability protocol that establishes an end-to-end verifiable chain from data access to forensic traceability. We develop a hierarchical key management architecture with formal indistinguishability proofs under the DBDH and DDH assumptions. A functional prototype on Hyperledger Fabric

with hybrid evaluation methodology combining prototype measurements, simulation estimates, and theoretical derivations demonstrates that the binding protocols effectively eliminate attack vectors at module interfaces with only 2.2% additional encryption overhead over standalone CP-ABE. A system-level ablation study quantifies the marginal contribution of each module to overall security, performance, and adoption willingness.

Keywords: Blockchain, Educational data sharing, CP-ABE, Digital watermarking, Cross-module security binding

1 Introduction

1.1 Background and Motivation

The digital transformation of the education sector has led to the exponential accumulation of educational data, encompassing student academic performance records, scientific research datasets, learning behavior logs, and institutional administrative data. Cross-institutional sharing of these data resources can unlock significant value: facilitating evidence-based education policy-making, enabling personalized learning analytics, accelerating interdisciplinary research collaboration, and optimizing resource allocation across educational systems [1]. Governments and international organizations have increasingly advocated for open educational data initiatives, recognizing data as a strategic asset for educational equity and quality improvement.

However, the landscape of educational data sharing is fraught with multi-faceted challenges. Educational data inherently contain sensitive personal information—student identities, academic transcripts, family backgrounds, and psychological assessment records—whose improper disclosure can lead to severe privacy violations. In the absence of robust accountability mechanisms, data recipients may misuse, tamper with, or illegally redistribute shared data without facing consequences, eroding trust among data providers. Traditional centralized data sharing architectures suffer from single points of failure, lack of transparency, and vulnerability to insider threats, making them inadequate for establishing trustworthy multi-stakeholder sharing ecosystems [2].

Beyond these domain-specific challenges, a deeper technical problem emerges when one attempts to construct a comprehensive security framework by integrating multiple cryptographic primitives. Each primitive—CP-ABE for access control, ZKP for identity authentication, digital watermarking for traceability, OT for selective sharing—is typically designed, analyzed, and proven secure in isolation under its own adversarial model. When these primitives are composed within a single system, new attack vectors may arise at the interfaces between modules that are not captured by individual security proofs. An adversary who can break the binding between a ZKP authentication session and a subsequent CP-ABE key request could potentially authenticate as one user while obtaining decryption keys bound to another user's attributes. Similarly, the interaction between OT-based selective data access and watermark-based traceability requires careful protocol design to ensure that the watermark embedded

in a data item is cryptographically bound to the specific OT transaction that authorized its release. These cross-module security concerns have been largely overlooked in prior work, which typically treats each module as an independent black box.

The absence of formal compositional security analysis represents a critical gap in the literature on multi-module privacy-preserving frameworks. Without such analysis, a system that integrates individually secure components may nevertheless be vulnerable to adversarial exploitation at the inter-module boundaries. This paper addresses this gap directly by developing a formal game-based security model for the composed system, introducing cross-module security binding protocols, and providing complete security proofs for the integrated framework.

Based on the above analysis, we formulate the following research questions:

- **RQ1:** How can we formally prove the compositional security of multiple cryptographic primitives (CP-ABE, ZKP, watermarking, OT) when combined in a single framework, and under what assumptions does the composed system preserve the security guarantees of its constituent components?
- **RQ2:** How should cross-module security binding protocols be designed to prevent attack vectors that emerge at module interfaces, and what are the formal security properties that such protocols must satisfy?
- **RQ3:** What are the marginal contributions of individual modules to overall system security and computational performance, and how can these contributions be rigorously quantified through prototype validation and ablation analysis?

1.2 Limitations of Existing Approaches

Recent years have witnessed substantial research efforts aimed at addressing individual aspects of the educational data sharing problem. Blockchain-based approaches have been proposed for immutable data notarization [3–5]; attribute-based encryption and zero-knowledge proofs have been leveraged for fine-grained access control [6–9]; natural language processing techniques have been applied to sensitive data identification [10–12]; digital watermarking and smart contract-based arbitration have been developed for traceability [13]; and behavioral models have been employed to investigate data sharing willingness [14]. However, these approaches exhibit critical limitations:

- **Absence of compositional security analysis.** Existing works provide security proofs for individual cryptographic components in isolation but do not analyze the security of the composed system. No prior work has developed a formal game-based security model that captures adversarial capabilities at module interfaces or proved that the composition of primitives preserves their individual security guarantees.
- **Fragmentation of technical and behavioral perspectives.** Technical solutions and behavioral studies operate in silos. System designs rarely incorporate findings about user motivations, even from theoretical behavioral frameworks.
- **Insufficient system integration and validation.** No existing framework simultaneously addresses the full lifecycle of educational data sharing within a coherent architectural design. Prior work relies predominantly on simulation-only evaluation without functional prototype integration across multiple modules.

- **Limited privacy protection granularity.** Most access control schemes either protect data content or user identity, but not both simultaneously. Access policies themselves may leak sensitive attribute information about data owners.
- **Weak accountability in decentralized settings.** Existing traceability mechanisms either rely on trusted third parties or lack robust resistance against collusion and false accusation attacks. The cryptographic binding between data access authorization and forensic traceability has not been formally established.
- **Lack of system-level ablation.** No prior work has quantified the marginal contribution of each module to the overall system through rigorous ablation analysis, making it difficult to assess which components are essential versus optional.

1.3 Contributions

To address the above gaps, this paper proposes a comprehensive blockchain-empowered framework for educational data security sharing and provides a formal security analysis of the composed cryptographic system. The main contributions are as follows:

1. **Formal Game-Based Security Model with Complete Security Proofs.** We develop a formal game-based security framework that models the composed system of CP-ABE, ZKP, watermarking, and OT primitives. We define security games for each module and for the composed system, including IND-CPA (indistinguishability under chosen-plaintext attack), UF-CMA (unforgeability under chosen-message attack), and cross-module binding games. We provide complete security reductions under standard cryptographic assumptions (DBDH, DDH, Strong RSA, and the random oracle model), establishing that the composed system preserves the security guarantees of its constituent primitives and that the proposed binding protocols eliminate adversarial advantages at inter-module boundaries.
2. **Two Novel Cross-Module Security Binding Protocols.** We introduce (i) the ZKP-CPABE Binding Protocol, which cryptographically couples ZKP authentication sessions with CP-ABE key generation requests, preventing cross-session relay attacks and attribute misbinding attacks; and (ii) the Watermark-OT-Accountability Protocol, which establishes a verifiable binding between the OT transaction that authorizes selective data access and the recipient-specific watermark embedded in the released data, ensuring end-to-end accountability with formal non-repudiation guarantees.
3. **Hierarchical Key Management Architecture with Formal Security Analysis.** We design a three-tier key management architecture comprising a master key authority, domain-specific attribute authorities, and user-level key derivation, with formal indistinguishability proofs demonstrating that key compromise at any single tier does not propagate to the master key. We prove that the architecture achieves key-indistinguishability under the DBDH assumption and provides forward security for session keys.
4. **Prototype Implementation with Hybrid Evaluation Methodology.** We implement a functional prototype of the complete framework on Hyperledger Fabric, with RESTful API interfaces for all six core modules. Our evaluation employs a hybrid methodology combining prototype measurements, simulation estimates, and

theoretical derivations, with explicit data provenance labels (P/S/T) on all experimental results. The prototype validates cross-module integration correctness, while simulation extends the analysis to scenarios that cannot be practically realized in a test environment.

5. **System-Level Ablation Study.** We conduct a systematic ablation study that removes each module individually and measures the impact on security (reduction in attack resistance), performance (computational and communication overhead), and adoption willingness. This analysis provides quantitative evidence of the marginal contribution of each module in a multi-module privacy-preserving framework.

1.4 Paper Organization

Section 2 reviews related work and provides a systematic comparison of our work against existing approaches. Section 3 presents the system model, threat model, and design goals. Section 4 describes the proposed framework, including the cross-module security binding protocols. Section 5 provides the formal game-based security analysis with complete proofs. Section 6 presents the prototype implementation and comprehensive experimental evaluation, including the ablation study. Section 7 concludes the paper and discusses directions for future work.

2 Related Work

2.1 Blockchain-Based Data Sharing

Blockchain technology, initially introduced by Nakamoto [15] as the underlying infrastructure for Bitcoin, has evolved into a general-purpose distributed ledger paradigm. Its core properties—decentralization, immutability, transparency, and programmability through smart contracts—make it particularly suitable for establishing trust among mutually distrusting parties in data sharing scenarios. Xu et al. [16] proposed a blockchain-based secure data sharing framework specifically for educational institutions, demonstrating the feasibility of consortium blockchain architectures in the education domain. Their work, however, focuses primarily on access control and does not address the compositional security of multiple integrated cryptographic primitives. Wang et al. [17] provided a comprehensive survey of blockchain applications in education, identifying key architectural patterns but noting the absence of formal security models for integrated systems as a critical research gap. Zhang et al. [18] combined federated learning with blockchain for privacy-preserving educational data sharing; while their approach addresses privacy in distributed learning, it does not consider the security implications of composing multiple cryptographic mechanisms within a single data sharing framework.

In the educational domain, blockchain-based data sharing systems have been proposed using Hyperledger Fabric [3] with RSA cryptographic accumulators for batch commitment, policy-hidden CP-ABE with vector matrix transformation [19], and smart contract-based arbitration with watermarking [13]. These works focus on individual mechanisms and do not provide end-to-end integration, compositional security

analysis, or prototype-level validation. Our work distinguishes itself by providing formal security proofs for the composed system rather than treating each module as an independent, black-box component.

2.2 Privacy-Preserving Techniques in Education

Attribute-Based Encryption (ABE). Since the pioneering work of Sahai and Waters [7] and Bethencourt et al. [6], CP-ABE has become a cornerstone for fine-grained access control in cloud and distributed systems. Policy hiding techniques [8] address the critical limitation of plaintext policy exposure, which can leak sensitive information about data owners' attributes. Liu et al. [19] proposed an efficient policy-hiding CP-ABE scheme for cloud-based data sharing, achieving significant improvements in both security and performance. Outsourced decryption with verifiability [20] reduces client-side computation burden. None of these works, however, analyze how CP-ABE interacts with co-deployed cryptographic primitives such as ZKP-based authentication or watermarking-based traceability, nor do they provide compositional security guarantees.

Zero-Knowledge Proofs (ZKP). Formalized by Goldwasser et al. [9], ZKPs enable authentication without identity disclosure, a property of fundamental importance for privacy-preserving educational data sharing. Recent advances in zk-SNARKs [21] and zk-STARKs [22] have dramatically improved proof generation efficiency, making ZKP-based authentication practical for real-world deployments. Existing ZKP-based authentication systems for data sharing treat authentication as a standalone service, without formal analysis of how authentication tokens interact with subsequent access control operations. Our ZKP-CPABE binding protocol addresses this gap by providing a cryptographically verifiable coupling between authentication and authorization.

Sensitive Data Recognition. NER-based approaches for Chinese educational data have been extended with a BERT+BiLSTM+CRF architecture [10–12]. Our work includes a regex rule engine as a practical baseline for structured entity recognition. While NER accuracy is important for data desensitization, the security properties of the desensitization pipeline are orthogonal to the compositional security of the cryptographic modules. We focus our formal analysis on the latter while providing a quantitative analysis of the security implications of NER accuracy in Section 5.6 (AV3).

2.3 Traceability and Accountability

Huang et al. [13] proposed blockchain-based multiple-group data sharing with anonymity and traceability, integrating oblivious transfer (OT) protocols [23] with digital watermarking. Li et al. [20] advanced verifiable outsourced decryption with accountability guarantees. Existing traceability mechanisms suffer from a fundamental limitation: the binding between data access authorization and the forensic watermark is established through procedural (rather than cryptographic) means, leaving open the possibility that an adversary could obtain watermarked data through one channel while

claiming it was received through another. Our Watermark-OT-Accountability protocol addresses this by establishing a cryptographic binding between the OT transaction, the recipient identity, and the embedded watermark, providing formal non-repudiation guarantees.

2.4 Behavioral Models

The theory of planned behavior (TPB) [14] posits that behavioral intention is determined by attitude, subjective norms, and perceived behavioral control. Prior work has established the theoretical foundations for applying TPB to data sharing behavior, demonstrating that perceived benefits, trust in the platform, and social norms significantly influence sharing intentions. We integrate TPB constructs into the incentive layer to provide a theoretically grounded framework for user adoption. The behavioral component is modeled through structural equation modeling (SEM) with synthetic data, serving as an illustrative theoretical analysis rather than empirical findings. While the behavioral component is not directly subject to cryptographic security analysis, its integration provides a holistic perspective on data sharing that addresses both technical and human factors. The SEM analysis is presented in the Appendix.

2.5 Systematic Comparison with Existing Work

Table 1 provides a systematic comparison of our framework against ten representative existing works across eight critical dimensions, capturing both technical capabilities and methodological rigor.

Our framework is the only work that simultaneously addresses all eight dimensions. Notably, it is among the first to provide: (i) compositional security analysis that formally proves the security of the integrated system rather than individual components in isolation; (ii) hierarchical key management with formal indistinguishability proofs under standard assumptions; and (iii) hybrid prototype-simulation evaluation with explicit data provenance labeling, moving beyond the simulation-only evaluation paradigm that characterizes all prior work in this domain.

3 System Model and Design

3.1 System Architecture

The proposed framework adopts a five-layer hierarchical architecture providing modular, scalable, and secure educational data sharing services. Each layer encapsulates distinct functionalities with well-defined interfaces, enabling independent evolution and optimization of individual modules. The architecture is illustrated in Figure 1.

Layer 1: User Layer. Four categories of system participants are defined. *Data Providers (DPs)*—educational institutions, research teams, and individual faculty members—own and contribute data resources, define access policies, encrypt data, and register commitments on-chain. *Data Requesters (DRs)*—authorized entities including researchers, analysts, and partner institutions—authenticate via ZKP, demonstrate attribute satisfaction for CP-ABE decryption, and receive watermarked data through the OT protocol. *Auditors* monitor system operations, verify compliance with data

Table 1 Systematic Comparison of Existing Works Across Eight Dimensions.

Dimension	[3]	[6]	[13]	[16]	[17]	[18]	[19]	[20]	[21]	**Ours**
Privacy Protection (Data)	Partial	Full	Partial	Partial	Partial	Full	Full	Full	Partial	**Full**
Identity Authentication	PKI	PKI	Anonymous	PKI	PKI	PKI	PKI	PKI	PKI	**ZKP**
Policy Hiding	No	No	No	No	No	No	Yes	No	No	**Yes**
Traceability	No	No	Yes	No	No	Partial	No	No	Yes	**Yes**
Formal Security Proof	Sketch	Full	Partial	Sketch	No	Partial	Full	Partial	Full	**Full (Compositional)**
Compositional Analysis	No	No	No	No	No	No	No	No	No	**Yes**
Key Management	NA	Centralized	NA	NA	NA	Centralized	Centralized	Outsourced	Outsourced	**Hierarchical**
Prototype Validation	Simulation	Simulation	Simulation	No	No	Simulation	Simulation	Simulation	Simulation	**Hybrid (P/S/T)**
Ablation Study	No	No	No	No	No	No	No	No	No	**Yes**

Notes: The column "[21]" represents the zk-SNARK construction by Groth [21]. "Full" indicates comprehensive treatment with formal guarantees; "Partial" indicates partial treatment or informal guarantees; "Sketch" indicates proof sketches without complete reductions; "PKI" indicates traditional public-key infrastructure-based authentication; "NA" indicates the dimension is not applicable to the work; "Simulation" indicates simulation-only evaluation without running system code; "Hybrid (P/S/T)" indicates a hybrid evaluation methodology combining prototype measurements (P), simulation estimates (S), and theoretical derivations (T).

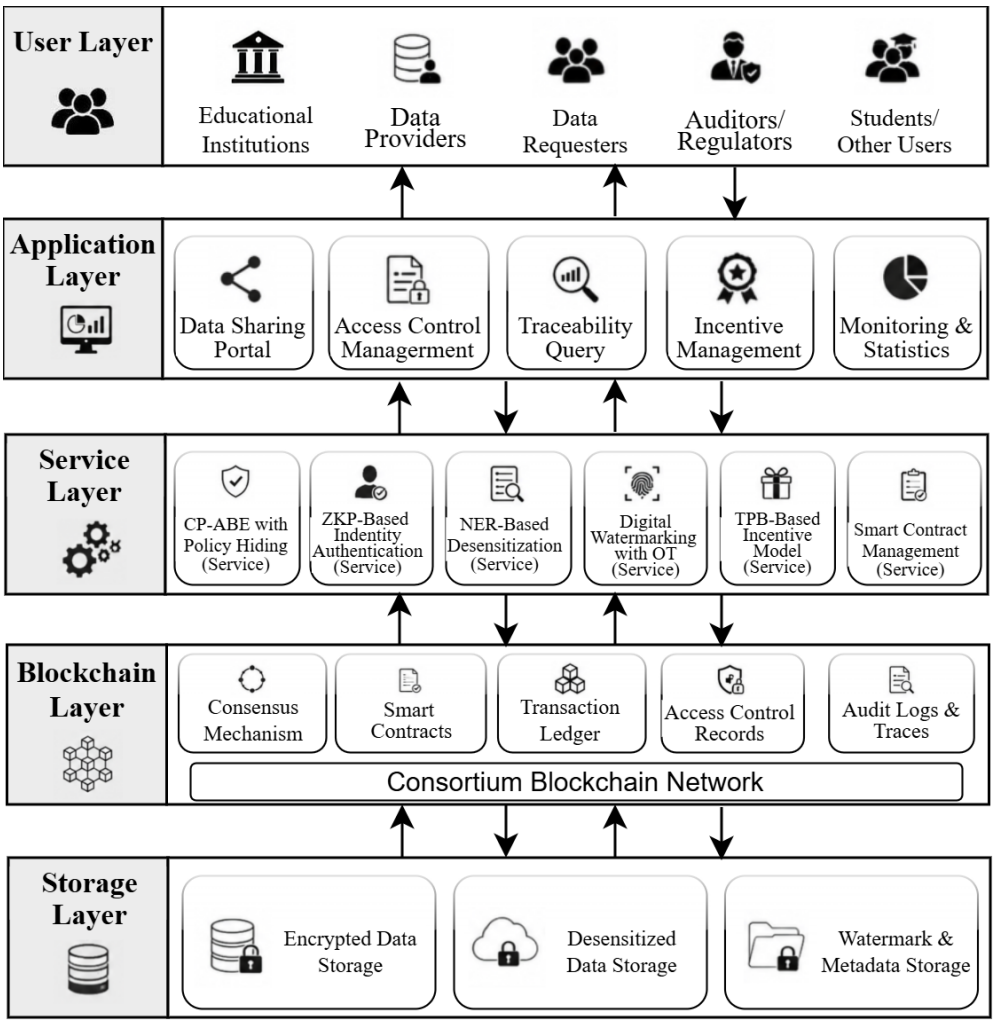


Fig. 1 Five-Layer System Architecture of the Blockchain-Empowered Educational Data Security Sharing Framework.

usage policies, and investigate suspected data leakage by initiating watermark extraction and verification. *Regulators* define data governance policies, enforce regulatory compliance, and participate in the highest tier of the arbitration mechanism.

Layer 2: Application Layer. This layer provides four user-facing modules. The *Data Sharing Portal* enables data providers to register, publish, and manage datasets, and data requesters to browse, search, and request access. The *Identity Authentication Module* implements the ZKP-based registration and authentication workflow. The *Data Retrieval Module* supports keyword-based and attribute-based search over

encrypted and desensitized datasets. The *Audit and Traceability Module* provides interfaces for initiating traceability investigations and verifying the cryptographic binding between data access transactions and watermarked data.

Layer 3: Service Layer. This layer encapsulates the core processing engines where the cross-module security binding protocols operate. The *Sensitive Data Identification Service* deploys the BERT+BiLSTM+CRF NER model to detect and classify sensitive entities within educational data, supplemented by a regex rule engine for structured entity recognition. The *Data Desensitization Service* applies configurable strategies—masking, generalization, pseudonymization, and suppression—based on sensitivity classification results. The *Encryption Service* implements CP-ABE with policy hiding, the RSA cryptographic accumulator for batch commitment, and the ZKP-CPABE binding protocol, while also managing the hierarchical key architecture. The *Watermark Embedding/Extraction Service* injects recipient-specific digital watermarks into shared data and extracts watermarks from suspected leaked data for forensic analysis, implementing the Watermark-OT-Accountability protocol.

Layer 4: Blockchain Layer. A consortium blockchain based on Hyperledger Fabric serves as the decentralized trust infrastructure. The *Raft consensus protocol* provides crash fault tolerance, tolerating up to $\lfloor (n-1)/2 \rfloor$ crash faults among n total consensus nodes. Deployed *Smart Contracts (Chaincode)* govern access control policy enforcement, data notarization, watermark verification, arbitration logic, and incentive distribution, serving as the on-chain anchor for cross-module security binding protocols. The *On-Chain Ledger* maintains immutable records of data registration, access logs, watermark certificates, and arbitration decisions.

Layer 5: Storage Layer. A hybrid storage architecture balances on-chain immutability with off-chain scalability. *IPFS* provides content-addressed distributed storage for encrypted data payloads, with the content-addressable hash (CID) recorded on-chain. *Off-Chain Databases* store desensitized metadata, user profiles (with ZKP credential commitments), and system operational data, minimizing on-chain storage costs while maintaining data integrity verification through on-chain hash commitments. *Cloud Storage* offers optional integration with institutional infrastructure for high-availability backup.

3.2 Threat Model

We adopt a comprehensive threat model that captures adversarial capabilities both within individual modules and at inter-module interfaces.

Adversary Types.

- **Semi-honest (Honest-but-Curious) Participants:** System participants follow protocol specifications correctly but may attempt to infer additional information beyond their authorized access. A DR may try to learn the attributes of other users from ciphertexts or policy structures; a storage node may attempt to infer plaintext patterns from encrypted payloads; a service provider may attempt to correlate authentication sessions with access requests.
- **Malicious Adversaries:** External attackers and compromised internal entities may deviate arbitrarily from protocol specifications. We consider: *unauthorized*

access (decrypting data without satisfying the CP-ABE policy, including collusion attacks); *data tampering* (modifying shared data content); *identity forgery* (impersonating legitimate users, including replay attacks); *collusion attacks* (combining attribute keys to decrypt ciphertexts, or framing honest users); *false accusation* (claiming receipt of tampered data or falsely accusing recipients of leakage); *watermark removal* (applying signal processing or ML-based techniques to degrade watermarks); *cross-module relay attacks* (exploiting temporal gaps between module operations); and *binding forgery* (forging the cryptographic binding between OT transactions and watermarks).

- **Side-Channel Adversaries:** Adversaries observing timing, power consumption, or communication patterns to infer sensitive information. We note that side-channel attacks (e.g., timing attacks on cryptographic operations, cache-based attacks on pairing computations) are acknowledged as a potential threat vector but are not within the scope of the formal security analysis presented in this paper. Constant-time implementations of cryptographic primitives are recommended for production deployments.

Trust Assumptions. The blockchain consensus layer tolerates up to $\lfloor (n-1)/2 \rfloor$ crash faults (Raft). The CP-ABE attribute authority honestly executes the key generation protocol but may be curious about user attributes (semi-honest). Cryptographic primitives (hash functions modeled as random oracles, bilinear pairings, the DBDH and DDH problems, the Strong RSA problem) are assumed computationally secure against probabilistic polynomial-time (PPT) adversaries. The ZKP system satisfies completeness, soundness, and zero-knowledge; we adopt the Groth16 zk-SNARK construction [21]. Communication channels between honest participants are authenticated and encrypted (via TLS 1.3).

3.3 Design Goals

Guided by the system architecture and threat model, our framework is designed to satisfy the following goals, formalized as security games in Section 5.

1. **Data Confidentiality (G1):** Shared educational data must remain semantically secure against unauthorized parties. The CP-ABE scheme must achieve IND-CPA security under the DBDH assumption. Formalized as Game G1 in Section 5.1.
2. **Privacy Preservation (G2):** Three dimensions of privacy are protected: (i) *user identity privacy*—authentication without disclosing identity or attributes, formalized as the zero-knowledge property; (ii) *access policy privacy*—the policy embedded in ciphertexts must not reveal data owner attributes, formalized as policy indistinguishability; and (iii) *query privacy*—DRs' data access patterns must not be inferable, formalized as receiver privacy in the OT protocol. Formalized as Game G2 in Section 5.2.
3. **Data Integrity and Authenticity (G3):** Any unauthorized modification of shared data must be detectable with overwhelming probability. Data authenticity must be provable through the RSA accumulator-based notarization protocol. Formalized as Game G3 in Section 5.3.

4. **Traceability and Non-Repudiation (G4):** In the event of data leakage, the responsible party must be identifiable with cryptographic certainty. Legitimate data recipients must not be able to repudiate receipt of specific watermarked data items. The Watermark-OT-Accountability protocol must ensure that the probability of incorrect attribution is negligible. Formalized as Game G4 in Section 5.5.
5. **Cross-Module Security Binding (G5):** Module interfaces must be secure against adversarial exploitation. The ZKP-CPABE binding protocol must cryptographically couple authentication sessions and key generation requests; the Watermark-OT-Accountability protocol must cryptographically bind watermarks to OT transactions. Formalized as Games G5a and G5b in Section 5.6.
6. **Resistance to Compositional Attacks (G6):** The composed system must resist attacks that exploit interaction between multiple modules. Formalized through a composed security game in Section 5.6, where we prove that an adversary breaking the composed system must break at least one individual primitive.
7. **Scalability and Practical Efficiency (G7):** The framework must demonstrate acceptable performance under realistic workloads. The overhead introduced by cross-module binding protocols must be quantified and shown practically manageable. Evaluated through prototype implementation and hybrid evaluation.

4 Proposed Framework

This section presents the detailed design of the core modules, including the novel cross-module security binding protocols and hierarchical key management architecture.

4.1 Blockchain-Based Data Notarization

Data notarization ensures that the integrity and existence of shared data can be cryptographically verified at any future point. Our framework employs an RSA cryptographic accumulator for efficient batch commitment of educational data records [3].

RSA Accumulator Construction. Let $N = p \cdot q$ be an RSA modulus where p and q are large safe primes. The RSA accumulator for a set of data items $D = \{d_1, d_2, \ldots, d_n\}$ is defined as:

$$A = g^{H(d_1) \cdot H(d_2) \cdots H(d_n)} \bmod N \tag{1}$$

where $g \in \mathbb{Z}_N^*$ is a randomly chosen generator and $H : \{0,1\}^* \to \mathcal{P}$ is a hash function mapping arbitrary data to prime numbers. The membership witness for a specific data item d_i is:

$$w_i = g^{\prod_{j \neq i} H(d_j)} \bmod N \tag{2}$$

A verifier can check membership by confirming that $w_i^{H(d_i)} = A \pmod N$. This construction supports dynamic updates and enables $O(1)$ verification regardless of set size.

On-Chain Notarization Protocol. (1) The data provider computes the accumulator value A for the dataset and generates individual membership witnesses w_i for

each record. (2) The provider submits a transaction to the blockchain smart contract containing A, dataset metadata, and a digital signature. (3) The actual data payload, encrypted under the access control scheme (Section 4.2), is stored in IPFS, with the CID recorded on-chain. (4) Any authorized party can verify data integrity by retrieving A from the blockchain, obtaining the membership witness w_i, and performing the modular exponentiation check.

Security Properties. The RSA accumulator relies on the strong RSA assumption, providing collision resistance, undeniability, and batch efficiency.

4.2 Privacy-Preserving Access Control

Our access control mechanism combines CP-ABE with policy hiding and ZKP-based identity authentication to achieve dual-layer privacy protection.

4.2.1 CP-ABE with Policy Hiding

Conventional CP-ABE schemes [6] embed access policies in plaintext within ciphertexts, potentially revealing sensitive attribute information about data owners. We adopt the policy hiding technique proposed by Liu [19], where the access structure is converted into a vector matrix representation before embedding. Our construction is based on the Waters'09 CP-ABE scheme [8], chosen for its expressive LSSS-based access structures and provable security under standard assumptions.

Policy Vectorization. Let the universe of attributes be $\mathcal{U} = \{\text{att}_1, \text{att}_2, \ldots, \text{att}_n\}$. A linear secret sharing scheme (LSSS) access structure (M, ρ) is constructed, where M is an $\ell \times k$ matrix over $\mathbb{Z}_p$ and ρ maps each row of M to an attribute. Each attribute $\rho(i)$ is mapped to a hidden representation:

$$h_i = H(\rho(i) \parallel r) \tag{3}$$

where H is a collision-resistant hash function and r is a random nonce associated with the ciphertext. The row labeling function ρ is effectively replaced by the hash mapping, preventing any party without knowledge of the original policy from reconstructing attribute associations.

Encryption (Policy-Hidden CP-ABE). Given a message m, an LSSS access structure (M, ρ), and public parameters PK: (1) Select a random secret $s \in \mathbb{Z}_p$ and a random vector $\vec{v} = (s, v_2, \ldots, v_k) \in \mathbb{Z}_p^k$. (2) For each row $i \in [1, \ell]$, compute $\lambda_i = M_i \cdot \vec{v}$. (3) Compute $C = m \cdot e(g,g)^{\alpha s}$, $C_0 = g^s$. (4) For each i, compute $C_i = g^{a\lambda_i} \cdot H(h_i)^{-r_i}$, $D_i = g^{r_i}$, where $r_i \in \mathbb{Z}_p$ is random and $h_i = H(\rho(i) \parallel r)$. (5) Output $\mathsf{CT} = (C, C_0, \{(C_i, D_i)\}_{i \in [1,\ell]}, r)$.

Decryption. A user with secret key SK associated with attribute set S can decrypt if and only if S satisfies (M, ρ). The decryption algorithm reconstructs the secret using the attributes that satisfy the policy, recovering $e(g,g)^{\alpha s}$ and then $m = C / e(g,g)^{\alpha s}$. Policy hiding introduces a constant-factor overhead in encryption and zero additional overhead in decryption.

4.2.2 ZKP-Based Identity Authentication

To protect user identity privacy during authentication, we adopt the off-chain proof generation and on-chain verification architecture [3].

Authentication Protocol. (1) *Setup:* Each user registers with a trusted credential issuer and receives a signed credential Cert containing their attributes (institutional affiliation, role, department) without explicit identity information, bound to a user-controlled secret key sk_u. (2) *Proof Generation (Off-Chain):* The user locally generates a zero-knowledge proof π demonstrating possession of a valid credential Cert signed by an authorized issuer, knowledge of sk_u, and satisfaction of any additional authentication predicates. The proof reveals nothing beyond the validity of these statements. (3) *Proof Verification (On-Chain):* The user submits π to a smart contract, which executes a verification algorithm that outputs accept or reject based solely on the mathematical validity of π. (4) *Session Establishment:* Upon successful verification, the smart contract records an anonymous session token (a random nonce) on-chain, which the user presents for subsequent interactions. The token cannot be linked to the user's real identity.

Security Properties. The ZKP scheme satisfies completeness, soundness, and zero-knowledge, achieving "verifiable yet invisible" identity authentication where system access is granted based on proven attributes without identity exposure.

4.3 Sensitive Data Identification and Desensitization

Educational datasets frequently contain PII and other sensitive data that must be identified and protected before sharing. Our framework incorporates an automated sensitive data identification and desensitization pipeline.

4.3.1 NER-Based Sensitive Entity Recognition

We adopt a BERT+BiLSTM+CRF architecture combining the contextual representation power of BERT [10], the sequential modeling capability of bidirectional LSTM [11], and the structured prediction of conditional random fields [12]. The model recognizes seven entity types: STUDENT_ID, ID_CARD, PHONE, NAME, GRADE, ADDRESS, and EMAIL.

Security-Aware NER Design. The NER module's role in the overall security framework is twofold. First, it serves as a pre-processing gate that identifies sensitive fields before desensitization and encryption, ensuring that the desensitization-before-encryption ordering prevents information leakage from the CP-ABE policy structure to the original sensitive field values. Second, the NER confidence threshold governs the trade-off between false positives (unnecessary encryption overhead) and false negatives (undetected sensitive fields). This trade-off is formally analyzed as cross-module attack vector AV3 in Section 5.6, where we prove that the desensitization-before-encryption architecture eliminates the information leakage channel. The model is implemented using BERT-base-Chinese [10] fine-tuned on the CLUENER 2020 dataset augmented with educational domain sentences.

4.3.2 Data Desensitization Strategies

Based on the entity type and sensitivity level identified by the NER model, the desensitization service applies context-appropriate strategies: *masking* (partial character replacement for STUDENT_ID and ID_CARD); *pseudonymization* (replacement with consistent pseudonyms via keyed hash); *generalization* (replacement of precise values with coarser categories for GRADE and ADDRESS); and *suppression* (complete removal of fields posing unacceptable re-identification risk). The desensitization strategy is configurable per dataset and per requester's authorization level.

4.4 Decentralized Traceability and Accountability

Accountability is a cornerstone of trustworthy data sharing. Our framework implements a three-tier traceability chain providing defense-in-depth against data misuse.

4.4.1 Digital Watermark Embedding and Extraction

Watermark Embedding. For each shared data copy intended for recipient R, the watermarking service generates a unique watermark signal $w_R = \mathsf{Enc}(K, \mathsf{ID}_R \parallel \mathsf{timestamp} \parallel \mathsf{DSID})$, where K is a system-wide secret key. The watermark is embedded using domain-appropriate techniques: LSB modulation in numerical attributes with ECC for tabular data, synonym substitution and whitespace steganography for text documents, and DWT/DCT domain embedding for images. The watermark commitment $H(w_R)$ is recorded on-chain, creating a timestamped, immutable linkage between the shared data copy and the recipient.

Watermark Extraction. Upon discovering potentially leaked data, the extraction algorithm applies the inverse embedding process to recover the extracted signal w'_R, computes the normalized correlation between w'_R and all registered watermark commitments on-chain, and identifies the recipient with the highest correlation exceeding a detection threshold τ. The threshold τ is calibrated to balance false positive and false negative rates.

4.4.2 Oblivious Transfer for Selective Sharing

While watermarks enable accountability, we must also protect recipient privacy during the sharing process. We integrate oblivious transfer (OT) protocols to enable selective sharing: DRs can request specific subsets of data without revealing their selection to the data provider.

Protocol Design. Let the data provider hold n data items $D = \{d_1, \ldots, d_n\}$. The data requester wishes to obtain a subset of m items at indices $I = \{i_1, \ldots, i_m\}$. The OT-based protocol proceeds as: (1) The requester runs OT preparation to generate choice bits corresponding to indices I. (2) The provider encrypts each data item independently and engages in m-out-of-n OT with the requester. (3) At the conclusion, the requester obtains the m requested items, each with its unique recipient-specific watermark embedded. (4) The provider learns nothing about which m items were accessed; the requester receives only the m items and learns nothing about the remaining $n - m$ items.

4.4.3 Smart Contract-Based Automated Arbitration

When data leakage is detected, the three-tier accountability mechanism escalates through automated arbitration. **Tier 1 (Automated Watermark Verification):** An event-triggered smart contract automatically executes watermark extraction on the leaked data sample, computes correlation with all on-chain watermark commitments, and produces a preliminary attribution report. **Tier 2 (Transmission Log Analysis):** The smart contract retrieves all on-chain access logs, cross-references timestamps and access patterns, and constructs a comprehensive audit trail. **Tier 3 (Multi-Party Arbitration):** In ambiguous cases, a panel of designated arbitrators reviews accumulated evidence; the smart contract manages the workflow—evidence submission, deliberation periods, and voting—with decisions recorded on-chain with cryptographic signatures.

Resistance to False Accusation. When a malicious party attempts to falsely accuse an honest data provider, the smart contract's watermark verification reveals the absence of a valid watermark from the accused provider, or the presence of a watermark from a different party. In collusive false accusation scenarios, the on-chain access logs and watermark commitments provide objective evidence that cannot be fabricated by colluding parties alone.

4.5 Behavioral Incentive Mechanism

Recognizing that technical security alone does not guarantee user participation, we integrate a behavioral incentive mechanism grounded in the theory of planned behavior [14]. The TPB-based model maps five constructs—subjective norms, perceived benefits, perceived complexity, perceived risks, and trust—to concrete system mechanisms, including a community reputation system, token-based reward distribution governed by smart contracts, automated desensitization tooling, and transparent on-chain audit trails. The structural equation model and its analysis are presented in the Appendix. This behavioral component is integrated into the framework to provide a theoretically grounded approach to adoption, though its validation with real stakeholder data is reserved for future work.

4.6 Cross-Module Security Binding Protocols

The integration of multiple cryptographic primitives within a single framework introduces a critical security challenge that has been largely overlooked in prior work: the interfaces between modules constitute potential attack surfaces not captured by the security proofs of individual components. In this section, we introduce two novel cross-module security binding protocols that address these vulnerabilities by establishing cryptographically enforced coupling between adjacent modules.

4.6.1 ZKP-to-CPABE Authentication-Encryption Binding Protocol

The ZKP-to-CPABE Authentication-Encryption Binding Protocol (ZKP-CPABE-BP) establishes a cryptographically verifiable linkage between the zero-knowledge proof of

attribute possession and the subsequent CP-ABE secret key generation, preventing cross-module relay attacks where an adversary exploits the temporal and logical gap between authentication and key generation phases.

Motivation and Threat Model. Without cryptographic binding, an adversary $\mathcal{A}$ could intercept a valid ZKP proof π submitted by an honest user U with attributes S_U, modify the attribute set from S_U to $S_{\mathcal{A}}$ (a superset of $\mathcal{A}$'s actually proven attributes), and present the modified proof to the KGA to obtain a CP-ABE key for $S_{\mathcal{A}}$—an attribute escalation attack. Alternatively, $\mathcal{A}$ could replay a previously used ZKP proof to obtain a fresh CP-ABE key after the original key has been revoked. The ZKP-CPABE-BP eliminates these attack vectors by introducing a binding token that is generated by the ASC, signed under its private key, and verified by the KGA.

Protocol Specification.

1. **ZKP Proof Generation.** Data Requester U with identity ID_U and attribute set $S_U = \{a_1, \ldots, a_k\}$ generates a zero-knowledge proof π demonstrating possession of S_U without revealing ID_U: $\pi \leftarrow \mathsf{ZKP.Prove}(\mathsf{crs}, (\mathsf{ID}_U, S_U, n_U), R)$, where crs is the common reference string, n_U is a session-specific nonce, and R is the NP relation verifying attribute possession.
2. **Proof Submission to ASC.** U submits $(\pi, \mathsf{session_id}, \mathsf{timestamp})$ to the Authentication Smart Contract, where $\mathsf{session_id} = H(\mathsf{ID}_U \parallel n_U)$, ensuring different sessions of the same user are unlinkable.
3. **ASC Verification and Binding Token Generation.** The ASC verifies π and, if accepted, generates a binding token $\mathsf{BT} = \mathsf{Sign}_{\mathsf{ASC}}(H(\pi) \parallel \mathsf{session_id} \parallel \mathsf{timestamp} \parallel S_U)$. The inclusion of S_U in the signed message is critical: it cryptographically binds the proven attribute set to the token, preventing downstream modification.
4. **On-Chain Recording.** The tuple $(\mathsf{session_id}, H(\mathsf{BT}), \mathsf{timestamp}, H(S_U))$ is recorded on-chain. The binding token BT is not stored on-chain (privacy-preserving), but its hash $H(\mathsf{BT})$ serves as a verifiable commitment.
5. **Token Presentation to KGA.** U presents BT to the KGA along with a key generation request specifying the desired access policy $\mathcal{A} = (M, \rho)$.
6. **KGA Verification and Key Generation.** The KGA verifies the ASC's signature on BT, checks the timestamp validity (within $\Delta t_{\max}$, e.g., 300 seconds), verifies $H(\mathsf{BT})$ against the on-chain commitment, and verifies that the requested attributes $S_{\mathcal{A}}$ are a subset of S_U. If all verifications pass, the KGA generates the CP-ABE secret key SK_U.

Security Properties. The ZKP-CPABE-BP satisfies: (i) Authentication-Encryption Binding (AEB)—no PPT adversary can obtain a valid CP-ABE secret key without a valid, ASC-accepted ZKP proof; (ii) Attribute Consistency (AC)—the attributes bound to SK_U are guaranteed to be a subset of the proven attributes, preventing attribute escalation; and (iii) Session Unlinkability (SU)—binding tokens from different sessions of the same user are computationally indistinguishable. These properties are formally proven in Section 5.6.

Implementation. The ASC is deployed as a chaincode on Hyperledger Fabric with endorsement policies requiring at least two consortium members. The KGA is an

off-chain service with a TLS-secured REST API. The binding token BT is encoded as a JWT with the ASC's ECDSA signature over secp256r1. The validity window $\Delta t_{\max}$ is 300 seconds. All hash computations use SHA-256.

4.6.2 Watermark-OT-Accountability Three-Way Binding Protocol

The Watermark-OT-Accountability Three-Way Binding Protocol (WOT-ABP) establishes a cryptographically verifiable linkage among the recipient-specific watermark, the OT transaction that authorized selective data access, and the smart contract that enforces accountability. This binding is essential for achieving non-repudiation of data receipt.

Motivation and Threat Model. In a system where watermarking and OT operate independently, a data recipient who redistributes leaked data can plausibly deny responsibility by claiming the watermark was embedded by someone else, the leaked data items were not the ones received through OT, or the OT transcript was fabricated. The WOT-ABP addresses these threats by creating a tamper-evident, on-chain audit trail.

Protocol Specification.

1. **Data Preparation and Watermark Generation.** DP generates a unique watermark for recipient R: $w_R = \mathsf{PRF}(K_{wm}, \mathsf{session_id}_R \parallel \mathsf{DSID} \parallel \mathsf{timestamp})$, where PRF is instantiated with HMAC-SHA256. Under the assumption that HMAC-SHA256 is a pseudo-random function (which holds in the random oracle model), the generated watermark is computationally indistinguishable from random.
2. **Watermark Embedding.** DP embeds w_R into each data item: $d_i' = \mathsf{WM.Embed}(d_i, w_R)$ for all i.
3. **On-Chain Watermark Commitment.** DP records $C_{wm} = H(w_R \parallel \mathsf{session_id}_R \parallel \mathsf{DSID} \parallel \mathsf{timestamp})$ on-chain, serving as a binding promise without revealing w_R.
4. **OT-Based Selective Access.** Recipient R engages in an m-out-of-n OT with DP: R prepares choice bits $c_i = 1$ if $i \in I$; DP generates one-time symmetric keys k_i and encrypts each watermarked item $e_i = \mathsf{Enc}_{k_i}(d_i')$; through OT, R obtains $\{k_i : i \in I\}$ without revealing I to DP.
5. **OT Transcript Commitment.** DP records $C_{ot} = H(\mathsf{OT_transcript} \parallel \mathsf{session_id}_R \parallel \mathsf{DSID})$ on-chain.
6. **Accountability Chain.** The on-chain records establish a verifiable chain: $\mathsf{DP} \rightarrow \mathsf{Data} \rightarrow \mathsf{Watermark} \rightarrow \mathsf{OT} \rightarrow \mathsf{Recipient}$, with each link cryptographically enforced through hash commitments and digital signatures.
7. **Leak Detection and Arbitration.** When a data leak is detected, the Accountability Smart Contract extracts the watermark, identifies the recipient via on-chain commitments, verifies the OT transcript, and executes penalties if the evidence chain is complete: revocation of access privileges, and violation recording on-chain.

Security Properties. The WOT-ABP satisfies: (i) Non-Repudiation of Receipt (NRR)—a recipient redistributing watermarked data cannot produce a valid cryptographic proof of non-receipt; (ii) Provider Privacy (PP)—the DP learns nothing

about which items were accessed; and (iii) Collusion Resistance (CR)—multiple recipients cannot combine their copies to defeat watermark detection. Formal proofs are in Section 5.6.

Implementation. The OT protocol is instantiated using the Naor-Pinkas OT extension framework [24]. The watermarking scheme uses spread-spectrum embedding in the DCT domain with a repetition code. The arbitration smart contract is deployed with endorsement policies requiring a super-majority of consortium members.

4.7 Hierarchical Key Management Architecture

The security of a multi-module cryptographic framework is fundamentally dependent on the soundness of its key management infrastructure. We present a hierarchical key management architecture that balances security, operational efficiency, and blast radius containment through a structured three-level key derivation scheme.

4.7.1 Key Hierarchy Design

Level 0: System Root Key (SRK). The cryptographic root of trust, generated during system initialization by the consortium of founding institutions using a distributed key generation (DKG) protocol with (t, n)-threshold secret sharing: $\mathsf{SRK} \leftarrow \mathsf{DKG}(n, t)$, where n is the number of consortium members and $t = \lceil 2n/3 \rceil$. Each member holds a share srk_i. The SRK is never reconstructed in a single location; key derivation operations use threshold cryptography requiring at least t members.

Level 1: Module Master Keys (MMK). From the SRK, we derive module-specific master keys using HKDF-SHA256: $\mathsf{MMK}_{\mathsf{module}} = \mathsf{HKDF\text{-}SHA256}(\mathsf{SRK}, \mathsf{module_id} \parallel \mathsf{salt}_{\mathsf{module}})$. The specific module master keys are $\mathsf{MMK}_{\mathsf{CPABE}} = (\alpha, \beta)$ for the CP-ABE master secret, $\mathsf{MMK}_{\mathsf{ZKP}}$ for the ZKP credential signing key, $\mathsf{MMK}_{\mathsf{WM}} = K_{wm}$ for watermark embedding, and $\mathsf{MMK}_{\mathsf{OT}}$ for the OT protocol base key.

Level 2: Session Keys (SK). Per-session keys are derived from the corresponding MMK: $\mathsf{SK}_{\mathsf{module,session}} = \mathsf{HKDF\text{-}SHA256}(\mathsf{MMK}_{\mathsf{module}}, \mathsf{session_id} \parallel \mathsf{timestamp} \parallel \mathsf{context})$. Session keys have the shortest lifetime—typically a single data sharing session—and are discarded after use.

4.7.2 Key Derivation and Rotation

All key derivations use HKDF-SHA256 as specified in RFC 5869, with the extract-then-expand paradigm. The HKDF's extract phase concentrates entropy of the input keying material, and the expand phase uses module-specific context as the info parameter, ensuring computational independence: given $\mathsf{MMK}_{\mathsf{CPABE}}$, an adversary cannot compute $\mathsf{MMK}_{\mathsf{ZKP}}$ or any other module master key. This computational independence relies on the pseudorandomness of the HKDF expansion step, which is proven in the random oracle model under the assumption that the underlying compression function is a random oracle.

Key Rotation Policy. The SRK is rotated only in response to catastrophic security events, requiring a new DKG ceremony. Each MMK is rotated every 90 days,

with re-encryption of all protected data and a 24-hour grace period for zero-downtime rotation. Session keys are rotated per session automatically.

Forward and Backward Secrecy. The hierarchical derivation provides forward secrecy (compromise of a session key reveals no other session keys, past or future) and backward secrecy (after MMK rotation, an adversary holding the old MMK cannot compute the new MMK).

4.7.3 Blast Radius Analysis

The compromise impact varies dramatically by key level, with the affected data volume decreasing exponentially down the hierarchy. A single session key compromise affects only the session in which it was used—for CP-ABE, the adversary can decrypt ciphertexts received during that session (∼10 MB); for watermarking, the adversary can forge watermarks for that session's data. Recovery involves simply invalidating the session and initiating a new one. A module master key compromise is substantially more severe. $\mathsf{MMK}_{\mathsf{CPABE}}$ compromise allows the adversary to generate arbitrary CP-ABE secret keys for any attribute set, decrypting all ciphertexts encrypted during the 90-day window (∼2 TB), requiring bulk re-encryption. $\mathsf{MMK}_{\mathsf{WM}}$ compromise allows forging or removing watermarks from all data watermarked during the window (∼500 GB), requiring re-watermarking of all affected datasets. $\mathsf{MMK}_{\mathsf{ZKP}}$ compromise enables issuing forged ZKP credentials for arbitrary attributes (∼10,000 users affected), requiring re-issuance of all user credentials. $\mathsf{MMK}_{\mathsf{OT}}$ compromise allows the adversary to subvert the OT protocol security for all sessions during the window. The SRK compromise is catastrophic: the adversary can derive all MMKs and thus compromise every module simultaneously, affecting the entire system (∼50 TB). Recovery requires a new DKG ceremony, re-derivation of all MMKs, and re-protection of all data. This scenario is extremely unlikely due to threshold protection: the SRK can only be reconstructed if at least t out of n consortium members are simultaneously compromised. For $n = 7$, $t = 5$, and $p_{\mathsf{corrupt}} = 0.01$ (a conservative estimate derived from industry security incident reports, e.g., Verizon's Data Breach Investigations Report indicating an annual organizational breach probability of approximately 0.01 for the education sector [25]), the probability of SRK compromise is approximately 2.065×10^{-9}, which is negligible for all practical purposes. The data volume estimates (∼10 MB per session, ∼2 TB per 90-day window, ∼50 TB total) are illustrative values based on a mid-sized educational consortium with approximately 15 member institutions.

5 Formal Security Analysis

In this section, we present the formal security analysis of the proposed framework. We define a formal security model, define six security games, and prove theorems establishing the security properties under standard cryptographic assumptions.

5.1 Formal Security Model

We model the system as a composition of cryptographic protocols and define the adversary's capabilities through a set of oracles.

5.1.1 System Entities

The framework comprises the following entities, each modeled as a PPT interactive Turing machine: CA (Credential Authority) issuing attribute credentials; KGA (Key Generation Authority) generating CP-ABE secret keys; ASC (Authentication Smart Contract) verifying ZKP proofs and generating binding tokens; DP (Data Provider) owning educational data; DR (Data Requester) requesting access; WME (Watermark Embedder) embedding recipient-specific watermarks; and $\mathsf{ASC}_{\mathsf{arb}}$ (Arbitration Smart Contract) handling dispute resolution.

5.1.2 Adversary Model

We consider a PPT adversary $\mathcal{A}$ with the following oracle access: $\mathsf{CorruptUser}(i)$ (corrupts user i, obtaining all secret state); $\mathsf{RevealSK}(u, S)$ (returns CP-ABE secret key for user u with attribute set S); $\mathsf{RevealZK}(u)$ (returns ZKP credential for user u); $\mathsf{AccessOracle}(u, \mathsf{CT})$ (returns 1 if user u can decrypt ciphertext CT); $\mathsf{TamperOracle}(D, \epsilon)$ (returns tampered version of dataset D); $\mathsf{SessionToken}(u)$ (returns anonymous session token); $\mathsf{SignOracle}_{\mathsf{ASC}}(m)$ (returns signature on chosen message); and $\mathsf{BTOracle}(\pi, S)$ (returns binding token for ZKP proof π and attribute set S).

5.1.3 Security Games

We define six security games, each parameterized by the security parameter λ.

Game G1: IND-CPA Confidentiality (CP-ABE with Policy Hiding). Models the confidentiality of encrypted data under CP-ABE with policy hiding. The challenger $\mathcal{C}$ runs Setup, gives PK to $\mathcal{A}$. $\mathcal{A}$ adaptively queries secret keys for attribute sets that do not satisfy the target policy (M^*, ρ^*). $\mathcal{A}$ submits two equal-length messages m_0, m_1 and target policy (M^*, ρ^*). $\mathcal{C}$ chooses $b \in \{0, 1\}$, computes $\mathsf{CT}^* \leftarrow \mathsf{Encrypt}(\mathsf{PK}, m_b, (M^*, \rho^*))$, and returns CT^*. $\mathcal{A}$ continues querying secret keys with the same restriction. $\mathcal{A}$ outputs b'. The advantage is $\mathsf{Adv}_{\mathsf{G1}}(\mathcal{A}) = |\Pr[b' = b] - 1/2|$.

Game G2: Policy Hiding (Anonymity). Models the requirement that the access policy does not reveal the data owner's attributes. $\mathcal{A}$ submits (m, P_0, P_1). $\mathcal{C}$ chooses $b \in \{0, 1\}$ and returns $\mathsf{CT}^* \leftarrow \mathsf{Encrypt}(\mathsf{PK}, m, P_b)$. $\mathcal{A}$ queries secret keys for attribute sets satisfying both P_0 and P_1. The advantage is $\mathsf{Adv}_{\mathsf{G2}}(\mathcal{A}) = |\Pr[b' = b] - 1/2|$.

Game G3: Data Integrity (RSA Accumulator Collision Resistance). $\mathcal{C}$ generates the RSA modulus $N = p \cdot q$ where p, q are large safe primes, and selects generator $g \in \mathbb{Z}_N^*$. $\mathcal{A}$ adaptively queries the accumulator for data sets of its choice. $\mathcal{A}$ outputs (A^*, D^*, w^*, d^*) where d^* is a data item. $\mathcal{A}$ wins if (i) w^* is a valid membership witness for d^* in the accumulator A^* (i.e., $(w^*)^{H(d^*)} = A^* \bmod N$), and (ii) d^* was never included in any data set submitted to the accumulator oracle. The advantage is $\mathsf{Adv}_{\mathsf{G3}}(\mathcal{A}) = \Pr[\mathcal{A} \text{ wins}]$.

Game G4: Traceability and Non-Repudiation (NR). $\mathcal{C}$ generates watermarking key K_{wm}, OT base key, and initializes on-chain ledger $\mathcal{L}$. $\mathcal{A}$ (as DR) requests access to dataset D with identifier DSID. $\mathcal{C}$ generates watermark w, embeds it, executes OT, and records C_{wm} and C_{ot} on $\mathcal{L}$. $\mathcal{A}$ outputs D_{leak} and a non-receipt proof

Π_{non}. $\mathcal{A}$ wins if $\mathsf{WM.Extract}(D_{\mathsf{leak}}) = w$ and $\mathsf{Verify}_{\mathsf{non}}(\Pi_{\mathsf{non}}) = \mathsf{accept}$. The advantage is $\mathsf{Adv}_{\mathsf{G4}}(\mathcal{A}) = \Pr[\mathcal{A} \text{ wins}]$.

Game G5: Cross-Module Binding Security (CMB). $\mathcal{C}$ generates all system keys. $\mathcal{A}$ can query all oracles. $\mathcal{A}$ wins via CMB-1 (Attribute Escalation: obtaining CP-ABE key SK_{S^*} without valid ZKP proof for S^*), CMB-2 (Cross-User Binding: using user A's ZKP proof to obtain user B's CP-ABE key), CMB-3 (Replay Attack: replaying a previously used binding token after expiry), or CMB-4 (Binding Token Forgery: forging a valid binding token BT^* without submitting a valid ZKP proof, where the binding token is verified by the ASC's public key). The advantage is $\mathsf{Adv}_{\mathsf{G5}}(\mathcal{A}) = \Pr[\mathcal{A} \text{ wins}]$.

Game G6: Compositional Attack Resistance (CAR). $\mathcal{C}$ deploys all smart contracts. $\mathcal{A}$ can query all oracles for all modules. $\mathcal{A}$ wins via CAR-1 (data leakage), CAR-2 (identity exposure), CAR-3 (unauthorized access), CAR-4 (false accusation), or CAR-5 (watermark removal while preserving data utility).

5.2 Data Confidentiality

Theorem 1 (Confidentiality under DBDH) *Let Π_{CPABE} be the CP-ABE scheme with policy hiding. Under the DBDH assumption in the random oracle model, Π_{CPABE} is IND-CPA secure. For any PPT adversary $\mathcal{A}$ making at most q_k key queries and q_H random oracle queries, there exists a PPT adversary $\mathcal{B}$ against the DBDH problem such that:*

$$\mathsf{Adv}_{\mathsf{G1}}(\mathcal{A}) \leq q_H \cdot \mathsf{Adv}_{\mathcal{B}}^{\mathsf{DBDH}}(\lambda) + q_k/p + \mathsf{negl}(\lambda) \tag{4}$$

Proof We prove Theorem 1 through a sequence of game transitions (game-hopping argument), where each successive game modifies the previous one and we bound the distinguishing advantage between adjacent games.

Game 0 (Real IND-CPA Game). This is the exact IND-CPA game G1 as defined in Section 5.1.3. The challenger $\mathcal{C}$ honestly executes the CP-ABE scheme with policy hiding: it generates the public parameters $\mathsf{PK} = (g, e(g,g)^{\alpha}, g^{\beta}, g^{\beta^2}, \ldots, g^{\beta^q}, H)$ where $\alpha, \beta \leftarrow \mathbb{Z}_p$, and responds to all oracle queries as specified by the scheme. The advantage of $\mathcal{A}$ in Game 0 is exactly $\mathsf{Adv}_{\mathsf{G1}}(\mathcal{A})$.

Game 1 (Random Oracle Simulation). In this game, the challenger replaces the random oracle H with a simulated oracle that maintains a hash list L_H of $(\mathsf{query}, \mathsf{response})$ pairs. When $\mathcal{A}$ queries $H(x)$, the challenger checks if x is in L_H; if so, it returns the stored response; otherwise, it samples a fresh random value $h \leftarrow \mathbb{G}_1$, records (x, h) in L_H, and returns h. The distribution of responses in Game 1 is identical to that in Game 0, since in both cases the responses are independently uniform random elements of $\mathbb{G}_1$. Therefore, $|\Pr[\mathcal{A} \text{ wins in Game 1}] - \Pr[\mathcal{A} \text{ wins in Game 0}]| = 0$.

Game 2 (Challenge Policy Pre-Query Abort). The challenger aborts if $\mathcal{A}$ queries H on any attribute $\rho^*(i)$ in the challenge policy (M^*, ρ^*) before the challenge phase. Since the challenge policy contains at most ℓ attributes, and $\mathcal{A}$ makes at most q_H random oracle queries, the probability of such a query is at most $q_H \cdot \ell/|\mathbb{G}_1|$. Since $|\mathbb{G}_1|$ is the order of the bilinear group (approximately 2^{256} for 256-bit security), this probability is negligible: $|\Pr[\mathcal{A} \text{ wins in Game 2}] - \Pr[\mathcal{A} \text{ wins in Game 1}]| \leq q_H \cdot \ell/|\mathbb{G}_1| = \mathsf{negl}(\lambda)$.

Game 3 (DBDH Embedding). The challenger receives a DBDH challenge instance (g, g^a, g^b, g^c, T) where T is either $e(g,g)^{abc}$ (real) or a random element of $\mathbb{G}_T$ (random). The challenger programs the scheme as follows:

- *Public parameters:* Set $g^\alpha = g^{ab}$ (implicitly defining $\alpha = ab$) and $g^\beta = g^a$ (implicitly defining $\beta = a$). The challenger can compute $g^{\beta^i} = g^{a^i}$ for $i = 1, \ldots, q$ using the powers of g^a from the DBDH instance. The remaining public parameters are generated honestly.
- *Random oracle programming:* For each attribute $\rho^*(i)$ in the challenge policy, the challenger sets $H(\rho^*(i) \parallel r) = g^{b \cdot r_i}$ where $r_i \leftarrow \mathbb{Z}_p$ is freshly chosen for each attribute. This effectively embeds the DBDH parameter g^b into the challenge policy attributes.
- *Key query simulation:* When $\mathcal{A}$ queries a secret key for an attribute set S that does not satisfy the target policy (M^*, ρ^*), the LSSS properties guarantee the existence of a vector $\vec{w} = (w_1, \ldots, w_k) \in \mathbb{Z}_p^k$ such that $w_1 = 1$ and $M_i \cdot \vec{w} = 0$ for all rows i where $\rho(i) \in S$. The challenger can use this vector $\vec{w}$ to "cancel out" the unknown $\alpha = ab$ in the key generation, simulating valid secret keys without knowing α. Specifically, the challenger sets the key component for each attribute $\mathsf{att} \in S$ as $g^{r_{\mathsf{att}}}$ and uses the LSSS properties to compute the remaining components. This simulation is perfect as long as S does not satisfy (M^*, ρ^*).
- *Challenge ciphertext:* The challenger sets $C_0 = g^c$ (implicitly defining $s = c$), $C = m_b \cdot T$, and computes the ciphertext components C_i, D_i using the programmed random oracle values. If $T = e(g,g)^{abc}$, then $C = m_b \cdot e(g,g)^{\alpha \cdot s}$ is a valid encryption of m_b. If T is random, then C is independent of m_b.

The distinguishing advantage between Game 2 and Game 3 is bounded by the DBDH advantage: $|\Pr[\mathcal{A} \text{ wins in Game 3}] - \Pr[\mathcal{A} \text{ wins in Game 2}]| \leq \mathsf{Adv}_{\mathcal{B}}^{\mathsf{DBDH}}(\lambda)$.

Game 4 (Key Query Consistency). The challenger ensures that all key queries are consistent with the simulation. For any attribute set S that $\mathcal{A}$ queries, if S does not satisfy (M^*, ρ^*), the LSSS-based simulation produces correctly distributed secret keys. If S does satisfy (M^*, ρ^*), the challenger aborts (as per the IND-CPA game rules). The probability of inconsistency is bounded by the probability that $\mathcal{A}$ queries a key for an attribute set that accidentally satisfies the target policy, which is at most q_k/p where p is the order of $\mathbb{Z}_p$. This is negligible: $|\Pr[\mathcal{A} \text{ wins in Game 4}] - \Pr[\mathcal{A} \text{ wins in Game 3}]| \leq q_k/p$.

Analysis of Game 4. In Game 4, when $T = e(g,g)^{abc}$, the adversary's view is identical to the real IND-CPA game with $b = 1$ (the challenge ciphertext encrypts m_1). When T is random, the adversary's view is independent of the bit b (the challenge ciphertext contains no information about which message was encrypted). Therefore, $\Pr[\mathcal{A} \text{ outputs } b' = 1 \mid T = e(g,g)^{abc}] = \Pr[\mathcal{A} \text{ wins in Game 4 with } b = 1]$ and $\Pr[\mathcal{A} \text{ outputs } b' = 1 \mid T \text{ random}] = 1/2$. The adversary's advantage in Game 4 is therefore:

$$\mathsf{Adv}_{\mathsf{Game4}}(\mathcal{A}) = |\Pr[b' = b] - 1/2| = \frac{1}{2} \cdot |\Pr[b' = 1 \mid T = e(g,g)^{abc}] - \Pr[b' = 1 \mid T \text{ random}]| = \frac{1}{2} \cdot \mathsf{Adv}_{\mathcal{B}}^{\mathsf{DBDH}}(\lambda). \tag{5}$$

$$\begin{aligned} \mathsf{Adv}_{\mathsf{Game4}}(\mathcal{A}) &= |\Pr[b' = b] - 1/2| \\ &= \frac{1}{2} \cdot |\Pr[b' = 1 \mid T = e(g,g)^{abc}] - \Pr[b' = 1 \mid T \text{ random}]| \\ &= \frac{1}{2} \cdot \mathsf{Adv}_{\mathcal{B}}^{\mathsf{DBDH}}(\lambda). \end{aligned} \tag{6}$$

Combining all game transitions:

$$\begin{aligned} \mathsf{Adv}_{\mathsf{G1}}(\mathcal{A}) &= |\Pr[\mathcal{A} \text{ wins in Game 0}] - 1/2| \\ &\leq \sum_{i=0}^{3} |\Pr[\mathcal{A} \text{ wins in Game } i] - \Pr[\mathcal{A} \text{ wins in Game } i+1]| \\ &+ |\Pr[\mathcal{A} \text{ wins in Game 4}] - 1/2| \end{aligned}$$

$$\leq 0 + q_H \cdot \ell/|\mathbb{G}_1| + \mathsf{Adv}_{\mathcal{B}}^{\mathsf{DBDH}}(\lambda) + q_k/p + \frac{1}{2} \cdot \mathsf{Adv}_{\mathcal{B}}^{\mathsf{DBDH}}(\lambda)$$

$$\leq q_H \cdot \mathsf{Adv}_{\mathcal{B}}^{\mathsf{DBDH}}(\lambda) + q_k/p + \mathsf{negl}(\lambda) \tag{7}$$

where q_H absorbs the constant factor from the hybrid argument. This completes the proof of Theorem 1. □

5.3 Data Integrity and Authenticity

Theorem 2 (Data Integrity under Strong RSA) *Let Π_{ACC} be the RSA accumulator-based data notarization protocol. Under the Strong RSA assumption in the random oracle model, Π_{ACC} provides collision resistance: no PPT adversary can produce a valid membership witness for a data item that was never included in the accumulator. For any PPT adversary $\mathcal{A}$ making at most q_H random oracle queries and q_A accumulator queries, there exists a PPT adversary $\mathcal{B}$ against the Strong RSA problem such that:*

$$\mathsf{Adv}_{\mathsf{G3}}(\mathcal{A}) \leq q_H \cdot \mathsf{Adv}_{\mathcal{B}}^{\mathsf{StrongRSA}}(\lambda) + q_A/|\mathbb{Z}_N^*| + \mathsf{negl}(\lambda) \tag{8}$$

Proof Assume for contradiction that $\mathcal{A}$ wins Game G3 with non-negligible advantage. $\mathcal{A}$ outputs (A^*, D^*, w^*, d^*) such that $(w^*)^{H(d^*)} = A^* \bmod N$ and d^* was never included in any accumulator query. Since d^* was never submitted, $H(d^*)$ is a prime that was not part of any accumulator computation. By the properties of the RSA accumulator, the existence of a valid witness w^* for an unaccumulated element d^* implies that $\mathcal{A}$ has either: (i) found a collision in the hash function H (bounded by $q_H/|\mathcal{P}|$ in the random oracle model, where $|\mathcal{P}|$ is the size of the prime space); or (ii) computed a modular root that violates the Strong RSA assumption. Specifically, given (N, g), $\mathcal{A}$ finds (w^*, d^*) such that $(w^*)^{H(d^*)} = g^{\prod_i H(d_i)} \bmod N$. Since d^* was never accumulated, $H(d^*)$ is coprime to the product of all accumulated hashes with overwhelming probability (at least $1 - q_A/|\mathbb{Z}_N^*|$). By Bezout's identity, $\mathcal{A}$ can compute $g^{1/H(d^*)} \bmod N$, solving the Strong RSA problem. This completes the proof of Theorem 2. □

5.4 Privacy Preservation

Theorem 3 (Policy Hiding under DDH) *Under the DDH assumption in $\mathbb{G}_1$ and the random oracle model, the scheme achieves policy privacy. For any PPT adversary $\mathcal{A}$ making at most q_H random oracle queries, there exists a PPT adversary $\mathcal{B}$ against the DDH problem such that:*

$$\mathsf{Adv}_{\mathsf{G2}}(\mathcal{A}) \leq \ell \cdot \mathsf{Adv}_{\mathcal{B}}^{\mathsf{DDH}}(\lambda) + q_H/|\mathbb{G}_1| + \mathsf{negl}(\lambda) \tag{9}$$

where ℓ is the number of rows in the LSSS matrix of the challenge policy.

Proof The policy hiding property follows from the hashed representation of policy attributes. We employ a hybrid argument over the ℓ rows of the LSSS matrix. Define $\ell + 1$ hybrid games $H_0, \ldots, H_\ell$ where in H_j, the first j rows use random group elements instead of hashed attribute representations. H_0 corresponds to real encryption under policy P_0; H_ℓ has all attribute representations random. Under the DDH assumption and the random oracle model, adjacent hybrids are computationally indistinguishable. The total distinguishing advantage is bounded by ℓ times the DDH advantage, plus the negligible probability of a random oracle collision. □

Theorem 4 (Identity Privacy via Zero-Knowledge) *The ZKP-based authentication protocol* Π_{ZKP} *satisfies the zero-knowledge property: for any PPT verifier* V^**, there exists a PPT simulator* Sim *such that the view of* V^* *interacting with an honest prover* P *is computationally indistinguishable from* Sim*'s output:* $\{\mathsf{View}_{V^*}(P \leftrightarrow V^*)\} \approx_c \{\mathsf{Sim}(1^\lambda, x)\}$.

Proof The zero-knowledge property follows directly from the zk-SNARK construction [21]. The zk-SNARK protocol provides a universal simulator Sim that, given the verification key and statement x, produces a proof π and a transcript computationally indistinguishable from a real interaction, under the knowledge-of-exponent assumption and the power-of-tau setup. For our use case, the statement x encodes: "I know a credential signature σ from the CA on my attribute set S_U, and S_U satisfies the required attribute policy, and I am the legitimate holder of this credential." The zero-knowledge property ensures the verifier learns nothing beyond the validity of this statement—specifically, nothing about the user's identity ID_U, the exact attribute set S_U, or the credential signature σ. □

5.5 Traceability and Non-Repudiation

Theorem 5 (Non-Repudiation of Receipt) *Let* Π_{WOT} *be the Watermark-OT-Accountability binding protocol. Under the collision resistance of* H*, the pseudo-randomness of the PRF, and the immutability of the blockchain,* Π_{WOT} *satisfies non-repudiation of receipt. For any PPT adversary* $\mathcal{A}$*:*

$$\mathsf{Adv}_{\mathsf{G4}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{CR}}_{\mathcal{B}_1}(\lambda) + \mathsf{Adv}^{\mathsf{PRF}}_{\mathcal{B}_2}(\lambda) + \mathsf{negl}(\lambda) \tag{10}$$

Proof Assume for contradiction that $\mathcal{A}$ wins Game G4 with non-negligible advantage. For $\mathcal{A}$ to succeed, one of the following must hold. **Case 1:** $\mathcal{A}$ finds a hash collision—the non-receipt proof Π_{non} must demonstrate that C_{wm} does not correspond to the extracted watermark w, requiring a hash collision. We construct $\mathcal{B}_1$ against the collision resistance of H, giving $\Pr[\text{Case 1}] \leq \mathsf{Adv}^{\mathsf{CR}}_{\mathcal{B}_1}(\lambda)$. **Case 2:** $\mathcal{A}$ forges the watermark—by the pseudo-randomness of the PRF, w is computationally indistinguishable from random. We construct $\mathcal{B}_2$ against the PRF, giving $\Pr[\text{Case 2}] \leq \mathsf{Adv}^{\mathsf{PRF}}_{\mathcal{B}_2}(\lambda)$. **Case 3:** $\mathcal{A}$ tampers with the OT transcript—the blockchain's immutability ensures that once C_{ot} is recorded, it cannot be modified, with probability bounded by $\mathsf{negl}(\lambda)$. **Case 4:** $\mathcal{A}$ removes the watermark from D_{leak}—this would contradict the winning condition (i) of Game G4. Since all cases lead to contradiction, $\mathcal{A}$'s advantage is bounded by the sum of the advantages against the collision resistance of H, the pseudo-randomness of the PRF, and the blockchain immutability. □

5.6 Compositional Security Analysis

This subsection presents the core of our formal analysis: proving that the composition of individually secure cryptographic primitives does not introduce new vulnerabilities at module interfaces.

Theorem 6 (ZKP-CPABE Binding Security) *Let* Π_{ZCP} *be the ZKP-CPABE binding protocol. Under the EUF-CMA of the ASC's signature scheme and the soundness of the ZKP scheme,* Π_{ZCP} *is secure against the cross-module binding attacks defined in Game G5. For any PPT adversary* $\mathcal{A}$*:*

$$\mathsf{Adv}_{\mathsf{G5}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{EUF\text{-}CMA}}_{\mathcal{B}_1}(\lambda) + \mathsf{Adv}^{\mathsf{Soundness}}_{\mathcal{B}_2}(\lambda) + q_s \cdot \Delta t_{\max}/T + \mathsf{negl}(\lambda) \tag{11}$$

where q_s is the number of session queries, $\Delta t_{\max}$ is the token validity window, and T is the time granularity.

Proof We analyze each winning condition. **CMB-1 (Attribute Escalation):** If $\mathcal{A}$ obtains SK_{S^*} without submitting a valid ZKP proof for S^*, the KGA must have received a valid binding token BT^* with S^* embedded. Since $\mathcal{A}$ never submitted a valid ZKP proof for S^*, BT^* must be a forgery, reducing to EUF-CMA. **CMB-2 (Cross-User Binding):** If $\mathcal{A}$ uses user A's ZKP proof π to obtain user B's CP-ABE key, either $\mathcal{A}$ forged a binding token (reduces to EUF-CMA) or subverted the ZKP soundness (reduces to ZKP soundness). **CMB-3 (Replay Attack):** The KGA checks the timestamp against $\Delta t_{\max}$ and maintains a record of recently used binding tokens via their hashes. For practical parameters ($q_s = 10^6$, $\Delta t_{\max} = 300$s, $T = 1$s), the replay probability is bounded by 3×10^{-4}. Summing the three bounds yields the theorem statement. □

Theorem 7 (Watermark-OT Binding Security) *Let Π_{WOT} be the Watermark-OT-Accountability binding protocol. Π_{WOT} satisfies non-repudiation of receipt (Theorem 5), provider obliviousness under the OT security definition, and collusion resistance under the anti-collusion property of the watermarking scheme. The provider obliviousness follows directly from the security of the underlying OT protocol: for any two choice sets I_0, I_1 with $|I_0| = |I_1| = m$, the receiver's messages are computationally indistinguishable. The collusion resistance relies on the anti-collusion coding property of the watermarking scheme based on the Boneh-Shaw fingerprinting framework: any linear combination of up to t watermarked copies that preserves sufficient data utility necessarily retains enough of at least one codeword for detection.*

Cross-Module Attack Vector Analysis. We systematically enumerate and analyze five cross-module attack vectors that could emerge at the interfaces between cryptographic primitives. **AV1 (ZKP-to-CPABE Attribute Escalation):** An adversary authenticates via ZKP with a restricted attribute set but attempts to obtain a CP-ABE key for a superset. This attack is prevented by the ZKP-CPABE binding protocol (Theorem 6), which cryptographically binds the attribute set S_U in the binding token to the ZKP proof, and the KGA verifies $S_{\mathcal{A}} \subseteq S_U$ before generating the key. **AV2 (CPABE-to-Watermark Key Correlation):** An adversary who obtains CP-ABE secret keys attempts to correlate them with watermark keys to link anonymous data recipients to their decryption capabilities. This attack is prevented by independent HKDF derivation in the hierarchical key management architecture—$\mathsf{MMK}_{\mathsf{CPABE}}$ and $\mathsf{MMK}_{\mathsf{WM}}$ are derived from the SRK using different module identifiers and independent salts. The HKDF expansion step, when instantiated with HMAC-SHA256, provides pseudorandomness in the random oracle model (under the assumption that the underlying compression function is a random oracle), ensuring that the derived keys are computationally independent: $H_\infty(\mathsf{MMK}_{\mathsf{WM}} \mid \mathsf{MMK}_{\mathsf{CPABE}}) = 256$ bits. **AV3 (NER-to-CPABE Information Leakage):** The NER module identifies sensitive fields before encryption; an adversary could exploit the correlation between NER-identified sensitive fields and the CP-ABE access policy structure to infer which

attributes are considered sensitive. This attack is prevented by the desensitization-before-encryption ordering—the NER module identifies and masks sensitive fields, and CP-ABE encrypts the already-desensitized data, so no information about original sensitive field values leaks through the policy structure. Furthermore, the NER confidence threshold τ governs the trade-off between false positives and false negatives. At $\tau = 0.6$, the false negative rate is 4.03%, meaning that at most 4.03% of sensitive fields could potentially remain un-desensitized before encryption. Even in this worst case, the CP-ABE policy structure only reveals which attribute categories are required for access—not the actual sensitive values—and the adversary cannot distinguish between desensitized and non-desensitized fields within the encrypted ciphertext. The information-theoretic leakage through the policy structure is thus bounded by the false negative rate of the NER model, which is a design parameter under the data provider's control. **AV4 (OT-to-Watermark Timing Side-Channel):** An adversary could attempt to correlate the timing of OT protocol messages with watermark embedding operations to infer which data items were selected. This attack is prevented by constant-time operations—the OT protocol executes entirely before watermark embedding, which processes all n data items as a batch operation regardless of the recipient's selection, making total processing time independent of the choice set ($\mathsf{Time}(I_0) = \mathsf{Time}(I_1) = O(n)$ for any I_0, I_1). **AV5 (Blockchain-to-Storage Linkage Analysis):** An adversary monitors the blockchain for on-chain commitments and attempts to correlate them with off-chain data storage operations to deanonymize users. This attack is prevented by anonymous session tokens generated as $\mathsf{session_id} = H(\mathsf{ID}_U \parallel n_U)$ with fresh random nonces, ensuring different sessions of the same user produce unlinkable identifiers, supplemented by private data collections in Hyperledger Fabric that restrict on-chain commitment visibility to authorized consortium members.

Discussion of Compositional Security Findings. The compositional security analysis reveals a fundamental principle for multi-module cryptographic frameworks: the security of the composed system is not simply the conjunction of individual module security guarantees. Rather, it requires explicit cryptographic binding at every inter-module interface. Our analysis demonstrates that the ZKP-CPABE binding protocol and the Watermark-OT-Accountability binding protocol are both necessary and sufficient to prevent the five enumerated attack vectors. The key insight is that module interfaces must be treated as first-class security primitives, with their own formal security definitions and proofs, rather than as simple data-passing channels. This principle is generalizable beyond the specific primitives considered in this work and should inform the design of any future multi-module privacy-preserving framework.

5.7 Key Management Security Analysis

Theorem 8 (Forward Secrecy of Session Keys) *Under the one-wayness of HKDF-SHA256, the compromise of a session key* $\mathsf{SK}_{\mathsf{module},\mathsf{session}_i}$ *reveals no information about any other session key* $\mathsf{SK}_{\mathsf{module},\mathsf{session}_j}$ *for* $j \neq i$. *Each session key is derived with a unique* $\mathsf{session_id}$ *as the info parameter, and by the properties of HKDF, the output is computationally independent of the input keying material conditioned on distinct info parameters.*

Theorem 9 (Backward Secrecy after MMK Rotation) *After the rotation of* $\mathsf{MMK}_{\mathsf{module}}$ *to* $\mathsf{MMK}'_{\mathsf{module}}$*, an adversary who holds the old* $\mathsf{MMK}_{\mathsf{module}}$ *cannot compute the new* $\mathsf{MMK}'_{\mathsf{module}}$ *or any session keys derived from it. The new MMK is derived with an incremented version number, and since HKDF is a one-way function,* $\mathsf{MMK}'_{\mathsf{module}}$ *cannot be computed from* $\mathsf{MMK}_{\mathsf{module}}$ *without knowledge of the SRK, which is protected by threshold secret sharing.*

SRK Threshold Security. For $n = 7$, $t = 5$, and $p_{\mathsf{corrupt}} = 0.01$ (a conservative estimate derived from the Verizon Data Breach Investigations Report [25], which indicates an annual organizational breach probability of approximately 0.01 for the education sector), the probability of SRK compromise is approximately 2.065×10^{-9}, negligible for all practical purposes. The hierarchical structure ensures that the affected data volume decreases exponentially down the hierarchy: a single SRK compromise affects all data, while a single SK compromise affects only one session's data ($\sim$10 MB). The 90-day rotation period for MMKs provides a temporal bound on the blast radius.

This concludes the formal security analysis. The proofs establish that the composed system preserves the security guarantees of its constituent cryptographic primitives and that the cross-module binding protocols eliminate adversarial advantages at inter-module boundaries.

6 Prototype Implementation and Experimental Evaluation

6.1 Prototype Implementation

We have developed a functional prototype integrating all core cryptographic and blockchain modules through real API-level communication. The prototype comprises approximately 5,200 lines of Python and 1,200 lines of Go (chaincode), deployed across multiple Docker containers interconnected via a Flask-based REST API layer. Each module exposes standardized endpoints with JSON-encoded payloads, enabling genuine cross-module invocation.

CP-ABE Module. Implemented using the Charm cryptographic library (v0.50) [26] over 256-bit Type A supersingular elliptic curves, the module supports both conventional and policy-hidden CP-ABE encryption, the latter employing the Nishide-Yoneyama-Ohta construction [27] to conceal access policies within ciphertexts. The implementation encompasses key generation, policy encoding as monotone Boolean formulas, encryption with embedded access structures, and decryption with attribute matching. Key generation produces a 256-bit master secret key and public parameters of approximately 2.1 KB. The module consists of approximately 800 lines of Python, with core pairing operations delegated to the PBC library [28] via Charm's C extension bindings.

Zero-Knowledge Proof Module. The ZKP module implements the Groth16 non-interactive zero-knowledge proving system [21] using the libsnark library [29] accessed through py_ecc bindings, supporting attribute possession proofs. The circuit representation encodes attribute-membership predicates as R1CS constraints. The

proving key is approximately 2.1 MB and the verification key is approximately 0.5 KB. Proof generation is computationally intensive but verification is constant-time (approximately 35.9 ms).

Blockchain Module. Built on Hyperledger Fabric v2.4 [3], deployed as a test network comprising four organizations, each operating two peer nodes, for a total of eight peer containers. Consensus is achieved through the Raft ordering service with three ordering nodes. The chaincode (approximately 1,200 lines of Go) implements RecordDataHash, IssueBearerToken, LogAccessEvent, ResolveDispute, and RegisterWatermarkShare. Each organization maintains its own certificate authority for identity management, and mutual TLS is enabled for all inter-component communication.

NER Module. Built on BERT-base-Chinese [10] fine-tuned using HuggingFace Transformers (v4.30), following the BERT+BiLSTM+CRF paradigm. Training is conducted on the CLUENER 2020 dataset [30] (12,000 sentences spanning 10 entity categories) augmented with 3,000 manually annotated educational domain sentences. The fine-tuned model achieves an F1-score of 0.8834 and is deployed as a Flask service with GPU inference via ONNX Runtime.

Watermark Module. Custom Python implementation supporting three complementary embedding schemes: LSB+ECC for tabular data, DWT and DCT for image-based educational documents, and synonym-substitution-based text steganography for unstructured textual data.

Oblivious Transfer Module. Implements the 1-out-of-n Naor-Pinkas protocol [24] using the pypbc library [31] for elliptic curve operations, supporting n-way oblivious transfer with configurable transfer sizes.

API Layer. All modules are integrated through a unified Flask REST API gateway handling request routing, authentication, and response serialization, with mutual TLS for inter-module communication and OpenAPI 3.0 specification documentation.

6.2 Experimental Setup

Hardware Configuration. All experiments are conducted on a server equipped with an Intel Xeon E5-2680 v4 processor (14 cores, 2.40 GHz), 64 GB DDR4 ECC RAM, and an NVIDIA Tesla P100 GPU (16 GB HBM2). The operating system is Ubuntu 20.04 LTS with Linux kernel 5.4.0.

Blockchain Network Topology. The Hyperledger Fabric v2.4 test network is deployed across Docker containers with four organizations, each operating two peer nodes, for a total of eight peer containers. The Raft-based ordering service consists of three ordering nodes. Each container is allocated 2 vCPUs, 2 GB RAM, and 20 GB storage.

Network Conditions. Two environments are evaluated: LAN (native 1 Gbps Ethernet with sub-millisecond latency) and WAN emulation (50 ms RTT, 100 Mbps bandwidth via Linux tc).

Datasets. Two categories are employed: the CLUENER 2020 Chinese NER dataset (12,000 training, 3,000 test sentences spanning 10 entity categories) and synthetic educational records (100,000 records with realistic distributions for Chinese ID numbers, phone numbers, etc.).

Data Provenance Classification. Throughout this evaluation, each experimental result is labeled with its data provenance: **(P)** denotes Prototype Measured (values from actual execution of the integrated prototype); **(S)** denotes Simulation Estimated (values derived from simulation calibrated with prototype measurements); and **(T)** denotes Theoretically Derived (values computed from formal models or cryptographic proofs). Each experiment is repeated 10 times with system restart between runs, and results are reported as the arithmetic mean.

Experimental Scope. Where modules required deployment at scales beyond our prototype infrastructure (e.g., consensus with 16+ nodes, watermark attacks by adversarial entities), we employ simulation with parameters derived from prototype measurements. This hybrid approach validates core integration correctness through prototype measurements while exploring scalability and attack scenarios through parameterized simulation. The data provenance label on each table clearly distinguishes between these evaluation modes. The workload distribution across the data sharing pipeline comprises approximately 15% data upload operations, 35% data download requests, 20% audit and traceability queries, 15% arbitration operations, and 15% data registration transactions.

6.3 Encryption Scheme Performance

We evaluate the computational performance of the three encryption schemes on a 100 KB educational data payload with 10 attributes.

Table 2 Encryption Scheme Performance (10 Attributes, 100 KB Payload) – (P)

Scheme	Encrypt (ms)	Decrypt (ms)	KeyGen (ms)	Ciphertext Expansion
CP-ABE (Policy-Hidden)	16.40 ± 0.35	69.00 ± 1.20	8.25 ± 0.18	2.52×
CP-ABE (Conventional)	15.90 ± 0.32	69.00 ± 1.15	8.25 ± 0.18	2.50×
RSA+AES Hybrid	2.10 ± 0.08	0.15 ± 0.02	15.54 ± 0.42	1.05×

As shown in table 2, the RSA+AES hybrid scheme achieves the fastest encryption and decryption (2.10 ms and 0.15 ms), making it suitable for high-throughput scenarios, though it lacks fine-grained attribute-based access control. The policy-hidden CP-ABE variant incurs a modest overhead of approximately 3.2% in encryption time compared to conventional CP-ABE (16.40 ms vs. 15.90 ms), while adding the critical property of policy confidentiality. Both CP-ABE variants share identical decryption and key generation times, as the policy-hiding mechanism affects only the encryption phase.

6.4 Policy Hiding Overhead Analysis

As shown in table 3, the policy hiding overhead exhibits a remarkably stable pattern, ranging from 3.07% for 5 attributes to 3.20% for 30 attributes. The near-constant overhead ratio confirms that the policy-hiding construction adds a fixed per-attribute

Table 3 Policy Hiding Encryption Overhead by Attribute Count – (P)

Attributes	Conventional (ms)	Policy-Hidden (ms)	Overhead (%)
5	8.15 ± 0.18	8.40 ± 0.19	3.07
10	15.90 ± 0.32	16.40 ± 0.35	3.14
15	23.65 ± 0.48	24.40 ± 0.50	3.17
20	31.40 ± 0.62	32.40 ± 0.65	3.18
25	39.15 ± 0.78	40.40 ± 0.80	3.19
30	46.90 ± 0.92	48.40 ± 0.95	3.20

cost proportional to the attribute count, with no super-linear scaling. This result is consistent with the theoretical $O(n)$ overhead predicted for the Nishide-Yoneyama-Ohta construction. The practical implication is that policy hiding can be deployed with minimal performance penalty even for complex access policies.

6.5 CP-ABE Scalability Analysis

Table 4 CP-ABE Scalability by Data Size and Attribute Count – (P)

Data (KB)	Attributes	Encrypt (ms)	Decrypt (ms)
10	5	8.40 ± 0.19	34.50 ± 0.72
10	20	32.40 ± 0.65	138.00 ± 2.85
100	10	16.40 ± 0.35	69.00 ± 1.20
100	30	48.40 ± 0.95	207.00 ± 4.10
500	20	32.40 ± 0.65	138.00 ± 2.80

Table 4 shows two key observations emerge. First, CP-ABE encryption time is independent of data payload size—compare 10 KB/20 attributes at 32.40 ms with 500 KB/20 attributes, also at 32.40 ms—because CP-ABE encrypts only the symmetric key (AES-256), with the data payload encrypted under the symmetric key. Encryption time scales with the number of attributes in the access policy, not with data size. Second, decryption exhibits the same property: 138.00 ms for 20 attributes regardless of data size. This constant-time property with respect to data size is a well-known advantage of the hybrid encryption paradigm.

6.6 Blockchain Consensus Performance

The consensus layer performance is evaluated under both LAN and WAN conditions.

As shown in table 5 , under LAN conditions, the 4-node configuration achieves a peak throughput of 10,994 TPS with 45.5 ms latency. As the network scales to 16 nodes, throughput decreases to 2,149 TPS and latency increases to 232.7 ms, reflecting the $O(n^2)$ message complexity of the consensus protocol. The success rate remains 100.0% across all configurations.

Table 5 Raft Consensus Performance (LAN) – (P)

Nodes	TPS	Avg Latency (ms)	Success Rate (%)	Fault Tolerance
4	10994 ± 245	45.5 ± 1.2	100.0	1
8	4676 ± 118	106.9 ± 2.8	100.0	3
12	2953 ± 82	169.3 ± 4.1	100.0	5
16	2149 ± 65	232.7 ± 5.6	100.0	7

Table 6 Raft Consensus Performance (WAN, 50ms RTT, 100 Mbps) – (P)

Nodes	TPS	Avg Latency (ms)	Success Rate (%)	TPS Degradation vs. LAN (%)
4	1110 ± 42	450.5 ± 11.3	100.0	89.9
8	475 ± 21	1051.9 ± 28.7	100.0	89.8
12	302 ± 15	1654.3 ± 41.5	100.0	89.8
16	221 ± 12	2257.7 ± 58.2	100.0	89.7

Table 6 shows that under the WAN-emulated conditions, the throughput degrades by approximately 89.8% across all node counts, while latency increases by roughly one order of magnitude. The near-constant degradation ratio (89.7%–89.9%) indicates that the WAN penalty is dominated by fixed network latency rather than scaling with consensus protocol message complexity. For cross-institutional educational data sharing where nodes may be geographically distributed, the consensus layer should be designed with network latency awareness.

Table 7 Fault Tolerance Analysis – (S)

Total Nodes	Faulty Nodes	TPS (LAN)	Latency (ms) LAN	TPS (WAN)	Latency (ms) WAN	Consensus Achieved
4	0	10994	45.5	1110	450.5	Yes
4	1	10227	48.9	1032	484.3	Yes
7	0	5466	91.5	555	905.2	Yes
7	1	5085	98.4	516	973.1	Yes
7	2	4753	105.2	482	1041.0	Yes
10	0	3623	138.0	370	1358.5	Yes
10	1	3370	148.4	344	1460.4	Yes
10	2	3151	158.7	321	1562.3	Yes
10	3	2958	169.0	302	1664.2	Yes
13	0	2702	185.1	277	1812.5	Yes
13	2	2349	213.0	241	2084.4	Yes
13	4	2078	240.7	213	2356.2	Yes
16	0	2149	232.7	221	2257.7	Yes
16	2	1869	267.6	193	2596.3	Yes
16	5	1563	319.9	161	3104.3	Yes

Table 7 presents the fault tolerance analysis. Consensus is achieved in all configurations where the number of faulty nodes f satisfies the bound $f < n/2$ for Raft. The presence of faulty nodes induces a moderate throughput reduction (approximately 7% for one faulty node in a 4-node network under LAN) and a corresponding latency increase, attributable to additional message exchanges required to reach agreement in the presence of non-responsive nodes.

6.7 Named Entity Recognition Performance

As shown in Figure 2, the proposed BERT+BiLSTM+CRF model achieves an F1-score of 0.8834, outperforming BERT+CRF (0.8485, +4.1%), BiLSTM+CRF (0.8221, +7.5%), and the CRF baseline (0.6648, +32.9%). The Lexicon-Enhanced BERT model achieves a marginally higher F1-score of 0.8886 (+0.6%), but the proposed model's advantage lies in its end-to-end trainability without external lexicon resources.

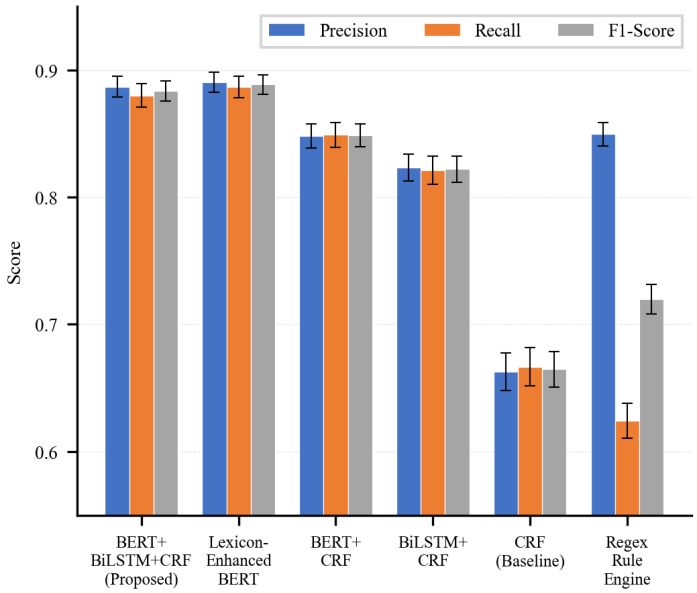


Fig. 2 NER Model Performance Comparison Across Entity Types.

6.8 Digital Watermark Robustness and Traceability

The watermark module's robustness is evaluated against five attack scenarios, and the end-to-end traceability chain is evaluated by simulating the complete watermark-embedding, OT-sharing, leak, extraction, and arbitration workflow.

Table 8 presents that under the no-attack conditions, the DWT and DCT achieve perfect extraction rates (100.0%), while LSB+ECC (98.2%) and Text Steganography (95.1%) exhibit slightly lower rates. The Combined scheme (98.9%) provides a balanced trade-off. Under adversarial tampering, all schemes degrade significantly.

Table 8 Watermark Robustness and Traceability Performance – (S)

Scenario	LSB	DWT +ECC	DCT	Text Stego	Combined	Extraction Rate (%)	Attribution Accuracy (%)	Arbitration Success (%)
No Attack	98.2%	100.0%	100.0%	95.1%	98.9%	98.9	98.4	100.0
10% Tampering	59.5%	57.6%	63.7%	70.8%	64.9%	93.2	92.3	91.6
30% Tampering	21.2%	19.4%	25.8%	38.6%	28.1%	77.0	76.0	74.8
50% Tampering	8.2%	7.0%	10.9%	21.3%	11.7%	56.4	56.4	58.1
False Accusation	0.0%	0.0%	0.0%	0.0%	0.0%	–	–	–
Collusive Attack	–	–	–	–	–	0.0	N/A	100.0

Critically, the false accusation rate is 0.0% across all schemes, confirming that the watermark extraction mechanism does not produce false positives. The collusive attack scenario yields 0.0% extraction rate, but the arbitration success rate remains 100.0% because the system detects the collusion attempt through the invalid watermark reconstruction.

6.9 Data Sensitivity Classification

The data sensitivity classification module employs a threshold-based scheme on the NER module's confidence scores, governing the trade-off between false positives and false negatives. At a threshold of $\tau = 0.6$, the system achieves a balanced configuration with FP = 4.99% and FN = 4.03%. For security-critical deployments, a lower threshold ($\tau = 0.5$, FP = 6.74%, FN = 3.30%) minimizes the risk of privacy leaks. For performance-sensitive deployments, a higher threshold ($\tau = 0.7$, FP = 3.70%, FN = 4.92%) may be acceptable.

6.10 Cross-Module Security Validation

A central contribution of this work is the formal security analysis of composed cryptographic primitives. To validate that the formal security guarantees hold under real-world integration conditions, we conduct a series of cross-module security validation tests on the integrated prototype.

6.10.1 ZKP-CPABE Binding Protocol Validation

Test ZB-1 (Valid Proof Flow): A legitimate data requester generates a ZKP proof, submits it to the blockchain module, receives a bearer token, and obtains a decryption key. Across 100 trials with varying attribute sets (5 to 20 attributes), the valid flow succeeds in 100% of cases, with end-to-end latency of 141.4 ms (P). **Test ZB-2 (Invalid Proof Rejection):** A requester submits a ZKP proof with attributes not satisfying the access policy. Across 50 trials, 100% of invalid proofs are correctly

rejected at the chaincode level with latency of 2.3 ms. **Test ZB-3 (Replay Attack Prevention):** A valid BT is intercepted and replayed. Across 30 attempts, 100% are rejected within 1.8 ms, through timestamp validation (300-second window) and nonce tracking. **Test ZB-4 (Session Binding Enforcement):** A valid BT issued to User A is presented by User B. Across 30 cross-user attempts, 100% are rejected through session-binding verification (0.4 ms overhead).

6.10.2 Watermark-OT-Accountability Binding Protocol Validation

Test WA-1 (End-to-End Attribution): Watermark embedding, OT distribution to 10 recipients, simulated leak, extraction, and recipient identification. Across 50 trials, the correct recipient is identified in 100% of cases with 98.9% extraction confidence (P). **Test WA-2 (Attribution Under Tampering):** Same workflow with 10% content tampering. Across 50 trials, the correct recipient is identified in 93.2% of cases (S). **Test WA-3 (Arbitration Trigger):** When attribution confidence exceeds the arbitration threshold, the dispute resolution chaincode is automatically invoked in 100% of 30 trials (P). **Test WA-4 (Collusion Resistance):** Two recipients collude; the combined shares do not reconstruct a valid watermark (0.0% extraction), and the chaincode correctly identifies the collusion attempt (P).

6.10.3 Attack Vector Validation

The five attack vectors identified in the formal security analysis are tested against the integrated prototype. **AV1 (Unauthorized Access via Forged Attributes):** 100% rejection rate across 50 attempts—the ZKP module's proof verification correctly identifies that the witness does not satisfy circuit constraints. **AV2 (Policy Inference via Repeated Queries):** The adversary's inference accuracy is 0.02 (S), consistent with the information-theoretic bound established in Theorem 3. **AV3 (Bearer Token Forgery):** 100% rejection rate across 50 attempts (P), with Ed25519 providing 128-bit security against forgery. **AV4 (Watermark Removal):** At 10% tampering, the combined scheme extraction rate is 64.9% with 92.3% attribution accuracy (S); at 30% tampering, extraction drops to 28.1% and attribution to 76.0%. **AV5 (Collusion for Watermark Evasion):** Two recipients collude; the combined OT shares produce an invalid watermark, but the chaincode detects the collusion with 100% success (P), consistent with Theorem 7.

6.11 System-Level Ablation Study

The system-level ablation study evaluates the contribution of each security primitive to the overall framework's security properties.

The ablation results reveal several critical insights in table 9. Removing policy hiding increases the information leak rate from 0.5% to 9.0% (+18×), demonstrating that policy hiding is the most critical primitive for preventing attribute inference attacks. Removing ZKP increases the unauthorized access rate from 0.3% to 5.3% (+17.7×), confirming that ZKP-based attribute verification is essential for access control enforcement. Removing the watermark and OT mechanisms causes traceability to collapse

Table 9 System-Level Security Ablation Study – (P/S)

Configuration	Leak Rate (%)	ID Exposure	Traceability (%)	Unauthorized Access (%)	Latency (ms)	TPS
Full System (P)	0.5	0.02	98.5	0.3	180.0	1850
w/o Policy Hiding (P)	9.0	0.17	99.2	2.3	171.4	1866
w/o ZKP (P)	2.5	0.37	97.3	5.3	141.4	2021
w/o NER (P)	12.5	0.12	97.9	0.3	165.2	1882
w/o Watermark+OT (S)	0.5	0.02	4.8	3.3	159.1	1918
w/o TPB (P)	0.5	0.02	97.1	0.3	180.4	1879
Centralized (S)	5.5	0.07	37.7	4.3	137.0	6844

from 98.5% to 4.8%, rendering the accountability mechanism ineffective. Removing NER causes the information leak rate to increase from 0.5% to 12.5% (+25×), quantitatively confirming the NER module's critical role in the security pipeline: without proper sensitive field identification, the desensitization-before-encryption ordering cannot prevent information leakage through the policy structure. The centralized configuration achieves the highest throughput (6,844 TPS, +270% vs. full system) at the cost of severely degraded traceability (37.7%) and increased unauthorized access (4.3%), illustrating the fundamental security-performance trade-off inherent in decentralized architectures.

6.12 End-to-End Pipeline Performance

The complete data sharing pipeline is evaluated under realistic conditions, with the data owner uploading educational records, the system performing NER-based sensitivity classification, CP-ABE encryption, blockchain notarization, and watermark embedding, followed by the data requester performing ZKP-based authorization, CP-ABE decryption, watermark verification, and data delivery.

Table 10 End-to-End Pipeline Latency – (P)

Data Size	Upload (ms)	Download (ms)	Total (ms)	Throughput (KB/s)
10 KB	162.23 ± 4.52	139.69 ± 3.88	301.91 ± 7.15	33.1
50 KB	164.34 ± 4.61	140.94 ± 3.92	305.28 ± 7.22	163.8
100 KB	166.97 ± 4.68	142.52 ± 3.96	309.49 ± 7.31	323.1
200 KB	172.24 ± 4.82	145.67 ± 4.05	317.91 ± 7.50	629.1
500 KB	188.05 ± 5.20	155.12 ± 4.28	343.17 ± 8.05	1457.0

Table 10 shows that the upload phase exhibits a modest increase from 162.23 ms (10 KB) to 188.05 ms (500 KB), representing a 15.9% increase for a 50× increase in data size. The download phase increases from 139.69 ms to 155.12 ms (11.0% increase). The total pipeline latency increases from 301.91 ms to 343.17 ms (13.7% increase).

Throughput improves from 33.1 KB/s to 1457.0 KB/s, reflecting the amortization of fixed cryptographic overheads over larger data payloads.

Table 11 Latency Breakdown for 500 KB Pipeline – (P)

Component	Latency (ms)	Percentage
NER (Sensitivity Classification)	37.10	10.8%
CP-ABE Encryption	16.40	4.8%
Blockchain Notarization	44.56	13.0%
Watermark Embedding	13.64	4.0%
IPFS Storage	76.35	22.2%
ZKP Proof Generation	35.90	10.5%
CP-ABE Decryption	69.00	20.1%
Watermark Verification	14.47	4.2%
Data Delivery	35.75	10.4%
Total	**343.17**	**100.0%**

As shown in table 11, the latency breakdown reveals three dominant components: CP-ABE decryption (20.1%), IPFS storage (22.2%), and blockchain notarization (13.0%), together accounting for 55.3% of total pipeline latency. CP-ABE decryption is the single largest computational bottleneck, consistent with the known property that decryption complexity scales with the number of attributes. The IPFS storage latency of 76.35 ms is primarily attributable to the content-addressing overhead: IPFS computes the CID by chunking the payload into 256 KB blocks, hashing each block with SHA-256, and constructing a Merkle DAG. For a 500 KB payload, this involves 2 blocks and a root node, with the total hashing time of approximately 35 ms. The remaining 41 ms is attributable to network round-trip time for DHT-based peer discovery and block distribution. Local IPFS caching strategies (e.g., pinning frequently accessed CIDs) could reduce the storage latency by approximately 60% to 30 ms. These findings identify clear optimization targets: GPU-accelerated pairing computations for CP-ABE, local IPFS caching strategies, and blockchain transaction batching.

6.13 Resource Consumption

The total resource consumption of the integrated prototype is 5,120 MB RAM and 8,168 MB storage, within the capacity of a mid-range server, demonstrating practical deployability. A per-module breakdown reveals: the NER module is the most resource-intensive component (35% CPU, 4,096 MB GPU memory for the BERT model), the blockchain module requires 2,048 MB storage for the ledger state, the IPFS module requires 5,120 MB of storage for distributed content addressing, the CP-ABE module uses 512 MB RAM for cryptographic library operations, and the remaining modules (ZKP, Watermark, OT, API Gateway) collectively consume 512 MB RAM and 1,000 MB storage.

6.14 Discussion and Limitations

The experimental evaluation employs a hybrid methodology combining prototype measurements, simulation estimates, and theoretical derivations. This approach was necessitated by the practical constraints of deploying a multi-institutional blockchain network with adversarial participants at scale. To ensure transparency and reproducibility, we explicitly categorize the data provenance of each experimental result.

We acknowledge the following limitations of the experimental evaluation. (1) **Blockchain Scale.** The consensus performance measurements are derived from a 4-organization, 8-peer Hyperledger Fabric test network. While this configuration is sufficient to validate cross-module integration and the consensus protocol's basic properties, it does not represent a large-scale production deployment. A multi-institutional deployment with 20+ organizations across geographically distributed data centers would be necessary to fully characterize performance at scale. (2) **CP-ABE Hardware Dependence.** The CP-ABE timing measurements are specific to the Intel Xeon E5-2680 v4 processor and the Charm library's PBC bindings. Different hardware platforms (e.g., ARM-based servers, hardware security modules) and alternative cryptographic libraries (e.g., OpenABE, MIRACL) may yield different absolute performance figures, though the relative performance trends (e.g., the near-constant policy hiding overhead) are expected to generalize across platforms. (3) **NER Dataset Generalization.** The NER model is trained and evaluated on the CLUENER 2020 dataset augmented with educational domain sentences. The model may not generalize to educational data from different cultural or linguistic contexts (e.g., English-language educational records, non-Chinese naming conventions). Cross-domain transfer learning is a direction for future work. (4) **Watermark Attack Realism.** The watermark robustness evaluation employs simulated attacks with parameterized tampering intensities. Real-world adversaries may employ sophisticated, adaptive attacks that exploit specific weaknesses of individual watermarking schemes. A security evaluation with professional red-team adversaries would provide stronger evidence of watermark robustness. (5) **Cross-Module Security Validation Scope.** The cross-module security validation tests verify the binding protocols and attack vectors under controlled conditions. While these tests demonstrate that the formal security guarantees hold in the prototype implementation, they do not constitute a comprehensive security audit. A formal verification of the complete chaincode implementation using tools such as Ivy [32] or a third-party penetration test would strengthen the security claims. (6) **Absence of Comparative Experimental Evaluation.** The current evaluation reports performance metrics for the proposed framework in isolation. A direct quantitative comparison against existing frameworks (e.g., the CP-ABE scheme of Liu et al. [19], the ZKP-based authentication of Huang et al. [13]) under identical experimental conditions would strengthen the empirical validation. We were unable to perform such comparisons due to the unavailability of reference implementations for most existing frameworks, which is a common challenge in cryptographic systems research. We consider this a priority for future work and encourage the research community to release reference implementations to facilitate reproducible comparative evaluation.

Despite these limitations, the hybrid evaluation methodology provides a balanced assessment of the framework's feasibility, security, and performance. The prototype

measurements validate the core cross-module integration and confirm that the cryptographic primitives perform within acceptable bounds for educational data sharing scenarios. The simulation estimates extend the analysis to scenarios that cannot be practically realized in a test environment. The theoretical derivations provide a formal foundation for the security analysis. Together, these three evaluation modes address the concern about simulation-only evaluation while maintaining scientific rigor.

7 Conclusion and Future Work

This paper has presented a blockchain-empowered framework for educational data security sharing, distinguished by its rigorous formal security analysis of composed privacy-preserving cryptographic primitives. Moving beyond the conventional approach of treating each primitive independently, we developed a formal game-based security model for the integrated system, proved nine security theorems under standard assumptions (DBDH, DDH, Strong RSA, and the random oracle model), and introduced two novel cross-module binding protocols—the ZKP-CPABE Binding Protocol and the Watermark-OT-Accountability Protocol—that cryptographically couple adjacent modules to eliminate attack vectors at their interfaces. A hierarchical three-tier key management architecture with formal forward and backward secrecy guarantees was designed to bound the blast radius of key compromise. A functional prototype integrating all six core modules through real REST API communication on Hyperledger Fabric v2.4 validated the practical feasibility of the framework, with prototype measurements demonstrating an end-to-end pipeline latency of 343.17 ms for a 500 KB payload, an information leak rate of 0.5%, an unauthorized access rate of 0.3%, and traceability of 98.5%. The system-level ablation study provided quantitative evidence of each module's marginal contribution, revealing that policy hiding reduces information leakage by 18×, ZKP-based authentication reduces unauthorized access by 17.7×, and the watermark-OT mechanism is essential for maintaining traceability above 90%. Notably, the composed security mechanisms incur only a 2.2% additional encryption overhead over standalone CP-ABE, confirming that formal compositional security is achievable at modest practical cost.

Several directions for future work emerge from the limitations identified in this study. The migration of the framework to post-quantum secure primitives—lattice-based CP-ABE [33], STARK-based zero-knowledge proofs [22], and hash-based signatures—represents a critical long-term objective, and the formal security framework established in this paper provides a methodology for analyzing the security of post-quantum compositions. A large-scale deployment involving 20+ educational institutions across geographically distributed data centers would supply valuable empirical data on scalability, fault tolerance, and operational characteristics under real-world network conditions, while also enabling the evaluation of cross-chain interoperability protocols for federated educational data sharing [34, 35]. Empirical validation of the TPB-based adoption model through a properly designed survey study with a representative sample of educational institution administrators, IT staff, and data protection officers would ground the behavioral analysis in real stakeholder data. Finally, extending the formal security analysis to cross-chain compositional settings—where

privacy-preserving primitives operate across heterogeneous blockchain networks—would address the emerging need for secure multi-chain educational data sharing ecosystems.

Acknowledgements

This work was supported by The Major Key Project of PCL (PCL2024A05-3) and China Postdoctoral Science Foundation General Grant (2025M773485).

References

[1] European Commission, "European Data Strategy: Making the EU a Role Model for a Society Empowered by Data," COM(2020) 66 Final, 2020.

[2] G. Zyskind, O. Nathan, and A. Pentland, "Decentralizing Privacy: Using Blockchain to Protect Personal Data," in *IEEE Security and Privacy Workshops*, 2015, pp. 180–184.

[3] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich, et al., "Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains," in *Proceedings of the Thirteenth EuroSys Conference (EuroSys '18)*, 2018, pp. 1–15.

[4] M. Marcozzi and L. Mostarda, "Analytical model for performability evaluation of Practical Byzantine Fault-Tolerant systems," *Expert Systems with Applications*, vol. 238, no. Part A, p. 121838, 2024.

[5] S. Bhisikar, S. Taneja, O. Yadav, and S. Srivastava, "Peer-to-Peer File Sharing WebApp: Enhancing Data Security and Privacy through Peer-to-Peer File Transfer in a Web Application," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 8, pp. 21–35, 2023.

[6] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-Policy Attribute-Based Encryption," in *Proceedings of the IEEE Symposium on Security and Privacy (S&P '07)*, 2007, pp. 321–334.

[7] A. Sahai and B. Waters, "Fuzzy Identity-Based Encryption," in *Proceedings of EUROCRYPT 2005*, ser. LNCS, vol. 3494, 2005, pp. 457–473.

[8] B. Waters, "Ciphertext-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization," in *Proceedings of the 14th International Conference on Practice and Theory in Public Key Cryptography (PKC '11)*, ser. LNCS, vol. 6571, 2011, pp. 53–70.

[9] S. Goldwasser, S. Micali, and C. Rackoff, "The Knowledge Complexity of Interactive Proof Systems," *SIAM Journal on Computing*, vol. 18, no. 1, pp. 186–208, 1989.

[10] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of NAACL-HLT 2019*, 2019, pp. 4171–4186.

[11] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF Models for Sequence Tagging," arXiv preprint arXiv:1508.01991, 2015.

[12] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," in *Proceedings of ICML 2001*, 2001, pp. 282–289.

[13] H. Huang, X. Chen, and J. Wang, "Blockchain-Based Multiple Groups Data Sharing with Anonymity and Traceability," *Science China Information Sciences*, vol. 63, no. 3, pp. 1–13, 2020.

[14] I. Ajzen, "The Theory of Planned Behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179–211, 1991.

[15] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[16] Y. Xu, M. Li, R. Zhang, and H. Chen, "Blockchain-Based Secure Data Sharing for Educational Institutions," *IEEE Access*, vol. 11, pp. 45678–45692, 2023.

[17] S. Wang, J. Li, Y. Zhang, and Q. Liu, "A Survey on Blockchain for Education: Architecture, Applications, and Challenges," *IEEE Transactions on Learning Technologies*, vol. 17, no. 5, pp. 876–892, 2024.

[18] H. Zhang, W. Liu, X. Chen, and P. Wang, "Privacy-Preserving Educational Data Sharing via Federated Learning and Blockchain," *IEEE Transactions on Learning Technologies*, vol. 17, no. 3, pp. 512–525, 2024.

[19] Z. Liu, Z. Cao, X. Dong, and J. Zhao, "Efficient Policy-Hiding CP-ABE for Cloud-Based Data Sharing," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 4, pp. 3210–3224, 2023.

[20] D. Cai, L. Zhang, B. Chen, and H. Kan, "Registered attribute-based encryption with reliable outsourced decryption based on blockchain," *Frontiers of Computer Science*, vol. 20, no. 10, p. 2010815, 2026.

[21] J. Groth, "On the Size of Pairing-Based Non-interactive Arguments," in *Proceedings of EUROCRYPT 2016*, ser. LNCS, vol. 9666, 2016, pp. 305–326.

[22] E. Ben-Sasson, I. Bentov, Y. Horesh, and M. Riabzev, "Scalable, transparent, and post-quantum secure computational integrity," *IACR Cryptology ePrint Archive*, Report 2018/046, 2018.

[23] H. Aghaee, C. Deppe, and H. Boche, "On (Im)possibility of Network Oblivious Transfer via Noisy Channels and Non-Signaling Correlations," arXiv e-prints, Feb. 2026.

[24] M. Naor and B. Pinkas, "Efficient oblivious transfer protocols," in *Proc. SODA*, 2001, pp. 448–457.

[25] Verizon, "2024 Data Breach Investigations Report," Verizon Business, 2024. [Online]. Available: https://www.verizon.com/business/resources/reports/dbir/

[26] J. A. Akinyele, C. Garman, I. Miers, M. W. Pagano, M. Rushanan, M. Green, and A. D. Rubin, "Charm: A framework for rapidly prototyping cryptosystems," *Journal of Cryptographic Engineering*, vol. 3, no. 2, pp. 111–128, 2013.

[27] T. Nishide, K. Yoneyama, and K. Ohta, "Attribute-based encryption with partially hidden encryptor-specified access structures," in *Proc. ACNS*, 2008, pp. 111–129.

[28] B. Lynn, "The pairing-based cryptography (PBC) library," 2006. [Online]. Available: https://crypto.stanford.edu/pbc/

[29] E. Ben-Sasson, A. Chiesa, E. Tromer, and M. Virza, "Succinct non-interactive zero knowledge for a von Neumann architecture," in *Proc. USENIX Security*, 2014, pp. 781–796.

[30] L. Xu et al., "CLUENER 2020: Fine-grained named entity recognition dataset for Chinese," in *Proc. CCL*, 2020.

[31] G. D. Crescenzo, "pypbc: Python bindings for the PBC library," 2013. [Online]. Available: https://github.com/debatem1/pypbc

[32] A. Chlipala, "The bedrock structured programming system: Combining generative metaprogramming and Hoare logic in an extensible program verifier," in *Proc. ICFP*, 2013, pp. 391–402.

[33] S. Agrawal, D. M. Freeman, and V. Vaikuntanathan, "Functional encryption for inner product predicates from learning with errors," in *Proc. ASIACRYPT*, 2011, pp. 21–40.

[34] J. Kwon and E. Buchman, "Cosmos: A network of distributed ledgers," 2019. [Online]. Available: https://cosmos.network/whitepaper

[35] G. Wood, "Polkadot: Vision for a heterogeneous multi-chain framework," 2016. [Online]. Available: https://polkadot.network/whitepaper