Privacy-Preserving Education Data Sharing Scheme Based on Consortium Blockchain

Jiaqi Guo[1], Zhuoran Wang[2], Ningning Liu[3]*
School of Cyberspace Security, Xi'an University of Posts and Telecommunications, Xi'an 710121, China[1,2]
Information Technology Center, Hebei Open University, Shijiazhuang 050080, China[3]

***Abstract*—With the growing emphasis on lifelong education and the rapid expansion of open education platforms, the secure and efficient management and sharing of lifelong learning data have become critical challenges. To address these issues, this paper proposes a Privacy-Preserving Educational Data Sharing (PPEDS) scheme based on blockchain technology. The PPEDS scheme employs attribute-based encryption with hidden attributes to achieve privacy-preserving and fine-grained access control. In addition, it incorporates multi-keyword searchable encryption to enable efficient encrypted data retrieval and combines private and consortium blockchains to ensure data authenticity and integrity across multiple educational institutions. The security analysis demonstrates that the scheme resists potential attacks and ensures confidentiality, access control, and search privacy under a semi-trusted model. Furthermore, performance evaluations conducted on real-world educational datasets show that the proposed scheme achieves efficient encryption, search, and decryption operations, with low computational overhead even in large-scale deployments. Overall, the PPEDS scheme provides a secure, scalable, and practical solution for privacy-preserving data sharing in lifelong education systems.**

Keywords*—Consortium blockchain; access control; secure search; life-long education; data sharing

I. Introduction

Lifelong education integrates various forms of learning that individuals receive at different stages of life and is becoming increasingly important in the rapidly developing knowledge-based society [1]. With the advancement of information technology, open education platforms have become a powerful tool for promoting lifelong learning, providing flexible access to educational resources from different institutions and regions [2], [3]. However, traditional approaches to educational data management are hampered by problems such as isolated data silos, limited transparency, and fragmented information systems [4]. These limitations pose significant obstacles to the recognition and portability of learning achievements, thereby affecting the overall effectiveness of the lifelong learning ecosystems. With the rapid development of lifelong education and open education platforms, the management and sharing of educational data have become increasingly complex and critical. The flow of students between different educational platforms and educational institutions, especially the needs of cross-school elective courses, credit certification, transcript verification, etc., has made cross-platform sharing of educational data an urgent need. Therefore, secure data sharing and privacy protection are essential to promote the healthy development of information technology in lifelong education.

*Corresponding author.

Blockchain technology has the characteristics of decentralization, immutability and transparency [5] and has been widely regarded as a transformative solution for secure data management and sharing. Blockchain can achieve secure storage of data, access control and privacy protection, while improving the transparency and credibility of data sharing. Recent studies have explored various applications in fields such as healthcare and finance, aiming to address challenges in data security, privacy, and cross-institutional data interoperability [6], [7].

Although extensive research has explored secure data sharing based on blockchain in domains such as healthcare, transportation, and the Internet of Things (IoT), relatively few studies have focused specifically on educational data sharing. Unlike other application areas, educational data presents unique challenges. It often involves multiple stakeholders—including students, teachers, educational institutions, and government bodies—with complex relationships and frequent data updates. Additionally, educational data is heterogeneous in format, privacy-sensitive, and typically distributed across different platforms, making secure, traceable, and fine-grained access control particularly difficult. Moreover, practical requirements such as cross-institutional credit recognition, transcript verification, and course-based access control necessitate flexible and scalable sharing mechanisms that are not well addressed by existing solutions. In recent years, some progress has been made toward blockchain-enabled educational systems. Huang et al. [8] developed a blockchain-based education data management system that integrates private and consortium blockchains to enhance data security and traceability. Marouan et al. [9] proposed a blockchain-based framework for managing educational credentials and supporting lifelong learning, emphasizing decentralized trust and certificate verification. However, these works mainly focus on identity and credential management, lacking technical details on fine-grained access control, privacy-preserving queries, and efficient data retrieval. Therefore, there is an urgent need to develop a comprehensive and secure data-sharing scheme tailored for educational environments.

Overall, existing research provides a variety of secure and private data sharing methods, but there is still a lack of specific solutions for educational data sharing, and these solutions have certain limitations, such as the lack of fine-grained access control, efficient user decryption, and practical multi-keyword search. In response to these limitations, we propose a privacy-preserving education data sharing scheme based on consortium blockchain, named PPDES, providing a comprehensive solution for secure and efficient lifelong education data sharing.

The main contributions and novel features of this paper are as follows:

- By integrating private blockchain and consortium blockchain, the PPEDS scheme addresses the limitations of traditional data management, ensuring data legitimacy, traceability, and enabling secure data sharing across educational platforms.
- The PPEDS scheme achieves fine-grained access control through multi-authority CP-ABE, allowing each educational platform to manage user attributes independently, while hidden attributes and user anonymity enhance access security.
- Based on public key searchable encryption, the PPEDS scheme supports efficient multi-keyword search. After searching, the consortium blockchain can perform pre-decryption for users who meet access policies, significantly reducing user computation.

The rest of this paper is organized as follows: Section II discusses the system model, threat model, and design goals. Section III introduces the problem formulation. Section IV details the preliminaries. The PEDD system is given in Section V. Sections VI and VII discuss security and performance analyses, respectively. Finally, Section VIII concludes the paper.

II. Related Work

In recent years, blockchain technology has been widely applied in the field of privacy-preserving data sharing across various domains, including healthcare, IoT, and transportation.

Zheng et al. [10] proposed a scalable privacy-preserving data sharing scheme based on blockchain, which utilizes a (p, t) threshold Paillier cryptosystem to enforce user access control. Li et al. [11] introduced a blockchain-based IoT data sharing scheme with privacy protection and a reward mechanism, where signature exchange is used to implement access control. Li et al. [12] designed a privacy-preserving method for IoT medical systems based on blockchain and lightweight secret sharing, combining decentralization and data confidentiality to facilitate secure sharing among institutions. In another work, Li et al. [13] proposed a privacy-preserving flight operation data sharing scheme using zero-knowledge proofs and proxy re-encryption to realize secure access control.

Although these schemes adopt lightweight access control mechanisms, they generally lack fine-grained expressiveness and exhibit limited efficiency when applied to educational environments involving a large number of users and roles.

To address fine-grained access control, Liang et al. [14] presented a personal data protection framework based on consortium blockchain and attribute-based encryption (ABE), enabling secure data sharing among multiple platforms. However, the issue of searchable encryption was not considered in this work, which is crucial for efficient data access in large-scale systems.

In terms of searchable encryption, Jiang et al. [15] proposed an efficient and privacy-preserving data sharing scheme in intelligent transportation systems, employing Bloom filters to support multi-keyword retrieval. Du et al. [16] designed a blockchain-based searchable encryption scheme using both public and private blockchains, where encrypted indexes reside on the private chain and documents are outsourced to the public chain. Their scheme also incorporates access control to improve query efficiency. Niu et al. [17] presented an attribute-searchable scheme based on CP-ABE and blockchain, supporting attribute hiding and multi-keyword search. Liu et al. [18] developed a privacy-preserving medical data sharing framework based on blockchain and ABE, which supports both fine-grained access control and secure keyword search among medical institutions.

However, most of the above solutions have the problem of low efficiency of user decryption or lack of the ability of attribute authorities to independently manage attributes.

While the above works span various application fields such as medical care, IoT, and transportation, there is relatively little research focused on blockchain-based educational data sharing. To fill this gap, Luo et al. [19] propose a blockchain-based education data management system that leverages the strengths of private and consortium blockchains to enhance system security. However, the scheme lacks basic access control and keyword retrieval. Huang et al. [8] developed a blockchain-based educational data management system that leverages both private and consortium blockchains to enhance security. In [9], Marouan et al. present a blockchain-based framework for educational credentials and lifelong learning. They explore the use of a decentralized ledger to issue, store, and verify educational credentials such as course completions, skills certificates, and learning records in a tamper-resistant manner. Their work emphasizes the trustworthiness and interoperability of credentials across institutions, highlighting how blockchain can mitigate fraud and simplify cross-platform verification. However, their schemes still lack mechanisms for fine-grained access control and efficient encrypted search. These limitations highlight the necessity of a specialized security framework tailored to educational data sharing scenarios. Our PPEDS scheme distinguishes itself by addressing these gaps through multi-authority CP-ABE, hidden attributes, efficient searchable encryption, and dual-chain integration.

III. Problem Formulation

A. System Model

The PPEDS scheme involves a large-scale open education platform composed of multiple provincial and municipal education platforms. Each educational platform has its own cloud server and local client, which is typically operated by teachers of the education platform. The registered students also have certain operational permissions, while other registered users only have the right to search and access some public courses. The transcripts and certificates of students, or some paid courses of teachers are encrypted and stored on the platform server. Each educational platform builds its own private blockchain to store the hashes of encrypted content and the keyword indexes. The private blockchains of all participating educational platforms together constitute a consortium blockchain, which is managed by all platforms for cross-institutional collaboration and data sharing. Each platform uploads part of its data, such as the identifier of the private blockchain and keyword indexes to the consortium blockchain.

The system mainly includes five entities: teachers, students, platform servers, private blockchains, and the consortium blockchain. The system model is illustrated in Fig. 1.

Teachers: The users of each education platform are simply divided into two roles: teachers and students. Teachers are typically the data managers of the platform, responsible for evaluating student assignments, exams, etc., and uploading encrypted transcripts and certificates to the platform server, along with generating corresponding keywords for retrieval. Teachers also need to upload their courses. Public courses do not need to be encrypted, and teachers generate keyword indexes and upload them directly. However, some courses are restricted to users with specific attributes. For example, a course of computer major may require students to satisfy {*"Major": "computer science" AND "Passed course": "C language"*} to access it. Teachers also need to generate hash values and signatures for encrypted transcripts or courses and upload them to the private blockchain to generate and broadcast this new transaction. Other nodes on the private blockchain are responsible for verifying the transaction, and if the verification passes, a new block is generated on the private blockchain.

Students: When students search for their grades or courses, they generate the corresponding search trapdoor using the keys assigned to them during registration and the related keywords. To achieve cross-platform interoperability, the search trapdoor is uploaded to the consortium blockchain, where nodes on the consortium execute the search algorithm. Students can only search and access their own transcripts and certificates, and apart from public courses, they can only access courses that are part of their study plan. Grades between different educational platforms can be mutually recognized. To avoid performing complex decryption operations locally, students can generate a temporary key during searches, allowing the platform server to perform pre-decryption operations for the authorized student. In practice, teachers or other registered users can also perform searches and access data, but they should all follow the access permissions of the data owner.

Platform Server (PS): Platform servers are typically large storage and computing servers in the cloud, which are responsible for storing all courses and related content. After storing encrypted transcripts or courses, the PS extracts the block identifiers from the private blockchain, along with the keyword indexes, and generates a signature to construct a new transaction on the consortium blockchain. Other nodes on the consortium blockchain are responsible for verifying the transaction, and if the verification passes, a new block is generated on the consortium blockchain. When the searched content is encrypted data, PS is responsible for pre-decrypting it for the user before returning the content.

Private Blockchain (PB): Teachers upload the hash values of encrypted transcripts or courses, keyword indexes, and their signatures to the PB, creating a new transaction. Nodes on the PB verify the signature to ensure it was legally generated by the teacher, and then generate a new block on the PB. The data structure of PB is shown in Fig. 2.

Consortium Blockchain (CB): The nodes on the CB are responsible for executing the search algorithm. If the search is successful, the nodes on CB extract the security index from the block to obtain the block identifier of the PB. Using the block identifier from the PBn, they retrieve the ciphertext hash and provide it to the corresponding PS. After confirming hash consistency, the PS uses the temporary key provided by the user to perform pre-decryption on the ciphertext and returns the pre-decrypted content to the consortium blockchain nodes, which then return to the search user. The data structure of CB is shown in Fig. 3.

B. Threat Model

Assuming the educational platform servers and the computers of teachers or students are semi-trusted, which means that they are trusted in most cases but may still be at risk of being compromised. Attackers may eavesdrop on the information being transmitted through communication channels, such as security indexes, encrypted course content, and search trapdoors. Additionally, clients and servers are not allowed to collude to infer the real identity of users.

The PB is managed by multiple administrators within the educational platform. This multi-administrator setting prevents single-point control and enhances the security of the private blockchain. In this semi-trusted environment, each administrator is assumed to be independently trusted, and they do not collude to tamper with data or engage in malicious behaviors.

The CB is managed by multiple educational platforms, with each platform participating in its management and maintenance, sharing part of the data, and achieving cross-platform interconnection. We assume that the nodes of these platforms do not collude with each other to ensure the security and integrity of the consortium blockchain data.

C. Design Goals

Based on the semi-trusted assumptions and potential attack risks described in Section 2.3, the security objectives are as follows:

1) Data security and access control: Sensitive data must be encrypted, and data owners should have the freedom to define access control policies, ensuring that only users who meet the specified criteria can access the data.

2) Privacy protection: Attributes in access control policies must not reveal sensitive user information, and ciphertext should be resistant to DDH testing to ensure recipient anonymity.

3) Secure search: Indexes and queries should remain encrypted, and searches should be performed by designated trusted nodes to prevent attackers from obtaining the true keywords during the search process.

4) System availability: Consistency in the private blockchain is achieved through the verification of signatures from internal administrators, while the consortium blockchain's consistency is maintained through platform signature verification, neither of which involves sensitive information.

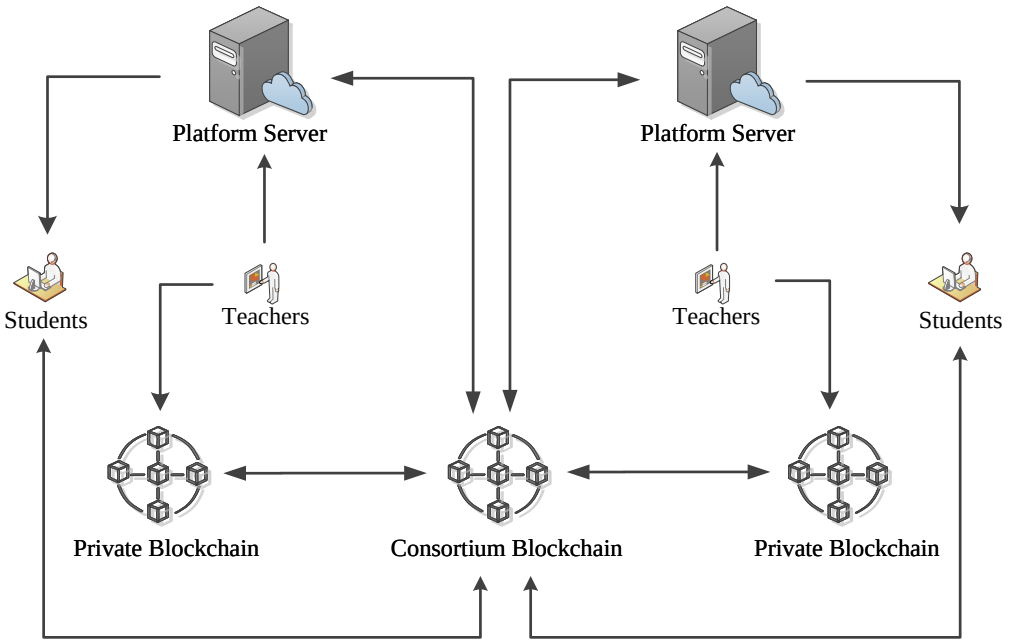

Fig. 1. System model.

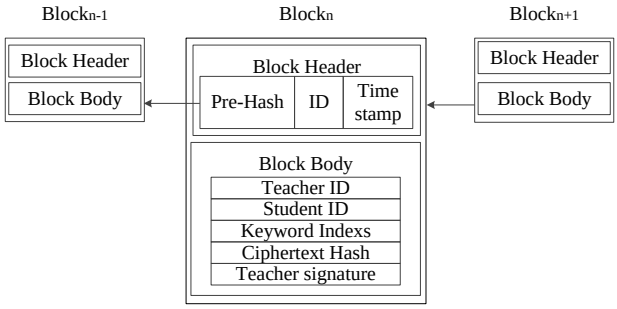

Fig. 2. Data structure of private blockchain.

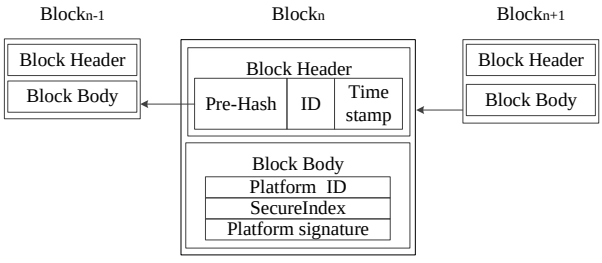

Fig. 3. Data structure of consortium blockchain.

IV. Preliminaries

A. Attribute-Based Encryption

Attribute-Based Encryption (ABE) [20] is a public key encryption scheme that allows users to encrypt and decrypt data based on a set of attributes. The main advantage of ABE is that it enables fine-grained access control and allows flexible data access management. There are two main types of ABE: Key-Policy ABE (KP-ABE) [20] [21] and Ciphertext-Policy ABE (CP-ABE)[22].

- KP-ABE: In KP-ABE, the access policy is embedded in the secret key, while the ciphertext is labeled with attributes.
- CP-ABE: In CP-ABE, the access policy is embedded in the ciphertext, while the secret key is associated with attributes. This allows the data owner to define who can access the encrypted data, making CP-ABE well-suited for fine-grained access control.

Formally, let G be a cyclic group of prime order p, and g be its generator. During the system setup, public parameters and a master key are generated. For each attribute i, a secret key component is generated using a random value. The decryption key is derived based on the attributes they possess. The encryption function takes a message M, an access policy A, and outputs a ciphertext C. The decryption function allows a user to recover the message M if their attributes satisfy the policy A.

B. Searchable Encryption

Searchable Encryption (SE) [23] is a cryptographic Primitive that allows encrypted data to still support keyword-based search. SE enables the platform to store data in encrypted form while allowing users to search over it without revealing sensitive information. The main components of SE include:

1) Index encryption: The data owner encrypts the index keyword w as I_w. The encrypted index I_w is then uploaded to the server.

2) Trapdoor generation: To search for a keyword w', the user generates a trapdoor $T_{w'}$.

3) Search: The server uses the trapdoor T_w to search through the encrypted index $\{I_w\}$. When $w = w'$, the server returns the matching content to the user without learning the specific information of w or w'.

C. Blockchain

Blockchain [24] is a distributed ledger technology that provides data integrity, security, transparency, and immutability through decentralized consensus mechanisms. It allows records to be securely stored in linked blocks, making it nearly impossible to alter data once added.

Blockchain can be classified into three main types: public blockchain, private blockchain, and consortium blockchain. Public blockchains are open to everyone, allowing anyone to participate in the consensus process. Private blockchains are restricted to specific participants, typically managed by a single organization, and provide more control over the data. Consortium blockchains are semi-decentralized, managed by a group of institutions, and are ideal for collaborative projects.

V. The Privacy-Preserving Education Data Sharing (PPEDS) Scheme

The PPEDS scheme specifically includes setup, key generation, encryption, blockchain generation, trapdoor generation, search, and decryption. The specific description of each algorithm is as follows.

A. Setup

The setup of the system is to generate some necessary public parameters. First, it selects a security parameter κ. Then, a multiplicative cyclic group G of prime number p order is selected, where g is its generator. Each user entering the system is assigned an identity ID, which can be composed of any string. The ID can be mapped to G using the function H_1. Each user role has its attributes, such as $\{\text{“}major\text{”} : \text{“}mathematics\text{”}\}$, and they are all expressed in the form of $\{attributename : attributevalue\}$. The function H_1 can also map the attribute value to G. Define a function H_2 to map the attribute back to the authority that manages the attribute. In general, the public parameters include $\{p, G, g, H_1, H_2\}$.

Next, each authority generates a pair of public and private keys. Assume that the authority picks four random numbers $\alpha, \beta, \gamma, \sigma \in Z_p$ as its private key. Then, it calculates $e(g,g)^{\alpha}, g^{\beta}, g^{\gamma}, g^{\sigma}$ as its public key and publishes them.

B. KeyGen

Each authority (i.e., each educational platform) distributes corresponding attribute keys to registered users with different roles (teachers, students, administrators, etc.) in the system.

Suppose an authority needs to generate an attribute key for a student named Alice, whose attribute is $v_\chi \in V_a$, where V_a is Alice's attribute set. Each authority responsible for V_a needs to generate a corresponding attribute key for the user. For example, if Alice is a teacher at platform PS_1 but is also a student in a specific course at platform PS_2, both PS_1 and PS_2 need to generate attribute private keys for Alice. Where ID_a is Alice's identifier, which can be composed of any string. We will use ID_a below to represent a student user like Alice.

$$\begin{aligned} & sk_{ID_a, v_\chi} \\ =&\{sk_1 = g^{\frac{\alpha}{\gamma}} H_1(ID_a)^{\frac{\beta}{\gamma}} H_1(v_\chi)^{\frac{\sigma}{\gamma}}, \\ & sk_2 = H_1(v_\chi)^{\frac{\sigma}{s}}, sk_3 = g^{\sigma}\} \end{aligned} \tag{1}$$

Obviously, all the keys related to $v_\chi \in V_a$ composes the attribute key $sk_a = \{sk_{ID_a,v}\}$. Each successfully registered student will receive a search key s distributed by the system to generate a search trapdoor.

C. Enc

The teacher extracts keywords (such as student name, ID, and subject) from course grades and encrypts them, and also extracts keywords from some non-public courseware and other learning content and encrypts them before uploading them to the platform server. Attribute-based encryption ensures that only students who meet the access policy specified by the teacher can decrypt the content.

1) FileEnc: Suppose the teacher needs to upload Alice's mathematics transcript T_a and only allow Alice and other teachers can access it. For the sake of generalization, assume that this specific access policy is represented by (A, f, Val_p). Here $A \in Z_p^{c \times d}$ is an access matrix, and f is a function that maps rows in A to attribute names. Let $Val_p = v_{f(1)}, \cdots, v_{f(c)}$ represent the corresponding attribute value. Among them, Val_p contains specific attribute values, such as the student's name or ID, which needs to be kept private, and the other parts of the access policy need to be uploaded with the ciphertext.

A function η is also defined to map the rows of A to the authorities, that is, $\eta(\cdot) = H_2(f(\cdot))$. Then, select some random numbers in Z_p and generate two vectors $\vec{a} = (z, a_2, \cdots, a_d)^{\top}$ and $\vec{b} = (0, b_2, \cdots, b_d)^{\top}$. According to LSSS [20], $a_i = A_i \vec{a}$ and $b_i = A_i \vec{b}$, where A_i represents the ith row of A.

For each $i \in [1, c]$ in A, randomly select $r_i, r_i' \in \mathbb{Z}_p$. The encrypted R is as follows:

$$\begin{aligned} & CT_0 = T_a e(g,g)^z, \\ \forall i \in [1,c], \quad & \{CT_{1,i} = g^{a_i} g^{\alpha_{\eta(i)} r_i}, CT_{2,i} = (g^{\gamma})^{-r_i}, \\ & CT_{3,i} = (g^{\delta})^{r_i - r_i'}, CT_{4,i} = g^{\beta_{\eta(i)} r_i} g^{b_i}, \\ & CT_{5,i} = H_1(v_{f(i)})^{r_i'}\} \end{aligned} \tag{2}$$

The ciphertext $CT = \{(A, f), CT_0, \{CT_{1,i}, CT_{2,i}, CT_{3,i}, CT_{4,i}, CT_{5,i}\}_{i \in [1,c]}\}$.

2) IndexEnc: The teacher selects several keywords such as $W = \{Alice, ID_a, Mathematics\}$, and choose a random number $\tau \in Z_p$ and encrypts them as:

$$\begin{aligned} I_1 &= e(g,g)^{\tau} \\ I_2 &= g^{s\tau} \\ I_w &= \{g^{\tau H(w)}\}_{w \in W} \end{aligned} \tag{3}$$

The encrypted index is $I = I_1, I_2, I_w$. Then the ciphertext CT is sent to PS, and the hash value of CT, the encrypted index I, and the signature of the teacher are all uploaded to the PB.

D. BlockGen

Teachers on each education platform upload data to their private chain, and each education platform integrates the data from the private chain into the alliance chain.

1) PriBlockGen: The teacher constructs a new transaction using the keyword index, teacher ID, and teacher signature, then broadcasts the transaction to the PB of the education platform. Upon receiving the new transaction, the verifiers on the PB verify the signature in the transaction. If it is valid and confirms that the content was legally created by an authorized teacher, a verification confirmation message is broadcast. When more than 2/3 of the verifiers confirm, the PB accepts the new transaction and generates a new block. Otherwise, the new block is rejected.

2) ConBlockGen: The platform server extracts the ID and keyword index from each new block of its PB to generate a security index. The PS constructs a new transaction using the security index, platform ID, and platform signature, then broadcasts the transaction to the CB. Upon receiving the new transaction, the verifiers on the CB verify the signature of PS. When more than 2/3 of the verifiers confirm, a new block is then generated, ensuring the legitimacy of the content across platforms.

E. TrapGen

When student ID_a wants to search for something of interest, such as her mathematics course transcript, a trapdoor T_w is generated, where w' is the keyword.

$$T_w = g^{\frac{1}{H(w')+s}} \tag{4}$$

If there is more than one keyword of interest, the trapdoor is generated as:

$$T_w = g^{\frac{1}{\sum_{w'=1}^{n} H(w')+s}} \tag{5}$$

If ID_a wants to get the pre-decrypted search results, she also needs to select a random number $t \in Z_p$ to generate a temporary key sk'_a and upload it with the trapdoor. Specifically, sk'_a is calculated as follows.

$$\begin{aligned} sk'_a =& \{sk'_{ID_a, v_\chi}\} \\ =& \{\{sk'_1 = sk_1^{\frac{1}{t}}, sk'_2 = sk_2^{\frac{1}{t}}, sk'_3 = sk_3^{\frac{1}{t}}\}, \\ & ID' = H_1(ID_a)^{\frac{1}{t}}, g^{\frac{1}{t}}\} \\ =& \{\{sk'_1 = g^{\frac{\alpha}{t\gamma}} H_1(ID_a)^{\frac{\beta}{t\gamma}} H_1(v)^{\frac{\sigma}{t\gamma}}, \\ & sk'_2 = H_1(v)^{\frac{\sigma}{t\delta}}, sk'_3 = g^{\frac{\sigma}{t}}\}, \\ & ID' = H_1(ID_a)^{\frac{1}{t}}, g^{\frac{1}{t}}\} \end{aligned} \tag{6}$$

F. Search

The node on the CB calculates and verifies whether the equation $e(I_2 \cdot \prod_{i=W} I_w, T_w) = I_1$ holds. If so, the Platform ID is obtained. The corresponding PB is accessed through the ID to obtain the hash value that can be used to generate the ciphertext. The platform server compares the ciphertext CT and returns it to the node on the CB. The specific calculation process is as follows.

$$\begin{aligned} & e(I_2 \cdot \prod_{w \in W} I_w, T_w) \\ =& e(g^{s\tau} \cdot \prod_{w \in W} g^{\tau H(w)}, g^{\frac{1}{\sum_{w' \in W} H(w')+s}}) \\ =& e(g,g)^{\frac{\tau(\sum_{w' \in W} H(w)+s)}{\sum_{w' \in W} H(w)+s}} \end{aligned} \tag{7}$$

If $w = w'$, the above equation is $e(g,g)^{\tau} = I_1$. The ciphertext corresponding to the index where the equation holds is the search result.

G. Dec

When the search is completed, if ID_a sends a temporary key for pre-decryption, the node on the CB pre-decrypts the ciphertext and sends it to ID_a for final decryption; otherwise, the search results are directly sent to ID_a for decryption by herself.

1) Pre-Dec: If the attribute value Val_a of ID_a satisfies the access policy (A, f, Val_p) in CT, then CB uses the temporary key sk'_a to help her pre-decrypt CT into $pre\text{-}CT$. Where $i \in minA$ is the minimum set of attributes that satisfy the access policy, and $x_i \in Z_p$ are some constants that satisfy $\sum_{i \in minA} x_i A_i = (1, 0, \cdots, 0)$.

$$\begin{aligned} & \prod_{i \in minA} (e(g^{\frac{1}{t}}, CT_{1,i}) \cdot e(sk'_1, CT_{2,i}) \cdot e(sk'_2, CT_{3,i}) \cdot \\ & e(sk'_3, CT_{5,i}) \cdot e(ID'_a, C_{4,i}))^{x_i} \\ =& \prod_{i \in minA} e(g^{\frac{1}{t}}, g^{a_i} g^{\alpha_{\eta(i)} r_i}) \cdot e(g^{\frac{\alpha}{t\gamma}} H_1(ID'_a)^{\frac{\beta}{t\gamma}} \\ & H_1(v_\chi)^{\frac{\sigma}{t\gamma}}, (g^{\gamma})^{-r_i}) \cdot e(H_1(v_\chi)^{\frac{\sigma}{t\delta}}, (g^{\delta})^{r_i - r'_i}) \cdot \\ & e(g^{\frac{\sigma}{t}}, H_1(a_{f(i)})^{r'_i}) \cdot e(H_1(ID_a)^{\frac{1}{t}}, g^{\beta_{\eta(i)} r'_i} g^{b_i}))^{x_i} \\ =& \prod_{i \in minA} (e(g,g)^{\frac{a_i}{t}} \cdot e(H_1(ID'_a), g)^{\frac{b_i}{t}})^{x_i} \\ =& e(g,g)^{\sum_{i \in minA} \frac{a_i x_i}{t}} \cdot e(H_1(ID'_a), g)^{\sum_{i \in minA} \frac{b_i x_i}{t}} \\ =& e(g,g)^{\frac{z}{t}} \end{aligned} \tag{8}$$

Then, $pre\text{-}CT = e(g,g)^{\frac{z}{t}}$ is sent to ID_a. Otherwise, the blockchain refuses to pre-decrypt for u. However, if Val_a does not satisfy the access policy contained in CT, CB refuses to pre-decrypt for ID_a.

2) Dec: When ID_a obtains the pre-ciphertext from CB, she just does a simple operation as $\frac{CT_0}{(pre\text{-}CT)^t} = \frac{CT_0}{(e(g,g)^{\frac{z}{t}})^t} = T_a$ to get her transcript.

VI. Security Analysis

Based on the security goals defined in Section III-C, we conduct a comprehensive security analysis of the PPEDS scheme.

A. Data Security and Access Control

The PPEDS scheme ensures data security and fine-grained access control through attribute-based encryption. For example, a teacher defines an access policy such as {*"Major": "computer science" AND "Name": "Alice"*}. With the multi-authority CP-ABE, only users named Alice and majoring in computer science can decrypt the data. The server can provide pre-decryption services for students if the specific access policy is met, ensuring that only authorized users receive the necessary data. Users who do not meet the access policy cannot achieve pre-decryption for them. The specific formal proof can be found in [25].

B. Privacy Protection

Privacy protection is achieved by minimizing the exposure of user attributes and ensuring anonymity. The system ensures that the attributes in the access policies do not leak sensitive user information. For example, the access policy is {*"Major": "computer science" AND "Name": "Alice"*}, where {*"Major": "computer science" AND "Name": "Alice"*} is uploaded to the education platform server along with the ciphertext, and the specific major information and name are not displayed in the ciphertext. Therefore, obtaining the ciphertext will not obtain the specific attribute value, thereby protecting user privacy. In addition, the ciphertext can also resist the Decisional Diffie-Hellman test (DDH-test) [26], making it infeasible for attackers to infer information about the recipient through brute-force methods. This ensures that users maintain anonymity during data access, preserving privacy even in a semi-trusted environment.

C. Secure Search

To maintain data confidentiality during searches, the PPEDS scheme uses encrypted indexes and query trapdoors, which are processed by trusted nodes. The secure search mechanism is built on the Identity-Based Encryption (IBE) scheme from [27]. According to the security analysis in [27], the PPEDS scheme also resist Chosen-Keyword Attacks (CKA). By encrypting both indexes and queries, the system prevents attackers from learning sensitive information during the search process. Only trusted nodes execute the search, ensuring that authorized parties can retrieve results while unauthorized entities remain unable to determine the actual keywords involved, thereby maintaining confidentiality.

D. System Availability

System availability is ensured through the use of blockchain-based consistency mechanisms. In the private blockchain, consistency is maintained by verifying the signatures of internal administrators, while in the consortium blockchain, consistency is ensured through platform-level signature verification. These verification mechanisms prevent unauthorized modifications while allowing the system to continue functioning even if some nodes are compromised. The use of consensus protocols helps guarantee that data and system integrity are preserved without leaking sensitive information, ensuring that the system remains available and reliable.

The PPEDS scheme effectively mitigates the risks brought by the semi-trusted environment and ensures data security, privacy protection, secure search, and high system availability. By adopting attribute encryption-based access control, secure search mechanism, and blockchain-based consistency check, the system provides strong protection against unauthorized data access and ensures user privacy even in the face of potential attacks.

VII. Performance Analysis

First, we compare and analyze the characteristic performance of several data sharing schemes in Table I. The comparison shows that only the PPEDS scheme fully supports multi-keyword search, pre-decryption, hidden attributes and decentralized authorities. Its features enhance its applicability in complex scenarios such as lifelong education platforms, which require flexible and secure data sharing.

Then, we analyze the theoretical time complexity of the major algorithms, and use several time-consuming operations to represent them. E represents the exponential operation in the group G, E_T represents the exponential operation in the group G_T, and P represents the bilinear pairing operation. As shown in Table II, n_a represents the number of attributes, and n_k represents the number of keywords. The time complexity of TrapGen, Search, and Dec is nearly fixed. The time complexity of EileEnc and pre-Dec is proportional to the number of attributes in the access policy, while the time complexity in IndexEnc is linearly related to the number of keywords to be encrypted.

TABLE I. Feature Comparison

schemes	F1	F2	F3	F4	F5	F6
[10]	personal data	×	×	×	×	×
[11]	IoT	×	×	×	×	×
[12]	Medical	×	×	×	×	×
[13]	Civil aviation	×	×	×	×	×
[14]	personal data	×	×	×	×	×
[15]	transportation	×	✓	×	×	×
[16]	personal data	×	✓	×	×	×
[17]	Medical	✓	✓	×	✓	×
[18]	Medical	✓	✓	✓	✓	×
[8]	Education	✓	×	×	×	×
PPEDS	Education	✓	✓	✓	✓	✓

Note: F1: Application Background; F2: Fine-grained access control; F3: Keyword search; F4: Pre-decryption; F5: Hidden attributes; F6: Decentralized authorities.

Compared with two typical scheme [17], [18], which are two schemes proposed in the medical field that supports keyword search and attribute hiding, the PPEDS scheme takes longer than the [18] scheme in the FileEnc stage. This is because the PPEDS scheme not only supports hiding attributes, but also supports distributed attribute authorities to manage attributes, and can realize user anonymity. Complex operations bring a safer and more private user experience. The two schemes embed access control into the search process, that is, only users whose attributes meet the policy can search. This results in the search part being secure but too inefficient in actual use, making it unsuitable for sharing educational

TABLE II. TIME COMPLEXITY COMPARISON

Algorithms	FileEnc	IndexEnc	TrapGen	Search	Dec
[17]	-	$(2n_a+6)E+P$	$(2n_a+n_k+2)E$	$(2n_a+n_k+2)P+E$	-
[18]	$E+E_T$	$(2n_a+n_k+1)E+E_T+n_kP$	$(n_a+4)E$	$(n_a+3)P+(n_a+n_k+2)E_T+E$	P
PPEDS	$7n_aE$+E_T	$(n_k+1)E+E_T$	E	P	E_T

- Indicates that this algorithm is not involved.

data. The high efficiency demonstrated by the PPEDS scheme highlights its practicality and suitability in the educational data sharing model.

We test the actual operational efficiency of various algorithms in the PPEDS scheme using Python on a real dataset. The data set is the encrypted transcripts of students majoring in computer science and technology, civil engineering, finance, law, and pharmacy at Hebei Open University in 2020. 100 students are randomly selected from each major, and 10 courses are selected for each major, totaling 5,000 transcripts. All experimental results were obtained using a personal computer with an Intel Core i5-1340P CPU (4.60GHz) and 16GB of memory.

A. FileEnc

Fig. 4 shows the time cost of FileEnc algorithm that the teacher executed using a multi-authority CP-ABE-based algorithm. The horizontal axis represents the number of attributes used in the access policy, ranging from 2 to 10, while the vertical axis represents the time required for encryption. As can be seen from the figure, the time required for file encryption increases linearly with the number of attributes. Although using more attributes in the policy can enhance the granularity and security of access control, it also leads to increased computational overhead. Therefore, a balance must be struck between security requirements and encryption efficiency to avoid performance bottlenecks, especially in scenarios involving a large number of attributes.

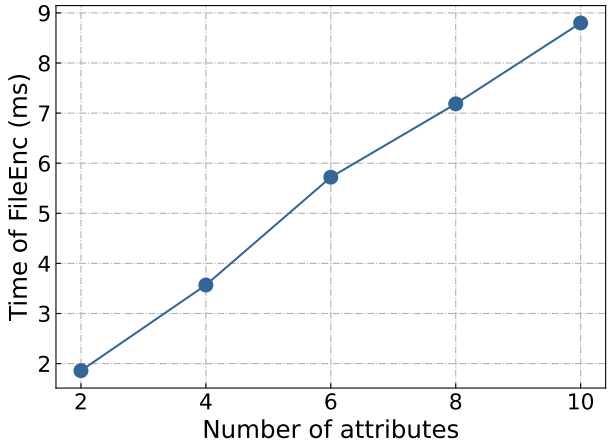

Fig. 4. Time cost for FileEnc.

B. IndexEnc

Fig. 5 presents the time cost of the IndexEnc algorithm. The horizontal axis represents the number of keywords involved in the index, ranging from 2 to 10, while the vertical axis represents the time required for encryption. The figure shows that the encryption time increases linearly as the number of keywords increases. That is, the computational complexity of index encryption is proportional to the number of keywords included in the index. Although adding more keywords can improve data retrieval efficiency and search accuracy, it also leads to increased computational overhead. Therefore, it is crucial to optimize the number of keywords used during indexing to balance search performance and encryption efficiency, especially when dealing with large amounts of encrypted data.

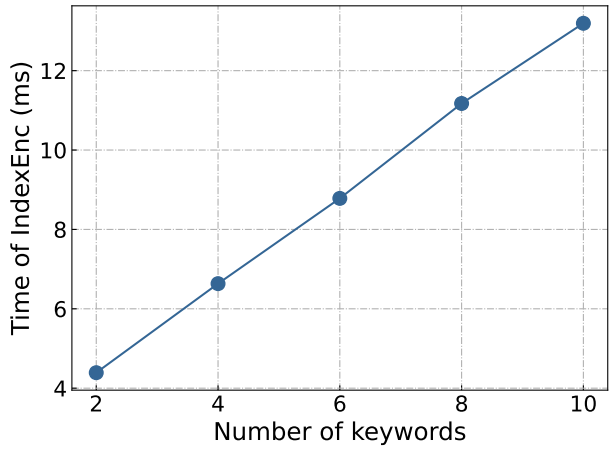

Fig. 5. Time cost for IndexEnc.

C. TrapGen

Fig. 6 illustrates the time cost of the TrapGen algorithm, which represents the time taken to generate a trapdoor for an encrypted search by the searcher. The horizontal axis shows the number of keywords, ranging from 2 to 10, while the vertical axis represents the time required to generate the trapdoor. As can be seen from the figure, the time cost of generating the trapdoor remains relatively constant regardless of the number of keywords involved. This shows that the trapdoor generation process in the PPEDS scheme is efficient and scalable even as the number of keywords increases. The stable performance of the TrapGen operation means that users can search with multiple keywords without significantly increasing the computational overhead. This makes the system well-suited for scenarios where users need to search for multiple attributes or terms simultaneously while maintaining efficient response times. However, too many keywords may result in fewer search results, so the users can adjust them according to their needs.

D. Search

Fig. 7 shows the time cost of executing the search algorithm in the PPEDS scheme. The horizontal axis represents the number of keywords used in the search, ranging from 2 to 10, while

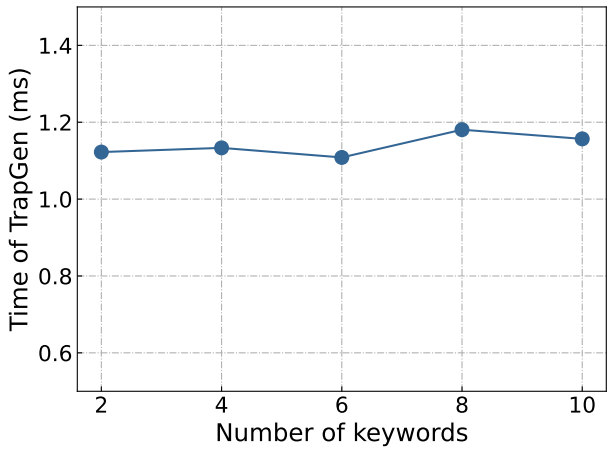

Fig. 6. Time cost for TrapGen.

the vertical axis represents the search time. The figure shows that the time cost of performing the search operation remains relatively constant as the number of keywords increases, and the search process is not significantly affected by the number of search keywords. This consistent performance shows that the proposed search algorithm is both efficient and scalable, even when processing multiple keywords simultaneously. Therefore, the scheme is very suitable for practical applications, and users searching encrypted data using multiple terms will not encounter obvious delays in search efficiency.

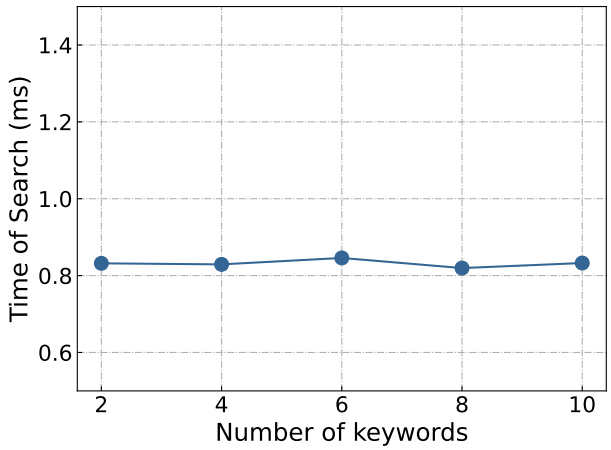

Fig. 7. Time cost for search.

E. Dec

Fig. 8 shows the time cost of the Dec algorithm, comparing the time required for the platform server to perform pre-decryption and the time required for the user to perform the final decryption. The figure includes two graphs: a zoomed-in view to highlight the comparison of the decryption time for the two stages, and a full-scale view that illustrates the overall decryption cost as the number of attributes increases.

The horizontal axis represents the number of attributes involved in decryption, while the vertical axis represents the decryption time. The comparison shows that the pre-Dec performed by the platform server significantly reduces the computational load for the user. In the zoomed-in view, it can be seen that the time cost of the final decryption performed by the user is small and remains almost constant regardless of the number of attributes. In the full-scale graph, the pre-Dec time grows as the number of attributes increases. However, the platform server has powerful computing power and performs the most computationally intensive part of the decryption process, leaving the user with only very small and fixed-time decryption steps. This ensures that the user experiences minimal computational overhead during the decryption process, thereby improving efficiency and maintaining security. By delegating the resource-intensive decryption part to the platform server, the system can strike a balance between user convenience and cryptographic security.

F. Verify by Blockchain

In the PPEDS scheme, the verification process of both private and consortium blockchains relies on the verification of signatures from teachers and educational platforms. We compare three commonly used signature schemes: ECDSA [28], Schnorr [29], and BLS [30] to determine the signature algorithm that best suits different scenarios.

Table III compares the properties of the three schemes, including signature size, speed, and support for aggregation. The ECDSA algorithm has a medium signature size and speed, without support for aggregation, making it suitable for environments where aggregation is not required and moderate signing and verification times are acceptable. Schnorr scheme have a small signature size and support aggregation (e.g., MuSig [31]). BLS scheme has a fixed-size signature and supports native aggregation, which is advantageous in distributed networks like Ethereum 2.0. However, due to high computational cost, BLS is more suitable for scenarios that prioritize aggregation and verification across multiple nodes rather than speed.

Table IV provides the time cost for signing and verification at a security parameter of 256 bits. Overall, for verifying teacher signatures on the private blockchain, Schnorr is preferred due to its small signature size, fast signing and verification times, and support for aggregation. This ensures that the verification process is both efficient and scalable. For the consortium blockchain, where educational platform signatures need to be verified across multiple institutions, BLS signature scheme is recommended. Although BLS has slower signing and verification times, its native support for aggregation makes it suitable for reaching consensus among distributed nodes. This trade-off is acceptable in scenarios where aggregation significantly reduces the number of signatures that need to be processed, thereby improving overall system efficiency.

VIII. Conclusions

This paper proposes the PPEDS scheme, a privacy-preserving data sharing scheme for large-scale, interconnected lifelong education platforms. By integrating private and consortium blockchains, the scheme enables secure, verifiable, and efficient cross-institutional data sharing while preserving user privacy.

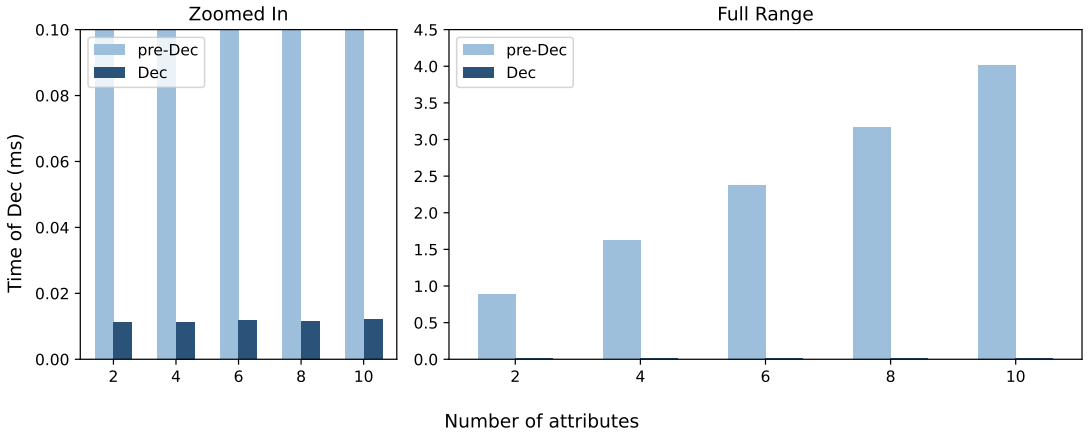

Fig. 8. Time cost for Dec.

TABLE III. FEATURE COMPARISON OF THREE SIGNATURE ALGORITHMS

Schemes	Signature Size	Speed	Aggregation Support	Application Scenarios
ECDSA	Medium	Medium	×	Bitcoin, SSL, TLS
Schnorr	Small	Fast	✓(supports MuSig)	Bitcoin (Taproot), Blockchain
BLS	Fixed-size	Slow	✓(native support)	Ethereum 2.0, Distributed networks

TABLE IV. PERFORMANCE ANALYSIS OF THREE SIGNATURE SCHEMES

Schemes	Sign (ms)	Verify (ms)
ECDSA	0.0903	0.3214
Schnorr	0.0051	0.0078
BLS	4.0647	4.5202

The PPEDS scheme introduces a multi-authority CP-ABE mechanism with hidden attributes to achieve fine-grained access control and user anonymity, allowing each authority to independently manage attribute policies. Additionally, the scheme incorporates public key-based multi-keyword searchable encryption with pre-decryption support via the consortium blockchain, significantly reducing computational overhead for end-users. These innovations collectively enhance the scalability, security, and privacy of educational data sharing.

From a practical perspective, PPEDS addresses real-world challenges in lifelong education systems, such as cross-platform credit recognition, transcript verification, and elective course access. The performance evaluations and comparisons results demonstrate that our pre-decryption mechanism reduces user-side decryption time to only 0.01ms compared to traditional CP-ABE-based systems. The trapdoor generation and search operations maintain sublinear growth with respect to the number of keywords, supporting real-time query capabilities even on resource-constrained devices. In addition, the encryption and index generation overheads remain stable as the number of access attributes increases, showcasing the scalability of the scheme. These metrics highlight the advantage of PPEDS in enabling efficient, scalable, and privacy-preserving data sharing for large-scale educational systems.

However, the current design assumes a semi-honest authority and does not account for collusion among multiple entities. Furthermore, while pre-decryption significantly improves search efficiency, the overhead of storing indexes on the blockchain remains a challenge, potentially impacting scalability over time.

Future work could consider integrating dynamic policy updates and revocation mechanisms to improve flexibility, and optimizing blockchain storage by compressing or pruning index structures to enhance scalability. Furthermore, exploring practical deployments and integration with decentralized identity frameworks will validate the practicality of this solution in real-world educational settings.

ACKNOWLEDGMENT

This work was supported by the Hebei Province Education Science "14th Five Year Plan Project" of 2023 (2303010) and University-level Research Project of Hebei Open University (ZD202203). The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

REFERENCES

[1] J. Field, "Lifelong education," *International Journal of lifelong education*, vol. 20, no. 1-2, pp. 3–15, 2001.

[2] A. Kuznetsov, E. Pyanykh, and M. Rodaikina, "Digital transformation in the context of improving the quality of lifelong education," 2021.

[3] S. F. M. Y. S. Salim, M. F. Mahmood, and A. B. Ahmad, "The importance of information literacy to support lifelong learning in convergence era," *International Journal of Academic Research in Progressive Education and Development*, vol. 7, no. 3, 2018.

[4] S. JANTHAPASS, N. CHANTHAPASSA, and S. KENAPHOOM, "The evolution of lifelong learning: From traditional classrooms to anywhere, anytime education," *Asian Education and Learning Review*, vol. 2, no. 1, pp. 42–54, 2024.

[5] J. Leng, M. Zhou, J. L. Zhao, Y. Huang, and Y. Bian, "Blockchain security: A survey of techniques and research directions," *IEEE Transactions on Services Computing*, vol. 15, no. 4, pp. 2490–2510, 2020.

[6] M. Krichen, M. Ammi, A. Mihoub, and M. Almutiq, "Blockchain for modern applications: A survey," *Sensors*, vol. 22, no. 14, p. 5274, 2022.

[7] N. Deepa, Q.-V. Pham, D. C. Nguyen, S. Bhattacharya, B. Prabadevi, T. R. Gadekallu, P. K. R. Maddikunta, F. Fang, and P. N. Pathirana, "A survey on blockchain for big data: Approaches, opportunities, and future directions," *Future Generation Computer Systems*, vol. 131, pp. 209–226, 2022.

[8] X. Huang, Y. Wang, H. Liang, Y. Ding, Q. Wu, Z. Zhang, and Q. Qu, "Educhain: A blockchain-based privacy-preserving lifelong education platform," in *International Conference on Database Systems for Advanced Applications*. Springer, 2023, pp. 701–706.

[9] A. Marouan, M. Badrani, N. Kannouf, and A. Chetouani, "Empowering education: leveraging blockchain for secure credentials and lifelong learning," in *Blockchain Transformations: Navigating the Decentralized Protocols Era*. Springer, 2024, pp. 1–14.

[10] B.-K. Zheng, L.-H. Zhu, M. Shen, F. Gao, C. Zhang, Y.-D. Li, and J. Yang, "Scalable and privacy-preserving data sharing based on blockchain," *Journal of Computer Science and Technology*, vol. 33, pp. 557–567, 2018.

[11] T. Li, H. Wang, D. He, and J. Yu, "Blockchain-based privacy-preserving and rewarding private data sharing for iot," *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 15 138–15 149, 2022.

[12] C. Li, M. Dong, X. Xin, J. Li, X.-B. Chen, and K. Ota, "Efficient privacy-preserving in iomt with blockchain and lightweight secret sharing," *IEEE Internet of Things Journal*, 2023.

[13] X. Li, H. Zhao, and W. Deng, "Bfod: Blockchain-based privacy protection and security sharing scheme of flight operation data," *IEEE Internet of Things Journal*, 2023.

[14] W. Liang, Y. Yang, C. Yang, Y. Hu, S. Xie, K.-C. Li, and J. Cao, "Pdpchain: A consortium blockchain-based privacy protection scheme for personal data," *IEEE Transactions on Reliability*, vol. 72, no. 2, pp. 586–598, 2022.

[15] S. Jiang, J. Cao, H. Wu, K. Chen, and X. Liu, "Privacy-preserving and efficient data sharing for blockchain-based intelligent transportation systems," *Information Sciences*, vol. 635, pp. 72–85, 2023.

[16] R. Du, C. Ma, and M. Li, "Privacy-preserving searchable encryption scheme based on public and private blockchains," *Tsinghua Science and Technology*, vol. 28, no. 1, pp. 13–26, 2022.

[17] S. Niu, M. Song, L. Fang, F. Yu, S. Han, and C. Wang, "Keyword search over encrypted cloud data based on blockchain in smart medical applications," *Computer Communications*, vol. 192, pp. 33–47, 2022.

[18] J. Liu, Y. Fan, R. Sun, L. Liu, C. Wu, and S. Mumtaz, "Blockchain-aided privacy-preserving medical data sharing scheme for e-healthcare system," *IEEE Internet of Things Journal*, vol. 10, no. 24, pp. 21 377–21 388, 2023.

[19] L.-j. Luo, H.-b. Lei[1], and Z. Wang, "Educhain: Blockchain-based informative platform for vocational education and training," in *Proceedings of the 2023 4th International Conference on Big Data and Informatization Education (ICBDIE 2023)*, vol. 178. Springer Nature, 2023, p. 454.

[20] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in *Advances in Cryptology–EUROCRYPT 2005: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005. Proceedings 24*. Springer, 2005, pp. 457–473.

[21] M. Chase, "Multi-authority attribute based encryption," in *Theory of Cryptography: 4th Theory of Cryptography Conference, TCC 2007, Amsterdam, the Netherlands, February 21-24, 2007. Proceedings 4*. Springer, 2007, pp. 515–534.

[22] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in *2007 IEEE symposium on security and privacy (SP'07)*. IEEE, 2007, pp. 321–334.

[23] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proceeding 2000 IEEE symposium on security and privacy. S&P 2000*. IEEE, 2000, pp. 44–55.

[24] H. Guo and X. Yu, "A survey on blockchain technology and its security," *Blockchain: research and applications*, vol. 3, no. 2, p. 100067, 2022.

[25] J. Guo, C. Tian, X. Lu, L. Zhao, and Z. Duan, "Multi-keyword ranked search with access control for multiple data owners in the cloud," *Journal of Information Security and Applications*, vol. 82, p. 103742, 2024.

[26] X. Li, D. Gu, Y. Ren, N. Ding, and K. Yuan, "Efficient ciphertext-policy attribute based encryption with hidden policy," in *Internet and Distributed Computing Systems: 5th International Conference, IDCS 2012, Wuyishan, Fujian, China, November 21-23, 2012. Proceedings 5*. Springer, 2012, pp. 146–159.

[27] C. Gentry, "Practical identity-based encryption without random oracles," in *Advances in Cryptology-EUROCRYPT 2006: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28-June 1, 2006. Proceedings 25*. Springer, 2006, pp. 445–464.

[28] D. Johnson, A. Menezes, and S. Vanstone, "The elliptic curve digital signature algorithm (ecdsa)," *International journal of information security*, vol. 1, pp. 36–63, 2001.

[29] C.-P. Schnorr, "Efficient identification and signatures for smart cards," in *Advances in Cryptology—CRYPTO'89 Proceedings 9*. Springer, 1990, pp. 239–252.

[30] D. Boneh, B. Lynn, and H. Shacham, "Short signatures from the weil pairing," in *International conference on the theory and application of cryptology and information security*. Springer, 2001, pp. 514–532.

[31] G. Maxwell, A. Poelstra, Y. Seurin, and P. Wuille, "Simple schnorr multi-signatures with applications to bitcoin," *Designs, Codes and Cryptography*, vol. 87, no. 9, pp. 2139–2164, 2019.