

Towards Personal Data Sharing Autonomy: A Task-driven Data Capsule Sharing System

Qiuyun Lyu, Yilong Zhou, Yizhi Ren, Zheng Wang, and Yunchuan Guo

Abstract—Personal data custodian services enable data owners to share their data with data consumers in a convenient manner, anytime and anywhere. However, with data hosted in these services being beyond the control of the data owners, it raises significant concerns about privacy in personal data sharing. Many schemes have been proposed to realize fine-grained access control and privacy protection in data sharing. However, they fail to protect the rights of data owners to their data under the law, since their designs focus on the management of system administrators rather than enhancing the data owners’ privacy. In this paper, we introduce a novel task-driven personal data sharing system based on the data capsule paradigm realizing personal data sharing autonomy. It enables data owners in our system to fully control their data, and share it autonomously. Specifically, we present a tamper-resistant data capsule encapsulation method, where the data capsule is the minimal unit for independent and secure personal data storage and sharing. Additionally, to realize selective sharing and informed-consent based authorization, we propose a task-driven data sharing mechanism that is resistant to collusion and EDoS attacks. Furthermore, by updating parts of the data capsules, the permissions granted to data consumers can be immediately revoked. Finally, we conduct a security and performance analysis, proving that our scheme is correct, sound, and secure, as well as revealing more advantageous features in practicality, compared with the state-of-the-art schemes.

Index Terms—personal data sharing autonomy, data capsule, task-driven.

I. INTRODUCTION

PERSONAL data, defined as “any information relating to an identified or identifiable natural person” [1], is gathered and maintained by numerous services, including online social networks, shopping platforms, consumer electronics [2], and VANETs [3], [4]. To efficiently manage and store this personal data, cloud-based Personal Data Storage (PDS [5]) systems (e.g., Mydex [6], WebBox [7], and HAT [8]) have been developed. These systems enable personal data owners to store their data and share it with service providers in exchange for various services [9].

However, according to the “cloud attacks rise but most sensitive data remains unencrypted” by Techeircle [10], only about 45% of sensitive data hosted in the cloud are encrypted. And the prevalence of data breaches, insider attacks, and the misuse of collected data by organizations also show that data privacy has emerged as a fundamental challenge in today’s digital landscape [11]. As a result, personal data owners are increasingly concerned about losing control of their data. This

concern is rooted in the fact that data hosted in the cloud is physically uncontrollable by owners [12].

To address the above challenges and ease the concern of data owners, governments and organizations have formulated many data privacy regulations, such as the General Data Protection Regulation (GDPR [13]), the Health Insurance Portability and Accountability Act (HIPAA [14]), and the California Consumer Privacy Act (CCPA [15]), to protect personal data owners’ right to their data. From the regulations, we can outline three requirements for data privacy protection in the process of personal data sharing, namely 1) data collecting and processing minimization, which requires data consumers only collect and process the minimum data strictly needed for the specified purposes; 2) explicit consent, requiring data consumers to obtain unambiguous consent from data owners before accessing data; and 3) the right to be forgotten, which necessitates data consumers delete the personal data after services are finished. These requirements can be concluded into personal data sharing autonomy which balances the fine-grained sharing of personal data with privacy protection.

In order to realize personal data sharing autonomy, technically, a privacy-enhancing data-sharing system needs to, at least, achieve the abilities of *selective sharing*, *informed consent-based authorization*, and *permission revocation*. *Selective sharing* is an approach to realize the “minimization of data collection and processing” and enables data owners to share their data (e.g., login credentials, transaction details, verification information) with selected service providers, providing only the information that is strictly needed by them, while specifying access times for the data. For example, in online forum applications, users can select any nickname or account to post, and in instant messaging software, users can share data with different friends and set the last accessible time of the data. It prevents the unwarranted disclosure of additional personal data, thereby protecting the privacy and control of the personal information. *Informed consent-based authorization* requires that data consumers obtain explicit, informed consent from data owners prior to accessing, modifying, deleting, and forwarding their personal data. For example, when users first use the software, they must read the software’s privacy policy in its entirety and click the “agree” button, and in online video conferences, participants must obtain consent from other attendees to record the meeting. This process ensures that data owners are fully aware of how their information will be processed. Since personal data is increasingly regarded as an individual’s private property, explicit consent should be obtained before processing it. This paper focuses on the access aspect during personal data sharing. And *permission*

This work was supported by the Natural Science Foundation of Zhejiang Province (No. LY23F020017) and the “Pioneer” and “Leading Goose” R&D Program of Zhejiang (Grant No. 2022C03174).

revocation is a compromised implementation of “the right to be forgotten” since the trusty erase of data is hard to realize [16]. Permission revocation usually makes the data’s decryption keys invalid in which data consumers cannot access the data anymore.

In fact, in order to protect personal data, numerous privacy-enhancing data sharing schemes have been proposed, which can be primarily categorized into two types: *traditional encryption data sharing schemes* and *attribute-based encryption (ABE) data sharing schemes*.

Traditional encryption data sharing schemes use traditional cryptographic tools, e.g., symmetric encryption [17] or public-key encryption [18], to encrypt data before storing it in cloud storage, thus ensuring confidentiality. However, such schemes fail to enable users to authorize multiple data consumers in a “one-to-many” access control manner. To address the above-mentioned issue, many traditional encryption data sharing schemes employ access control mechanisms such as role-based [19] or identity-based [20] access control to improve efficiency. Although these traditional data sharing systems achieve data confidentiality and “one-to-many” data sharing, they fall short of the above abilities: selective sharing, informed-consent based authorization, and permission revocation.

Attribute-based encryption (ABE) data sharing schemes have been proposed, firstly, to provide privacy-preserving, efficient, and fine-grained data sharing for data owners [21], [22], [23], [24], [25], [26], [27]. The ABE [28], [29] enables personal data owners to provide their information to various data consumers whose attributes satisfy the access policy [30]. The scheme [22] integrates ABE and identity-based encryption (IBE [31]) to build a privacy-preserving data sharing system that supports the revocation of data consumers by updating part of the ciphertext. However, its revocation process lacks efficiency and granularity. And the scheme [23] combines ABE and proxy re-encryption (PRE [32], [33], [34]) to design a two-factor protection mechanism to improve efficiency for decryption keys revocation of data consumers, but the scheme fails to allow the data owner to initiate the revocation process actively. Furthermore, the scheme [26] proposes an electronic medical record sharing system integrating tamper resistance and supporting malicious user revocation without affecting honest users. Though the above-mentioned schemes realize the permission revocation in many ways, they fail to support selective sharing and informed-consent based authorization.

To realize selective sharing and informed-consent based authorization, the scheme [27] combines searchable encryption (SE [35]) and ABE to design a selective data sharing system by selecting the keywords of data and sending them to data consumers. The keywords in the scheme [27] represent the data owner’s informed consent for data consumers to access their data. However, because of the separation between keywords and data, keywords can be forwarded by authorized consumers to unauthorized consumers maliciously without the data owners’ authorization. For instance, in the scheme [27], attackers holding a keyword can collude with another attacker holding a set of attributes that satisfy the access policy to decrypt the ciphertext which is called collusion attacks [30].

In addition, for efficient and convenient management of

data, the above-mentioned schemes employ the cloud storage as the data keeper. However, they are unable to resist Economic Denial of Sustainability (EDoS [36], [37]) attacks, where attackers excessively request the data of a data owner, leading to significant financial losses due to the overuse of cloud resources. To deal with this attack, the scheme [25] designs a dual access control for data sharing that enables sharing data securely and controls the download request of cloud resources, thus realizing EDoS resistance. However, it fails to realize personal data sharing autonomy.

In the above schemes, a single piece of data is not treated as the independent minimum unit for authorization and sharing, and the authorization to data consumers is not strictly and explicitly associated with the consent of data owners, mainly due to their design focus on the management of system administrators rather than enhancing the data owners’ privacy. This leads to misuse of data by unauthorized data consumers, tampering with data by data keepers, and resulting in data owners losing control over their own data. To solve the above issues, we introduce a novel task-driven data capsule sharing system (TD-DCSS) realizing personal data sharing autonomy for flexible, secure, and privacy-enhancing data sharing. Data capsules [11], [38], which store both the data and the access policy dictating their usage, offer the potential to act as the minimal units for independent and secure personal data storage and sharing. And the “task-driven” data sharing mechanism links a task to a consumer, with the task representing the owners’ consent and the permission granted by the owner to the consumer. The main contributions of the paper are stated as follows.

- *Data capsule encapsulation method with tamper resistance.* A method enabling personal data owners to encapsulate their data into data capsules is devised. Data capsules encapsulate the personal data, an access policy, and a bilinear pairing verification element. The proposed encapsulation method, building on the cumulative XOR random mask, bilinear pairing hash generation and ABE, generates a data capsule under an access policy. To protect the data capsules from tampering by semi-trusted data keepers, a bilinear pairing verification element, derived from the data capsule itself, is embedded within the capsule to ensure its integrity.
- *Task-driven data sharing mechanism with collusion and EDoS resistance.* We propose a novel task-driven data sharing mechanism for selective data sharing, and formulate an efficient task construction methodology. The “task-driven” mechanism comprises two components: a task sent to data consumers and two tokens sent to the data keeper. The task, signifying the data owner’s consent, includes the decryption keys for specific data granules. Since the task is associated with the data consumer’s ID, the data consumer cannot collude with other data consumers to decrypt the data capsules. The two tokens, the download token and the revocation token, serve distinct purposes in our system. Specifically, preventing EDoS attacks is achieved by the download token, and revoking the data consumers’ permissions is achieved by

the revocation token.

- *Personal data sharing autonomy.* The concept of personal data sharing autonomy is introduced, focusing on data owners' privacy. Personal data sharing autonomy, including selective sharing, informed-consent based authorization, and permission revocation, not only realizes the data privacy regulations: data collecting and processing minimization, explicit consent, and the right to be forgotten, but also enables data owners to have full control over their data.
- *Security and efficiency analysis.* We conduct a comprehensive security analysis, a theoretical analysis, and an experimental simulation. The security analysis shows that our scheme is correct, sound, and secure under the security model. The results of the theoretical analysis and the experimental simulation demonstrate that, compared with the state-of-the-art schemes, our scheme offers more advantageous features and stands as a highly efficient solution.

II. OVERVIEW AND DESIGN GOALS

In this section, we sketch the proposed data capsule encapsulation method and task-driven data sharing mechanism, providing intuition behind them. Then, we introduce the design goals that guide the development of the proposed scheme. In order to combine with the real scene, we model the role of service provider (SP) and cloud storage (CS) as data consumer and data keeper, respectively.

A. Overview of the Proposed Scheme

Our proposed task-driven data capsule sharing system (TD-DCSS) scheme is based on a fast attribute-based mechanism (FABEO [29]), and it can easily integrate with other CP-ABE schemes to support additional features (e.g., attribute revocation [39], accountability [40], and policy hiding [41]). We introduce a data capsule encapsulation method and task-driven data sharing mechanism in TD-DCSS for flexible, efficient, and secure data sharing. The details are described as follows.

- *Data capsule encapsulation method.* In real life, service providers may only need partial personal data from users to successfully offer services. For instance, when a hospital verifies a user's age, it is sufficient to provide the information from the ID related to the birth date. Providing the entire ID in this situation may pose a privacy risk. To ease this risk, we propose a data capsule encapsulation method as shown in Fig. 1. The main idea is that data owners can initially break down their data into granules, which can be shared in the future upon demand. Subsequently, owners can XOR all the data granules, including a secret granule generated under an access policy through encryption operations in ABE. Finally, the data capsule (DC) comprises a set of data granules dg_1, dg_2, dg_3, dg_4 and a secret granule sg , which is generated under the specified access policy. Specifically, the data capsule is represented as $dg_1 \oplus dg_2 \oplus dg_3 \oplus dg_4 \oplus sg$.

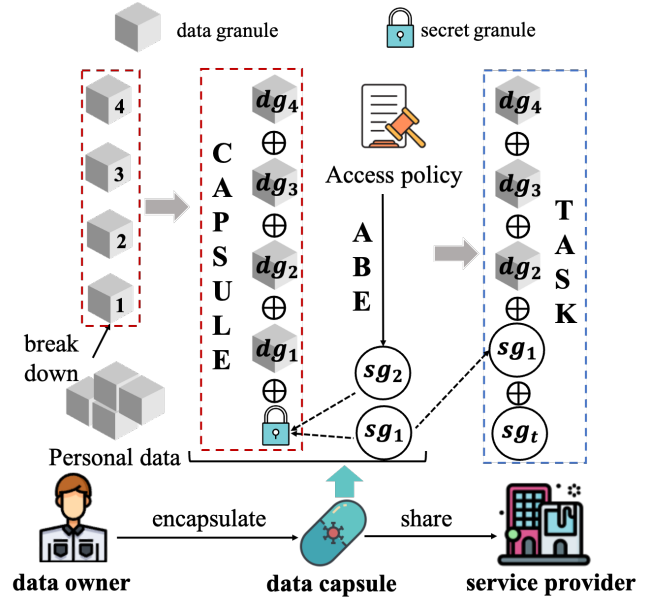


Fig. 1: Data capsule encapsulation and sharing.

- *Task-driven data sharing mechanism.* To realize selective sharing and informed-consent based authorization, we propose a novel task-driven data sharing mechanism, and the main idea is shown in Fig. 1. When the data owner wishes to share dg_1 in the data capsule selectively with the service provider, it issues a task and sends the task to the SP as the consent for accessing its data. Specifically, the sg consists of two distinct components, sg_1 and sg_2 , where sg_1 is randomly generated, while sg_2 is generated under the specified access policy. In this scenario, the task takes the form of $dg_2 \oplus dg_3 \oplus dg_4 \oplus sg_1 \oplus sg_t$, where sg_t is associated with the target ID of the SP. After retrieving the data capsule from the CS and receiving the task from the owner, the SP can recover dg_1 by reconstructing the sg_2 and sg_t , and then computing $dg_1 = DC \oplus \text{task} \oplus sg_2 \oplus sg_t$.

B. Design Goals

Our goal is to build a task-driven data capsule sharing system, realizing personal data sharing autonomy, for flexible, secure, and privacy-enhancing data sharing. There are four key objectives in our scheme: correctness, soundness, security, and efficiency.

- *Correctness:* A service provider with a data capsule encrypted under an access policy, a task bound to its ID, and its attributes that satisfy the access policy, can correctly recover the data encrypted in the data capsule and task.
- *Soundness:* Service providers should refrain from downloading data capsules from cloud storage when the task issuer is wrong, and should not receive tasks or data capsules that cannot be decrypted. The cloud storage should not learn any sensitive information about the data capsule when updating it.

- *Security*: Unauthorized service providers and cloud storage should have no knowledge of personal data. The task must not disclose any information beyond what is contained in both the data capsule and the task itself.
- *efficiency*: The computation and storage consumption of the scheme should not be too much.

III. PRELIMINARIES

A. Access Structure

Our access structure is denoted by $\mathbb{A} = (\mathbb{M}, \pi)$, where policy matrix $\mathbb{M} \subseteq \mathbb{Z}_p^{n_1 \times n_2}$ and mapping $\pi : [n_1] \rightarrow \mathcal{U}$, where $[n]$ denotes the set $\{1, 2, \dots, n\}$, and \mathcal{U} denotes the universe of attributes. Let $\mathcal{S} \models \mathbb{A}$ indicate that \mathcal{S} satisfies the access policy $(\mathbb{A} = \mathbb{M}, \pi)$, where $\mathcal{S} \in \mathcal{U}$ denotes a set of attributes. We use the same definitions $\rho(i) = |\{z | \pi(z) = \pi(i), z \leq i\}|$ and $\tau = \max_{i \in [n_1]} \rho(i)$ corresponding to the maximum number of times an attribute is used in \mathbb{M} as [29]. Note that any boolean formula can be transferred to (\mathbb{M}, π) in polynomial time [42].

B. Bilinear Pairing

Let λ be a security parameter and GroupGen be a probabilistic polynomial time algorithm that takes the 1^λ and outputs group description $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e : \mathbb{G}_1 \times \mathbb{G}_2 \rightarrow \mathbb{G}_T, g_1, g_2)$, where p is the group order of $\mathbb{G}_1, \mathbb{G}_2$ and \mathbb{G}_T , and g_1, g_2 is the generator of $\mathbb{G}_1, \mathbb{G}_2$ respectively. The mapping e has the following properties:

- **Bilinearity**: For all $u \in \mathbb{G}_1, v \in \mathbb{G}_2$ and $x, y \in \mathbb{Z}_p$, $e(u^x, v^y) = e(u, v)^{xy}$
- **Non-degeneracy**: $\exists u \in \mathbb{G}_1, v \in \mathbb{G}_2$, such that $e(u^x, v^y) = I_{\mathbb{G}_T}$, where $I_{\mathbb{G}_T}$ denotes the identity element of \mathbb{G}_T .
- **Computability**: e can be easily and efficiently computed.

C. Security Assumption

Definition 1 (DBDH): Let \mathcal{BG} be a type-III pair $(\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, p, e)$ with generators $g \in \mathbb{G}_1, h \in \mathbb{G}_2$. Given a DBDH tuple $(g, g^a, g^b, g^c, h, h^a, h^b, h^c, \mathcal{Z})$, where $a, b, c \in \mathbb{Z}_p$, the goal of the decisional bilinear Diffie-Hellman (DBDH) problem is to distinguish whether $\mathcal{Z} = e(g, h)^{abc}$ or $\mathcal{Z} = \mathcal{R}$, where \mathcal{R} is a random element of \mathbb{G}_T .

D. Notation

For the convenience, the frequently used symbols are presented in TABLE I.

IV. DEFINITION

A. Task-Driven Data Capsule Sharing System

Definition 2 (TD-DCSS): The task-driven data capsule sharing system consists of the following algorithms:

- $\text{Setup}(1^\lambda) \rightarrow (mpk, msk)$: The Setup algorithm takes a system security parameter $\lambda \in \mathbb{N}$ as input, and outputs a master public key mpk and a master secret key msk .
- $\text{KeyGenSP}(mpk, msk, ID_{SP}, \mathcal{S} \subseteq \mathcal{U}) \rightarrow sk_{SP}$: The KeyGenSP algorithm takes the master public key mpk , the master secret key msk , an ID_{SP} of the target service

TABLE I: Frequently Used Notations

Notation	Description
λ	System security parameter
ℓ	The length of data granules
\mathcal{DG}_n	A set of n data granules
$\mathbb{A} = (\mathbb{M}, \pi)$	An access policy
$\mathcal{T}, \mathcal{R}, \mathcal{D}$	a task, revocation token and download token
\mathcal{L}	The local secret data for issuing tasks
\mathcal{I}	A set of index of data granules in \mathcal{DG}_n
$h, \{\mathcal{H}_i\}_{i \in [3]}$	The efficient hash function used in system

provider and an attribute set $\mathcal{S} \subseteq \mathcal{U}$ as input, and outputs secret key sk_{SP} of the SP.

- $\text{GenSeed}(mpk, ID_{PDO}) \rightarrow (\gamma, \psi)$: The GenSeed algorithm takes the master public key mpk and an ID_{PDO} of the personal data owner as input, and outputs a pair of seed (γ, ψ) .
- $\text{PKeyGenPDO}(\psi) \rightarrow (pk_{PDO}, \beta)$: The PKeyGenPDO algorithm takes the seed ψ as input, and outputs a factor β as well as a pk_{PDO} .
- $\text{SKeyGenPDO}(\gamma, \beta) \rightarrow sk_{PDO}$: The SKeyGenPDO algorithm takes the seed γ and the factor β as input, and outputs a secret key sk_{PDO} of the PDO.
- $\text{Encapsulate}(mpk, sk_{PDO}, \mathcal{DG}_n, \mathbb{A}) \rightarrow (DCI, \mathcal{L}, DC)$: The Encapsulate algorithm takes the master public key mpk , the secret key sk_{PDO} , a set of n data granules \mathcal{DG}_n and an access policy \mathbb{A} as input, and outputs the data capsule identifier DCI , a local secret parameter \mathcal{L} and a data capsule DC corresponding to the DCI .
- $\text{TaskIssue}(mpk, sk_{PDO}, ID_{SP}, \mathcal{DG}_n, \mathcal{I}, \mathcal{L}, t) \rightarrow (\mathcal{T}, \mathcal{R}, \mathcal{D})$: The Taskissue algorithm takes the master public key mpk , the secret key sk_{PDO} , an ID_{SP} , a set of n data granules \mathcal{DG}_n , a set of indices \mathcal{I} for \mathcal{DG}_n ¹, a local secret parameter \mathcal{L} corresponding to \mathcal{DG}_n and a task expiry time t as input, and outputs a task \mathcal{T} , a revocation token \mathcal{R} and a download token \mathcal{D} .
- $\text{AccessDC}(mpk, sk_{SP}, DCI, \mathcal{T}, pk_{PDO}) \rightarrow P_{\mathcal{T},1}$: The AccessDC algorithm takes the master public key mpk , the secret key sk_{SP} , a data capsule identifier DCI , a task \mathcal{T} and the pk_{PDO} that issues the \mathcal{T} as input, and outputs a download parameter $P_{\mathcal{T},1}$.
- $\text{DownloadDC}(DCI, \mathcal{D}, P_{\mathcal{T},1}) \rightarrow DC$: The DownloadDC algorithm takes the data capsule identifier DCI , a download token \mathcal{D} corresponding to DCI and a download parameter $P_{\mathcal{T},1}$ as input, and outputs the DC .
- $\text{DecDC}(mpk, sk_{SP}, DCI, DC, \mathcal{T}, P_{\mathcal{T},1}) \rightarrow \{dg_w\}_{w \in \mathcal{I}}$: The DecDC algorithm takes the master public key mpk , the secret key sk_{SP} , a data capsule identifier DCI , a data capsule DC , a task \mathcal{T} and a download parameter $P_{\mathcal{T},1}$ as input, and outputs a set of data granules $\{dg_w\}_{w \in \mathcal{I}}$.
- $\text{UpdateDC}(mpk, DCI, DC, \mathcal{R}) \rightarrow (DCI', DC')$: The UpdateDC algorithm takes the master public key mpk ,

¹ \mathcal{I} represents a set of indices of data granules in \mathcal{DG}_n to be shared with the SP.

a data capsule identifier DCI , a data capsule DC and the revocation token \mathcal{R} as input, and outputs an updated DCI' and an updated DC' .

B. Definition of Correctness

The correctness of TD-DCSS is that when a service provider possesses a task corresponding to the data capsule and its attributes satisfy the access policy linked to the data capsule, it can successfully recover the data encrypted in both the data capsule and the task.

Definition 3 (Correctness): Our TD-DCSS scheme is correct, if $\forall \lambda \in \mathbb{N}$ and $S \models \mathbb{A}$, we have

$$\Pr[\text{DecDC}(mpk, sk_{SP}, DCI, DC, \mathcal{T}, P_{\mathcal{T},1}) = \{dg_w\}_{w \in \mathcal{I}}] = 1$$

where $P_{\mathcal{T},1} = \text{AccessDC}(mpk, sk_{SP}, DCI, \mathcal{T}, pk_{PDO})$ and the probability is taken with respect to the choice of

$$\begin{aligned} (mpk, msk) &\leftarrow \text{Setup}(1^\lambda), \\ (\gamma, \psi) &\leftarrow \text{GenSeed}(mpk, ID_{PDO}), \\ sk_{PDO}, pk_{PDO} &\leftarrow \text{PKeyGenPDO}(\psi), \text{SKeyGenPDO}(\gamma, \beta), \\ sk_{SP} &\leftarrow \text{KeyGenSP}(mpk, msk, ID_{SP}, S \subseteq \mathcal{U}), \\ (DCI, \mathcal{L}, DC) &\leftarrow \text{Encapsulate}(mpk, sk_{PDO}, \mathcal{D}\mathcal{G}_n, \mathbb{A}), \\ (\mathcal{T}, \mathcal{R}, \mathcal{D}) &\leftarrow \text{TaskIssue}(mpk, sk_{PDO}, ID_{SP}, \mathcal{D}\mathcal{G}_n, \mathcal{I}, \mathcal{L}, t). \end{aligned}$$

The correctness of the scheme ensures that authorized parties can recover the data as intended.

C. Definition of Soundness

In TD-DCSS, we define the soundness to cover the following three cases in which system participants do not act as expected, which are not addressed in security.

- *Case 1 (Soundness of AccessDC):* If the secret key inside the task does not match the public key inside the AccessDC algorithm, or the factor in \mathcal{L} inside the task does not match the factor in \mathcal{L} inside the data capsule identifier DCI , the AccessDC algorithm would be unable to correctly recover the download parameter.
- *Case 2 (Soundness of DecDC):* The download parameter and the result of ABE-related decryption are tied to the ID of an SP. This means that the probability of an SP holding a task colluding with another SP, whose attributes satisfy the access policy, to recover the data granules successfully is negligible.
- *Case 3 (No Leakiness of UpdateDC):* The revocation token is sent to the CS, which possesses partial information about the local parameter \mathcal{L} . However, the CS lacks complete knowledge of the parameters within \mathcal{L} , which means that the CS can successfully recover the important parameters inside \mathcal{L} with a negligible probability.

D. Definition of Security

The formal definition of security for the TD-DCSS scheme is described as follows.

Definition 4 (IND-CPA): The IND-CPA security model of $DCSS_{\mathcal{T}\mathcal{D}}$ can be described through a game involving an adversary \mathcal{A} and a challenger \mathcal{C} .

Init. \mathcal{A} submits a challenge access policy \mathbb{A}^* to \mathcal{C} .

Setup. With a system security parameter $\ell \in \mathbb{N}$, \mathcal{C} runs the system setup algorithm $\text{Setup}(1^\lambda)$ to generate a master public key mpk , a master secret key msk , and initializes the universe of attributes \mathcal{U} . \mathcal{C} sends mpk and \mathcal{U} to \mathcal{A} while securely retaining the msk .

Phase 1. \mathcal{A} can adaptively send a sequence of the following queries to \mathcal{C} .

- $\mathcal{O}_{\text{KSP}}(ID_{SP}, S \subseteq \mathcal{U})$: \mathcal{A} can submit a query to the key generation of SP oracle with the input message containing an $ID_{SP} \in \{0,1\}^*$, and a set of attributes $S \subseteq \mathcal{U}$. If $S \models \mathbb{A}^*$, \mathcal{C} outputs a \perp instead. Otherwise, \mathcal{C} runs the key generation of SP algorithm, $\text{KeyGenSP}(mpk, msk, ID_{SP}, S \subseteq \mathcal{U})$, to generate the secret key sk_{SP} and sends it to \mathcal{A} .
- $\mathcal{O}_{\text{GenSeed}}(ID_{PDO})$: \mathcal{A} can submit a query to the seed generation oracle with the input message containing an identifier $ID_{PDO} \in \{0,1\}^*$. \mathcal{C} runs the seed generation algorithm, $\text{GenSeed}(mpk, ID_{PDO})$, to generate two seeds γ, ψ and sends ψ to \mathcal{A} .
- $\mathcal{O}_{\text{PKPDO}}(\psi)$: \mathcal{A} can submit a query to the public key generation of PDO oracle with the input message containing a group element $\psi \in \mathbb{G}_2$. \mathcal{C} runs the public key generation of PDO algorithm, $\text{PKeyGenPDO}(\psi)$, to generate a public key pk_{PDO} , a factor β and sends pk_{PDO}, β to \mathcal{A} .
- $\mathcal{O}_{\text{SKPDO}}(ID_{PDO}, \beta)$: \mathcal{A} can submit a query to the public key generation of PDO oracle with the input message containing an $ID_{PDO} \in \{0,1\}^*$ and a factor $\beta \in \mathbb{Z}_p^*$. \mathcal{C} runs the secret key generation of PDO algorithm, $\text{SKeyGenPDO}(\gamma, \beta)$, to generate a secret key sk_{PDO} and sends sk_{PDO} to \mathcal{A} .

In phase 1, when \mathcal{A} queries to generate a pair of public key and secret key of PDO, the seed generation oracle $\mathcal{O}_{\text{GenSeed}}(ID_{PDO})$, the public key generation oracle $\mathcal{O}_{\text{PKPDO}}(\psi)$, and the secret generation oracle $\mathcal{O}_{\text{SKPDO}}(\gamma, \beta)$ must be queried in a sequential order.

Challenge. \mathcal{A} submits two sets of data granules (m_0, m_1) with the same length to \mathcal{C} . \mathcal{C} flips a random coin $b \in \{0,1\}$. \mathcal{C} then runs the data capsule encapsulation algorithm $\text{Encapsulate}(mpk, \cdot, m_b, \mathbb{A}^*)$, to generate $(DCI^*, DC^*, \mathcal{L}^*)$ and sends DCI^*, DC^* to \mathcal{A} .

Phase 2. Under the previous restrictions, \mathcal{A} continues to query \mathcal{C} .

Guess. \mathcal{A} outputs a guess b' . If $b' = b$, \mathcal{A} wins the game. The advantage of \mathcal{A} in this game is defined as

$$\text{Adv}_{DCSS_{\mathcal{T}\mathcal{D}}, \mathcal{A}}^{\text{IND-CPA}}(\lambda) = |\Pr[b = b'] - \frac{1}{2}|$$

V. SYSTEM ARCHITECTURE

A. System Model

The Fig. 2 describes a task-driven data capsule sharing system (TD-DCSS), in which personal data owners (PDOs) want to share personal data with multiple service providers for utilization of their services. PDOs encapsulate their data into

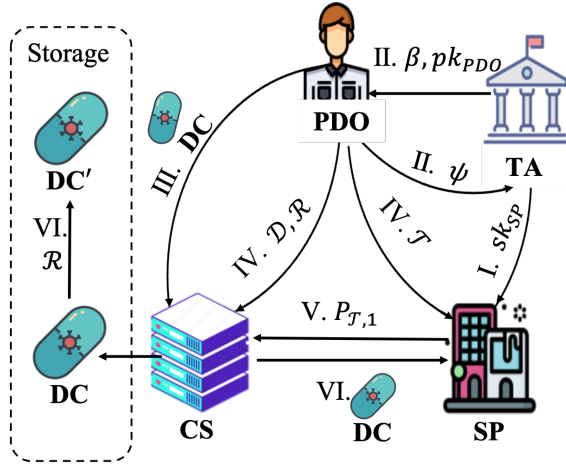


Fig. 2: System model.

data capsules (DCs) under an access policy and upload DCs to the cloud storage (CS). PDOs selectively share their data and grant authorization to service providers (SPs) by issuing tasks and sending tasks to SPs. Since PDOs require that the task be a one-time token, the CS revokes the permissions of SPs after they use the task by updating part of the data capsules. A TA, a CS, PDOs, and SPs as system entities are given the details below.

- 1) *TA*: The trusted authority (TA), such as the government, bears responsibility for initiating the system parameters, the universe of attributes and issues the secret keys to SPs (see I in Fig. 2) as well as the public keys to PDOs (see II in Fig. 2).
- 2) *CS*: The cloud storage (CS) as a semi-trusted entity provides abundant storage capacity and is responsible for archiving data capsules (see III in Fig. 2) and the associated revocation/download tokens (see IV in Fig. 2). It updates a data capsule according to the tokens provided by a PDO (see VI in Fig. 2) and rejects the request from expired tasks.
- 3) *PDO*: Personal data owners (PDOs) as the predominant participants in the system can break down their data into data granules, subsequently encapsulate these granules into a data capsule, and outsource it to the CS (see III in Fig. 2). Furthermore, PDOs can issue tasks to the designated service providers (see IV in Fig. 2) and submit the download token as well as the revocation token to the CS (see IV in Fig. 2).
- 4) *SP*: The service providers (SPs) who have many attributes and provide service for PDOs are the important system participants. They accept tasks (see V in Fig. 2) and decrypt the data capsules to access personal data that is necessitated for providing service.

B. System Phases

Our TD-DCSS scheme has the following several functions: Setup, KeyGenSP, GenSeed, PKeyGenPDO, SKeyGenPDO, Encapsulate, TaskIssue, AccessDC, DownloadDC, DecDC, and UpdateDC, which are defined in Section IV-A. These

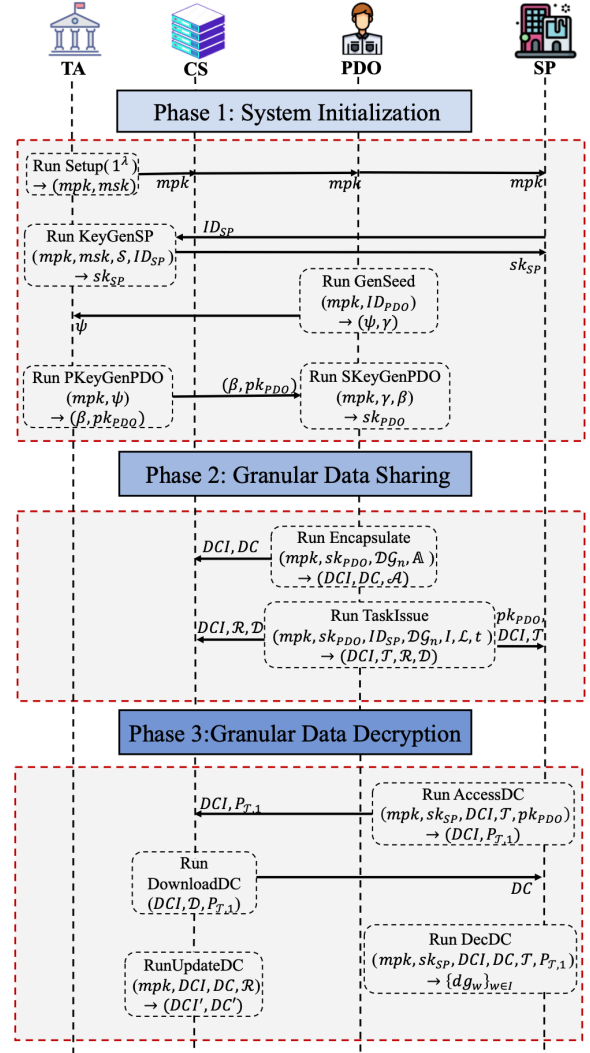


Fig. 3: Data sharing phases in our system.

functions are used in the following phases: System Initialization, Granular Data Sharing, and Granular Data Decryption. As shown in Fig.3, the three phases are described as follows:

- 1) *System Initialization*: As shown in Fig. 3, the TA generates the master public key mpk and master secret key msk by running $Setup(1^\lambda)$. TA runs $KeyGenSP(mp_k, msk, ID_{SP}, \mathcal{S} \subseteq \mathcal{U})$ with an ID_{SP} and a set of attributes $\mathcal{S} \subseteq \mathcal{U}$ to generate a secret key sk_{SP} for each SP in the system, where \mathcal{U} is the universe of attributes. PDO runs $GenSeed(mp_k, ID_{PDO})$ with its ID_{PDO} to get a pair of seed γ, ψ . Then, PDO sends ψ to TA. After receiving the ψ , TA selects a random mask $\beta \in \mathbb{Z}_p$ and runs $PKeyGenPDO(\psi)$ to get pk_{PDO} and sends β, pk_{PDO} to PDO. Then, PDO runs $SKeyGenPDO(\gamma, \beta)$ to generate sk_{PDO} .
- 2) *Granular Data Sharing*: As shown in Fig. 3, to encapsulate the personal data into a data capsule, PDO first selects an access policy \mathbb{A} . Next, PDO breaks down personal data into data granules DG_n with length ℓ . Then, PDO runs $Encapsulate(mp_k, sk_{PDO}, DG_n, \mathbb{A})$ to generate the data capsule identifier DCI , the

data capsule DC , and the local secret parameter \mathcal{L} which is stored locally. Finally, PDO sends DCI, DC to the CS. The CS stores DC indexed by DCI . When sharing with service providers (SPs), PDO runs $\text{TaskIssue}(mpk, sk_{PDO}, ID_{SP}, \mathcal{D}\mathcal{G}_n, \mathcal{I}, \mathcal{L}, t)$ to obtain a task \mathcal{T} , a revocation token \mathcal{R} , and a download token \mathcal{D} , with the input sk_{SP} , the data granules $\mathcal{D}\mathcal{G}_n$, the ID_{SP} of the target SP, an expiry time for task t , and a set of indices of data granules in $\mathcal{D}\mathcal{G}_n$ to be shared with the SP. Then PDO sends $pk_{PDO}, DCI, \mathcal{T}$ to the SP, and sends $DCI, \mathcal{R}, \mathcal{D}$ to the CS.

- 3) *Granular Data Decryption*: As shown in Fig. 3, after receiving $pk_{PDO}, DCI, \mathcal{T}$, SP first runs $\text{AccessDC}(mpk, sk_{SP}, DCI, \mathcal{T}, pk_{PDO})$ to obtain the download parameter $P_{\mathcal{T},1}$ and sends it to the CS. The CS checks the validity and timeliness of $P_{\mathcal{T},1}$ by running $\text{DownloadDC}(DCI, \mathcal{D}, P_{\mathcal{T},1})$. If DownloadDC returns DC , the CS sends it to the SP. Meanwhile, the CS runs $\text{UpdateDC}(mpk, DCI, \mathcal{R})$ with \mathcal{R} provided by the PDO to immediately update the DC and revoke the permission of the SP. With the received DC , the SP recovers the data granules by running $\text{DecDC}(mpk, sk_{SP}, DCI, DC, \mathcal{T}, P_{\mathcal{T},1})$.

C. Threat Model

In our task-driven data capsule sharing system, the TA and PDOs have full trust. The TA honestly sets up the system parameters, initializes the universe of attributes, issues secret keys to SPs, and publishes the public keys of PDOs. PDOs can encapsulate their data into data capsules, outsource data capsules into the CS, and issue tasks to SPs for selectively sharing their data with them through our proposed scheme.

The CS, as a semi-trusted party, has huge storage resources to store data capsules from PDOs. The CS also assists PDOs in updating the data capsules using the revocation token from PDOs and runs access control over data capsules when SPs access data capsules using a download parameter. However, the CS may try to learn sensitive information from data capsules, the revocation token, and the download token individually.

The SPs are untrusted and may try to learn unauthorized information from data capsules and tasks. Furthermore, the SPs can be categorized into three types: type 1, type 2, and type 3, where type 1 of SPs have the correct task but their attributes do not satisfy the access policy of the data capsules, type 2 of SPs have attributes that satisfy the access policy of data capsules but do not have the correct tasks, and type 3 of SPs do not have a correct task and their attributes also do not satisfy the access policy of the data capsules.

Moreover, a collusion attack in our system occurs when type 1 and type 2 of SPs combine their attributes and tasks to access data granules within data capsules that they cannot obtain independently.

The soundness and security model of TD-DCSS, as shown in the definition of soundness and security (Section IV), formalize the above threat model, respectively. Specifically, in the IND-CPA model and the soundness, the adversary simulates the CS and the SPs as follows:

- *Untrusted service providers*. The adversary can get many secret keys by querying the secret key generation of the SP oracle many times, which enables the adversary to simulate the three types of SPs. Specifically, the adversary can act as type 1 of SPs that hold a correct task but the attribute set of secret keys does not satisfy the access policy. And the adversary can act as type 2 of SPs, if the attribute set associated with the issued secret keys satisfies the challenge message but does not have a task. Moreover, the adversary can act as type 3 of SPs, if the attribute set associated with the issued secret keys do not satisfy the challenge message and do not have a correct task. After the adversary gets secret keys, the adversary can obtain a ciphertext in the challenge phase and update the ciphertext in phase 2. The adversary wins the IND-CPA game by correctly guessing the bits of the underlying message.
- *Semi-trusted CS*. The adversary can obtain the original ciphertext and the updated ciphertext without querying the secret key generation of the SP oracle. The adversary wins the IND-CPA game by guessing the underlying message of the ciphertext correctly.

VI. PROPOSED SCHEME

A. The Proposed TD-DCSS Scheme

Our TD-DCSS scheme is shown in Fig. 4 and Fig. 5. Specifically, when selectively sharing data with service providers, the personal data owner (PDO) first encapsulates its data into a data capsule, and then issues a task. Subsequently, the PDO sends the task to a service provider (SP) and its PK_{PDO} . The SP runs the AccessDC algorithm to check whether the PK_{PDO} matches the public key of the task issuer and recover the download token. Then the SP successfully downloads the data capsule from the CS if the task is within its validity period. Finally, the SP can successfully recover the data if its attributes satisfy the access policy associated with the data capsule. In our scheme, the function FABEO.Dec serves as a tool that decrypts the ABE ciphertext, and our scheme is also designed to adopt other CP-ABE schemes that support useful features easily.

B. Correctness of Proposed TD-DCSS Scheme

To decrypt a data capsule (or an updated data capsule generated by the UpdateDC algorithm) and recover the data granules, the SP first has to decrypt the task to obtain the download token $P_{\mathcal{T},1}$ through the AccessDC algorithm. Subsequently, the SP can recover the data granules $\{dg_w\}_{w \in \mathcal{I}}$ using the DecDC algorithm. The correctness of the AccessDC , DecDC , and UpdateDC algorithms is demonstrated below.

Correctness of AccessDC : $P_{\mathcal{T},1}^*$ can be computed in AccessDC algorithm as shown in Fig. 4.

If the task is associated with the ID_{SP} and the identifier DCI of DC , then

$$\begin{aligned} P_{\mathcal{T},1}^* &= \frac{e(sk_2, DCI) \cdot e(sk_4, pk_{PDO})}{e(\mathcal{T}_1, sk_3)} \\ &= \frac{e(sk_2, DCI)}{e(\mathcal{T}_1, sk_3)} \cdot e(sk_4, pk_{PDO}) \end{aligned}$$

Setup(1^λ): Initialize the universe of attributes \mathcal{U} . Run **GroupGen**(1^λ) to obtain $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e, g_1, g_2)$, where $e : \mathbb{G}_1 \times \mathbb{G}_2 \rightarrow \mathbb{G}_T$ is a non-degenerate bilinear map (Section III-B). Pick $\alpha \in \mathbb{Z}_p$. Select the following hash functions,

$$h : \{0, 1\}^* \rightarrow \mathbb{Z}_p^*, \quad \mathcal{H}_1 : \{0, 1\}^* \rightarrow \mathbb{G}_1, \quad \mathcal{H}_2 : \mathbb{G}_T \rightarrow \{0, 1\}^\ell, \quad \mathcal{H}_3 : \mathbb{G}_2^2 \times \{0, 1\}^\ell \times \mathbb{G}_1^{\tau+n_1} \rightarrow \mathbb{Z}_p^*,$$

where n_1 is the number of rows in the policy matrix \mathbb{M} and $\tau = \max_{i \in [n_1]} \rho(i)$, where

$\rho(i) = |\{z | \pi(z) = \pi(i), z \leq i\}|$ and $\pi : [n_1] \rightarrow \mathcal{U}$. Output the master public key and master private key as

$$mpk = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e, g_1, g_2, g_2^\alpha, h, \{\mathcal{H}_i\}_{i \in [3]}), \quad msk = \alpha.$$

KeyGenSP($mpk, msk, ID_{SP}, \mathcal{S} \subseteq \mathcal{U}$): Pick $r \in \mathbb{Z}_p$. For each $s \in \mathcal{S}$, compute $sk_{1,s} = \mathcal{H}_1(s)^r$. Parsing $msk = \alpha$, compute

$$sk_2 = \mathcal{H}_1(ID_{SP})^\alpha \cdot \mathcal{H}_1(|\mathcal{U}| + 1)^r, \quad sk_3 = g_2^r, \quad sk_4 = \mathcal{H}_1(ID_{SP})^r,$$

where $|\mathcal{U}|$ denotes the number of the universe of attributes. Output $sk_{SP} = (\mathcal{S}, \{sk_{1,s}\}_{s \in \mathcal{S}}, sk_2, sk_3, sk_4)$.

GenSeed(mpk, ID_{PDO}): Pick $\sigma \in \mathbb{Z}_p^*$, compute $\gamma = h(ID_{PDO} || \sigma)$ and $\psi = g_2^\sigma$. Output γ and ψ . Send ψ to TA.

PKeyGenPDO(ψ): With the input seed $\psi \in \mathbb{G}_2$, pick $\beta \in \mathbb{Z}_p$ and compute PDO's public key $pk_{PDO} = \psi^\beta$. Output pk_{PDO} and the factor β . Send pk_{PDO}, β to PDO.

SKeyGenPDO(γ, β): With the input seed $\gamma, \beta \in \mathbb{Z}_p$, set $\gamma = \gamma\beta$ and output $sk_{PDO} = \gamma$.

Encapsulate($mpk, sk_{PDO}, \mathcal{DG}_n, \mathbb{A}$): Initializing $c = 1$ and the length of data granule ℓ , for $k \in [c]$, pick $a_k \in \{0, 1\}^\ell, d_k \in \mathbb{Z}_p^*$. Compute $P_1 = \bigoplus_{k=1}^c a_k, DCI = \prod_{k=1}^c g_2^{d_k}$. Pick $y \in \mathbb{Z}_p$, compute $C_1 = g_2^y$. Then, compute

$$P_2 = \mathcal{H}_2(e(g_1^{sk_{PDO}}, C_1)), \quad C_2 = (\bigoplus \{\mathcal{DG}_n\}) \oplus P_1 \oplus P_2,$$

where $\bigoplus \{\mathcal{DG}_n\}$ denotes the XOR of all elements in the set \mathcal{DG}_n . Parse $\mathbb{A} = (\mathbb{M}, \pi)$. Suppose \mathbb{M} has the shape $(n_1 \times n_2)$. Pick $\mathbf{v} \in \mathbb{Z}_p^{n_2-1}, \mathbf{y}' \in \mathbb{Z}_p^\tau$. For $j \in [\tau]$, compute $C_{3,j} = g_2^{y'_j}$. For $i \in [n_1]$, compute

$$C_{4,i} = \mathcal{H}_1(|\mathcal{U}| + 1)^{\mathbb{M}_i \cdot (\mathbf{y}' || \mathbf{v})^\top} \cdot \mathcal{H}_1(\pi(i))^{y'[\rho(i)]}$$

Next, compute $\delta = \mathcal{H}_3(DCI, C_1, C_2, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]})$ and $V = g_1^{\delta \sum_{k=1}^c d_k}$. Output DCI ,

$$\mathcal{L} = (DCI, P_1 = \bigoplus_{k=1}^c a_k, d = \sum_{k=1}^c d_k, y), \quad DC = (\mathbb{A}, C_1, C_2, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]}, V).$$

TaskIssue($mpk, sk_{PDO}, ID_{SP}, \mathcal{DG}_n, \mathcal{I}, \mathcal{L}, t$): $\mathcal{I} \subseteq [n]$ represents the indices of data granules that PDO wants to share with the SP. Parse $\mathcal{L} = (DCI, P_1, d, y)$.

1) **Task Generation Phase**: Using sk_{PDO} , compute $\mathcal{T}_1 = \mathcal{H}_1(ID_{SP})^{sk_{PDO}} \cdot \mathcal{H}_1(|\mathcal{U}| + 1)^d$. Using g_2^α from mpk , compute

$$P_{\mathcal{T},1} = e(\mathcal{H}_1(ID_{SP})^d, g_2^\alpha), \quad P_{\mathcal{T},2} = e(\mathcal{H}_1(ID_{SP})^y, g_2^\alpha), \quad P_{\mathcal{T}} = P_{\mathcal{T},1} \cdot P_{\mathcal{T},2}.$$

Compute $\mathcal{T}_2 = P_{\mathcal{T}} \cdot e(g_1^{sk_{PDO}}, g_2^y)$. For $w \in \mathcal{I}$, pick $r_w \in \mathbb{Z}_p$ and compute

$$P_w = e(\mathcal{H}_1(ID_{SP})^{r_w}, g_2^\alpha), \quad \mathcal{T}_{w,1} = (\bigoplus \{\mathcal{DG}_n \setminus dg_w\}) \oplus P_1 \oplus \mathcal{H}_2(P_w), \quad \mathcal{T}_{w,2} = P_{\mathcal{T}} \cdot P_w.$$

2) **Parameter Generation Phase**: Pick $a_{c+1} \in \{0, 1\}^\ell, d_{c+1} \in \mathbb{Z}_p^*$. Compute

$$DCI' = DCI \cdot g_2^{d_{c+1}} = \prod_{k=1}^{c+1} g_2^{d_k}, \quad P'_1 = P_1 \oplus a_{c+1} = \bigoplus_{k=1}^{c+1} a_k, \quad d' = d + d_{c+1} = \sum_{k=1}^{c+1} d_k.$$

Update $\mathcal{L} = (DCI', P'_1, d', y)$. Compute $\mathcal{R}_1 = g_1^{d'} = g_1^{\sum_{k=1}^{c+1} d_k}$. Set $\mathcal{T}_w = (\mathcal{T}_{w,1}, \mathcal{T}_{w,2})$. Output DCI ,

$$\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3 = \{\mathcal{T}_w\}_{w \in \mathcal{I}}), \quad \mathcal{R} = (\mathcal{R}_1, \mathcal{R}_2 = DCI', \mathcal{R}_3 = a_{c+1}), \quad \mathcal{D} = (\mathcal{D}_1 = P_{\mathcal{T},1}, \mathcal{D}_2 = t).$$

AccessDC($mpk, sk_{SP}, DCI, \mathcal{T}, pk_{PDO}$)^a: Using \mathcal{T}_1 in \mathcal{T} and sk_2, sk_3, sk_4 in sk_{SP} , compute and output

$$P_{\mathcal{T},1} = \frac{e(sk_2, DCI) \cdot e(sk_4, pk_{PDO})}{e(\mathcal{T}_1, sk_3)}$$

DownloadDC($DCI, \mathcal{D}, P_{\mathcal{T},1}$): Determine whether $t_{now} < \mathcal{D}_2$ and $P_{\mathcal{T},1} = \mathcal{D}_1$ holds, where t_{now} denotes the current time. Return \perp if not. Otherwise, find the target DC by DCI and output DC .

^aService providers run the AccessDC algorithm to check the authenticity of the task issuer's identity, ensuring it corresponds to the expected ID (pk_{PDO}). Upon successful verification, the algorithm outputs a download parameter for further proceedings.

Fig. 4: Our construction of TD-DCSS.

DecDC($mpk, sk_{SP}, DCI, DC, \mathcal{T}, P_{\mathcal{T},1}$):

- 1) *DC Integrity Verification Phase.*: Parse $DC = (\mathbb{A}, C_1, C_2, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]}, V)$. Compute $\delta = \mathcal{H}_3(DCI, C_1, C_2, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]})$. Return \perp if $e(V, g_2) = e(g_1^\delta, DCI)$ not holds.
- 2) *Attribute-based Decryption Phase.*: Compute $P_{\mathcal{T},2} = \text{FABEO.Dec}(mpk, \mathbb{A}, \mathcal{S}, CT_{\text{FABEO}}, sk_{SP})$ if \mathcal{S} in sk_{SP} satisfies \mathbb{A} , where $CT_{\text{FABEO}} = \{C_1, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]}\}$. The detail of FABEO.Dec is shown as below: If \mathcal{S} in sk_{SP} satisfies \mathbb{A} , there exists constants $\{\gamma_i\}_{i \in [I]}$ s.t. $\sum_{i \in [I]} \gamma_i \mathbb{M}_i = (1, 0, \dots, 0)$, where $I = |\mathcal{S} \cap \{\pi(i) | i \in [n_1]\}|$.

$$P_{\mathcal{T},2} = e(sk_2, C_1) \cdot \frac{\prod_{j \in [\tau]} e(\prod_{i \in [I], \rho(i)=j} (sk_{1,\pi(i)})^{\gamma_i}, C_{3,j})}{e(\prod_{i \in [I]} (C_{4,i})^{\gamma_i}, sk_3)}$$

Return \perp if $P_{\mathcal{T},2}$ is invalid.

- 3) *Message Recovery Phase.*: Based on the results of above two phases, compute $P_{\mathcal{T}} = P_{\mathcal{T},1} \cdot P_{\mathcal{T},2}$. Reconstruct P_2 by computing

$$\mathcal{H}_2\left(\frac{\mathcal{T}_2}{P_{\mathcal{T}}}\right) = \mathcal{H}_2\left(\frac{e(\mathcal{H}_1(ID_{SP})^{\sum_{k=1}^c d_k}, g_2^\alpha) \cdot e(\mathcal{H}_1(ID_{SP})^y, g_2^\alpha) \cdot e(g_1^\gamma, g_2^y)}{e(\mathcal{H}_1(ID_{SP}), g_2)^{\alpha y} \cdot e(\mathcal{H}_1(ID_{SP})^\alpha, \prod_{k=1}^c g_2^{d_k})}\right)$$

For $w \in \mathcal{I}$, compute

$$P_w = \frac{\mathcal{T}_{w,2}}{P_{\mathcal{T}}}, \quad dg_w = C_2 \oplus \mathcal{T}_{w,1} \oplus \mathcal{H}_2(P_w) \oplus P_2.$$

Output $\{dg_w\}_{w \in \mathcal{I}}$.

UpdateDC($mpk, DCI, DC, \mathcal{R}$)^a: Update DC by setting $DCI' = \mathcal{R}_2$ and computing

$$C'_2 = C_2 \oplus \mathcal{R}_3, \quad \delta' = \mathcal{H}_3(DCI', C_1, C'_2, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]}), \quad V' = \mathcal{R}'_1.$$

Output the updated $DCI = DCI', DC = (\mathbb{A}, C_1, C'_2, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]}, V')$.

^aThe UpdateDC algorithm modifies segments within data capsules. Furthermore, upon updating, the new data capsule replaces the old one, thus maintaining storage space without expansion.

Fig. 5: Our construction of TD-DCSS (Cont).

For simplicity, we can divide the above equation into two parts: the left part L and the right part R . We can compute

$$\begin{aligned} L &= \frac{e(sk_2, DCI)}{e(\mathcal{T}_1, sk_3)} \\ &= \frac{e(\mathcal{H}_1(ID_{SP})^\alpha \cdot \mathcal{H}_1(|\mathcal{U}| + 1)^r, \prod_{k=1}^c g_2^{d_k})}{e(\mathcal{H}_1(ID_{SP})^\gamma \cdot \mathcal{H}_1(|\mathcal{U}| + 1)^{\sum_{k=1}^c d_k}, g_2^\alpha)} \\ &= \frac{e(\mathcal{H}_1(ID_{SP}), g_2)^\alpha \sum_{k=1}^c d_k \cdot e(\mathcal{H}_1(|\mathcal{U}| + 1), g_2)^r \sum_{k=1}^c d_k}{e(\mathcal{H}_1(ID_{SP}), g_2)^{\gamma r} \cdot e(\mathcal{H}_1(|\mathcal{U}| + 1), g_2)^{r \sum_{k=1}^c d_k}} \\ &= \frac{e(\mathcal{H}_1(ID_{SP}), g_2)^\alpha \sum_{k=1}^c d_k}{e(\mathcal{H}_1(ID_{SP}), g_2)^{\gamma r}} \end{aligned}$$

$$R = e(sk_4, pk_{PDO})$$

$$\begin{aligned} L \cdot R &= \frac{e(\mathcal{H}_1(ID_{SP}), g_2)^\alpha \sum_{k=1}^c d_k}{e(\mathcal{H}_1(ID_{SP}), g_2)^{\gamma r}} \cdot e(sk_4, pk_{PDO}) \\ &= \frac{e(\mathcal{H}_1(ID_{SP}), g_2)^\alpha \sum_{k=1}^c d_k}{e(\mathcal{H}_1(ID_{SP}), g_2)^{\gamma r}} \cdot e(\mathcal{H}_1(ID_{SP}), g_2)^{\gamma r} \\ &= e(\mathcal{H}_1(ID_{SP}), g_2)^\alpha \sum_{k=1}^c d_k = P_{\mathcal{T},1} \end{aligned}$$

Correctness of DecDC: $\{dg_w\}_{w \in \mathcal{I}}$ can be computed in DecDC algorithm as shown in Fig. 5. In detail, δ^* can be computed as $\delta^* = \mathcal{H}_3(DCI, C_1, C_2, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]})$ and the correctness of δ^* can be verified as

$$e(g_1^{\delta^*}, DCI) = e(g_1, g_2)^{\delta^* \sum_{k=1}^c d_k} = e(g_1^{\delta^* \sum_{k=1}^c d_k}, g_2) = e(V, g_2)$$

The correctness of FABEO.Dec algorithm has been demonstrated [29], and it outputs the correct $P_{\mathcal{T},2}$. Based on the above results, we can compute

$$\begin{aligned} \frac{\mathcal{T}_2}{P_{\mathcal{T}}} &= \frac{e(\mathcal{H}_1(ID_{SP})^{y + \sum_{k=1}^c d_k}, g_2^\alpha) \cdot e(g_1^\gamma, g_2^y)}{P_{\mathcal{T},1} \cdot P_{\mathcal{T},2}} \\ &= \frac{e(\mathcal{H}_1(ID_{SP})^{y + \sum_{k=1}^c d_k}, g_2^\alpha) \cdot e(g_1^\gamma, g_2^y)}{e(\mathcal{H}_1(ID_{SP}), g_2)^{\alpha \sum_{k=1}^c d_k} \cdot e(\mathcal{H}_1(ID_{SP}), g_2)^{\alpha y}} \\ &= e(g_1^\gamma, g_2^y) \end{aligned}$$

We have $P_2^* = \mathcal{H}_2\left(\frac{\mathcal{T}_2}{P_{\mathcal{T}}}\right)$. For $w \in \mathcal{I}$, dg_w^* can be computed as

$$\begin{aligned} dg_w^* &= C_2 \oplus \mathcal{T}_{w,1} \oplus \mathcal{H}_2\left(\frac{\mathcal{T}_{w,2}}{P_{\mathcal{T}}}\right) \oplus P_2^* \\ &= \oplus \{\mathcal{DG}_n\} \oplus P_1 \oplus P_2 \oplus \mathcal{T}_{w,1} \oplus \mathcal{H}_2\left(\frac{\mathcal{T}_{w,2}}{P_{\mathcal{T}}}\right) \oplus P_2^* \\ &= \oplus \{\mathcal{DG}_n\} \oplus P_1 \oplus \{\mathcal{DG}_n \setminus dg_w\} \oplus P_1 \oplus \mathcal{H}_2(P_w) \oplus \mathcal{H}_2(P_w)^* \\ &= dg_w \end{aligned}$$

Correctness of UpdateDC: The updated DCI' can be computed as $DCI' = g_2^{\sum_{k=1}^{c+1} d_k} = g_2^{d'}$. And the updated DC' can be computed by computing the updated C'_2 and V' . Specifically, the C'_2 can be computed as $C'_2 = C_2 \oplus \mathcal{R}_3 = (\oplus \{\mathcal{DG}_n\}) \oplus P_1 \oplus P_2 \oplus a_{c+1} = (\oplus \{\mathcal{DG}_n\}) \oplus P'_1 \oplus P_2$. And the V' can be computed as $V' = \mathcal{R}'_1 = g_1^{\delta' \cdot \sum_{k=1}^{c+1} d_k} = g_1^{\delta' \cdot d'}$, where $\delta' = \mathcal{H}_3(DCI', C_1, C'_2, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]})$.

TABLE II: Property & Functionality Comparisons

Scheme	Selective Sharing	Informed-consent based Authorization	Permission Revocation	CR	TR	ER	Groups
DYL+ [22]	✗	✗	✓	✓	✗	✗	Symmetric
ZSL+ [23]	✗	✗	✓	✓	✗	✗	Symmetric
NHS+ [25]	✗	✗	✗	✓	N/A	✓	Symmetric
YSX+ [27]	✓	✓	✗	✗	✓	✗	Symmetric
TD-DCSS	✓	✓	✓	✓	✓	✓	Asymmetric

“CR”, “TR”, and “ER” are the abbreviation of “Collusion Resistance”, “Tamper Resistance”, and “EDoS Resistance”, respectively.

C. Soundness of Proposed TD-DCSS Scheme

The soundness of our proposed TD-DCSS scheme encompasses three cases: soundness of AccessDC, soundness of DecDC, and no leakiness of UpdateDC.

For soundness of AccessDC, the PDO utilizes the secret key sk_{PDO} to generate \mathcal{T}_1 , and the PDO’s public key takes the form $g_2^{sk_{PDO}}$. If the sk_{PDO} used in \mathcal{T}_1 does not match $pk_{PDO} = g_2^{sk_{PDO}}$, the probability of recovering the correct $P_{\mathcal{T},1}$ is negligible. Furthermore, if an SP possesses an “old” or a “wrong” task and attempts to access the DC, the SP cannot successfully recover the correct $P_{\mathcal{T},1}$. In the first case, the \mathcal{T}_1 in the task is bound with the factor $d = \sum_{k=1}^c d_k$, the identifier DCI is bound with the factor $d^* = \sum_{k=1}^{c+1} d_k$, and thus the components $e(\mathcal{H}_1(|\mathcal{U}| + 1), g_2)^{r \cdot d}$ and $e(\mathcal{H}_1(|\mathcal{U}| + 1), g_2)^{r \cdot d^*}$ cannot be canceled. In the second case, if an SP* holds a “wrong” task generated under the ID_{SP} , the SP* cannot recover a correct $P_{\mathcal{T},1}$ since

$$P_{\mathcal{T},1}^* = \frac{e(\mathcal{H}_1(ID_{SP}^*)^\alpha, g_2^d) \cdot e(\mathcal{H}_1(ID_{SP}^*)^r, g_2^{sk_{PDO}})}{e(\mathcal{H}_1(ID_{SP})^{sk_{PDO}}, g_2^r)} \neq P_{\mathcal{T},1}$$

For soundness of DecDC, suppose an SP with an identifier ID_{SP} holds a task \mathcal{T} , and another SP* with an identifier ID_{SP^*} possesses sufficient attributes to satisfy the access policy associated with the DC. The SP first recovers $P_{\mathcal{T},1} = e(\mathcal{H}_1(ID_{SP}), g_2)^{\alpha \cdot d}$, and then the SP* recovers $P_{\mathcal{T},2}^* = e(\mathcal{H}_1(ID_{SP^*}), g_2)^{y \cdot \alpha}$. To recover data granules, they have to reconstruct P_2 by computing

$$P_2 = \mathcal{H}_2\left(\frac{\mathcal{T}_2}{P_{\mathcal{T},1} \cdot P_{\mathcal{T},2}^*}\right) = \mathcal{H}_2\left(\frac{e(\mathcal{H}_1(ID_{SP})^y, g_2^\alpha) \cdot e(g_1^\gamma, g_2^y)}{e(\mathcal{H}_1(ID_{SP}^*), g_2)^{\alpha y}}\right)$$

The probability of the SP using $P_{\mathcal{T},2}^*$ to successfully recover P_2 is negligible, as $P_{\mathcal{T},2} = e(\mathcal{H}_1(ID_{SP}), g_2)^{y \cdot \alpha} \neq P_{\mathcal{T},2}^*$. The soundness of DecDC indicates that the proposed scheme is resistant to collusion attacks.

For no leakiness of UpdateDC, let’s consider a scenario where the CS collects information during the update operation. The CS possesses information a_2, a_3, \dots, a_c but lacks any knowledge about a_1 and cannot recover P_1 with a non-negligible probability. Similarly, the CS is aware of $g_2^{d_1}, g_2^{d_2}, \dots, g_2^{d_k}, g_1^{\sum_{k=2}^c d_k}$ but lacks any knowledge about $g_1^{d_1}$ and therefore cannot successfully recover $\sum_{k=1}^c d_k$ with a non-negligible probability.

D. Security Analysis of Proposed TD-DCSS Scheme

Theorem 1: Assume that DBDH assumption holds, then the proposed $DCSS_{TD}$ is IND-CPA secure.

Proof: If a polynomial time adversary \mathcal{A} is capable of successfully breaching the TD-DCSS, then by interacting with \mathcal{A} , another algorithm \mathcal{C} can be easily constructed to exploit the DBDH assumption and break the IND-CPA security of the TD-DCSS. Given a DBDH tuple $(g^a, g^b, g^c, h^a, h^b, h^c, \mathcal{Z})$ to \mathcal{C} , where $g \in \mathbb{G}_1, h \in \mathbb{G}_2$, the goal of \mathcal{A} is to determine whether $\mathcal{Z} = e(g, h)^{abc}$ or $\mathcal{Z} = \mathcal{R}$, where \mathcal{R} is a random element of \mathbb{G}_T .

Init. \mathcal{A} submits a challenge access policy $\mathbb{A}^* = (\mathbb{M}^*, \pi^*)$ to \mathcal{C} .

Setup. \mathcal{C} runs $\text{GroupGen}(1^\lambda)$ algorithm to generate $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e, g_1, g_2)$. Then \mathcal{C} sets $g_1 = g^a$ and

$g_2 = h^b$. Pick $\alpha \in \mathbb{Z}_p^*$, compute $g_2^\alpha = h^{b\alpha}$. \mathcal{C} sends the $mpk = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e, g^a, h^b, h^{b\alpha}, h, \mathcal{H}_{i \in [3]})$ to \mathcal{C} , where $h, \mathcal{H}_{i \in [3]}$ are collision-resistant hash function. The master private key is set as $msk = (\alpha)$.

Phase 1. \mathcal{A} can adaptively send a sequence of the following queries to \mathcal{C} .

- $\mathcal{O}_{KSP}(ID_{SP}, \mathcal{S} \subseteq \mathcal{U})$: On input $ID_{SP} \in \{0, 1\}^*$ and $\mathcal{S} \subseteq \mathcal{U}$, \mathcal{C} checks whether $\mathcal{S} \models \mathbb{A}^*$ holds. If $\mathcal{S} \models \mathbb{A}^*$ holds, \mathcal{C} aborts and returns \perp . Otherwise, it picks $r \in \mathbb{Z}_p^*$ and sets $sk_{1,s} = \mathcal{H}_1(s)^r, sk_2 = \mathcal{H}_1(ID_{SP})^\alpha \cdot \mathcal{H}_1(|\mathcal{U}| + 1)^r, sk_3 = h^{br}, sk_4 = \mathcal{H}_1(ID_{SP})^r$, where $s \in \mathcal{S}$.
- $\mathcal{O}_{GenSeed}(ID_{PDO})$: On input $ID_{PDO} \in \{0, 1\}^*$, \mathcal{C} picks $\sigma \in \mathbb{Z}_p^*$ and computes $\gamma = h(ID_{PDO} || \sigma)$. Then \mathcal{C} sets $\psi = h^{b\gamma}$. \mathcal{C} inits a table D_1 and records the record (ID_{PDO}, γ) into D_1 . And then \mathcal{C} sends the ψ to \mathcal{A} .
- $\mathcal{O}_{PKPDO}(\psi)$: On input $\psi \in \mathbb{G}_2$, \mathcal{C} picks $\beta \in \mathbb{Z}_p^*$ and sets $pk_{PDO} = \psi^\beta = h^{b\gamma\beta}$. Then \mathcal{C} sends β, pk_{PDO} to \mathcal{A} .
- $\mathcal{O}_{SKPDO}(ID_{PDO}, \beta)$: On input $ID_{PDO} \in \{0, 1\}^*$ and $\beta \in \mathbb{Z}_p^*$, \mathcal{C} first searches table D_1 to retrieve γ . If there does not exist a record (ID_{PDO}, γ) in D_1 , it return \perp . Otherwise, \mathcal{C} sets $sk_{PDO} = \gamma \cdot \beta$. After that, it sends sk_{PDO} to \mathcal{A} .

Note that the seed generation oracle $\mathcal{O}_{GenSeed}(ID_{PDO})$, the public key generation oracle $\mathcal{O}_{PKPDO}(\psi)$, and the secret generation oracle $\mathcal{O}_{SKPDO}(\gamma, \beta)$ must be queried in a sequential order.

Challenge. \mathcal{A} submits two sets of data granules (m_0, m_1) with the same length ℓ to \mathcal{C} . \mathcal{C} first picks $P_1 \in \{0, 1\}^\ell$ and $d, y \in \mathbb{Z}_p^*$, and computes $DCI^* = h^{bd}$ and $C_1 = h^{by}$. Then, \mathcal{C} sets $P_2 = \mathcal{H}_2(\mathcal{Z}^y)$. \mathcal{C} flips a random coin $b \in \{0, 1\}$ and sets $C_2 = m_b \oplus P_1 \oplus P_2$. Parse $\mathbb{A}^* = (\mathbb{M}^*, \pi^*)$. Suppose \mathbb{M}^* has the shape $(n_1 \times n_2)$. \mathcal{C} picks $\mathbf{v} \in \mathbb{Z}_p^{n_2-1}, \mathbf{y}' \in \mathbb{Z}_p^r$. For $j \in [\tau]$, it computes $C_{3,j} = h^{by'[j]}$. For $i \in [n_1]$, it computes $C_{4,i} = \mathcal{H}_1(|\mathcal{U}| + 1)^{\mathbb{M}_{i \cdot (y||\mathbf{v})}^\top} \cdot \mathcal{H}_1(\pi(i))^{y'[\rho(i)]}$. Next, it computes $\delta = \mathcal{H}_3(DCI^*, C_1, C_2, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]})$ and $V = g^{a\delta \sum_{k=1}^c d_k}$. Then \mathcal{C} sends DCI^* and $DC^* = (\mathbb{A}^*, C_1, C_2, \{C_{3,j}\}_{j \in [\tau]}, \{C_{4,i}\}_{i \in [n_1]}, V)$ to \mathcal{A} .

Phase 2. \mathcal{A} continues to query \mathcal{C} under the previous restrictions.

Guess. \mathcal{A} outputs a guess b' . It returns 1 implying $\mathcal{Z} = e(g, h)^{abc}$ if $b' = b$. Otherwise, it returns 0 implying $\mathcal{Z} = \mathcal{R}$.

If $\mathcal{Z} = e(g, h)^{abc}$, then $sk_{PDO}^* = c$, where sk_{PDO}^* is the secret key that used in the encryption, the challenge ciphertext queried by \mathcal{A} originates from a distribution that is the same as in the construction. If $\mathcal{Z} = \mathcal{R}$, and since \mathcal{R} is a random element of \mathbb{G}_T , the challenge ciphertext $C_2 = m_b \oplus \mathcal{H}_2(\mathcal{R}^y) \oplus P_2$ is uniformly random. m_b is independent in the view of \mathcal{A} .

In this simulation, the advantage of \mathcal{A} is given by

$$\text{Adv}_{DCSS_{TD}, \mathcal{A}}^{\text{IND-CPA}}(\lambda) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \left(\frac{1}{2} + \epsilon\right) - \frac{1}{2} = \frac{\epsilon}{2}$$

VII. PERFORMANCE EVALUATION

In this section, many state-of-the-art schemes [27], [22], [23], [25] are compared with our scheme. Firstly, we analyze these several related schemes regarding properties and functionality. Then, we give comprehensive computation and storage cost comparisons

TABLE III: Computation cost comparisons

Scheme	Setup	KGDO	KGDU	Enc	PreWork	Dec	Revoke
DYL+ [22]	$ \mathcal{U} (e_1 + e_T) + p$	0	$2 \mathcal{S} e_1$	$3n_1e_1 + (2n_1 + 1)e_T$	-	$2Ip$	$2n_1e_T$
ZSL+ [23]	$3 \mathcal{U} e_1 + e_T + p$	0	$(3 \mathcal{U} + 1)e_1$	$(\mathcal{U} + 1)e_1 + e_T$	CS: $(\mathcal{U} + 1)e_1 + e_T$	$(\mathcal{U} + 1)p$	$3 \mathcal{U} e_1$
NHS+ [25]	$e_1 + e_T + p$	0	$(\mathcal{S} + 2)e_1$	$(3n_1 + 1)e_1 + e_T$	DU: $(\mathcal{S} + 2)e_1$ Other: $e_1 + (2I + 1)p$ S: $(\mathcal{S} + 3)e_1 + (2I + 1)p$	$(2I + 1)p$	-
YSX+ [27]	$4e_1 + e_T + p$	0	$(\mathcal{S} + 7)e_1$	$(3n_1 + 5)e_1 + 2e_T + p$	DU: $(\mathcal{S} + 12)e_1$	CS: $(2I + 7)p$ DU: $e_1 + 2e_T$ S: $e_1 + e_T + (2I + 7)p$	-
TD-DCSS	e_2	DO: e_2 TA: e_2 S: $2e_2$	$(\mathcal{S} + 3)e_1 + e_2$	$(2n_1 + 2)e_1 + (\tau + 2)e_2 + p$	DO: $(n + 1)e_1 + (n + 2)p$	$e_1 + (\tau + 4)p$	DO: $e_1 + e_2$ CS: e_1 S: $2e_1 + e_2$

“**KGDO**”/“**KGDU**” denote key generation for a data owner/data user; “**PreWork**” denotes the preparatory work for data sharing.

TABLE IV: Storage cost comparisons

Scheme	MPK	KeyDO	KeyDU	CT	PreWork Cost
DYL+ [22]	$ \mathcal{U} (\mathbb{G}_1 + \mathbb{G}_T)$	0	$ \mathcal{S} \mathbb{G}_1 $	$ \mathbb{A} + n_1(\mathbb{G}_1 + \mathbb{G}_T)$	-
ZSL+ [23]	$ \mathcal{U} (\mathbb{G}_1 + \mathbb{G}_T)$	0	$(3 \mathcal{U} + 1) \mathbb{G}_1 $	$ \mathbb{A} + \mathbb{G}_T + (\mathcal{U} + 1) \mathbb{G}_1 $	-
NHS+ [25]	$ \mathbb{G}_1 + \mathbb{G}_T $	0	$(\mathcal{S} + 2) \mathbb{G}_1 $	$ \mathbb{A} + \mathbb{G}_T + (3n_1 + 1) \mathbb{G}_1 $	$(\mathcal{S} + 2) \mathbb{G}_1 $
YSX+ [27]	$4 \mathbb{G}_1 + \mathbb{G}_T $	0	$2 \mathbb{Z}_p + (\mathcal{S} + 4) \mathbb{G}_1 $	$ \mathbb{A} + 3\ell + (2n_1 + 5) \mathbb{G}_1 $	$(\mathcal{S} + 8) \mathbb{G}_1 + \mathbb{Z}_p $
TD-DCSS	$ \mathbb{G}_2 $	$ \mathbb{Z}_p $	$(\mathcal{S} + 2) \mathbb{G}_1 + \mathbb{G}_2 $	$ \mathbb{A} + \ell + (n_1 + 2) \mathbb{G}_1 + (\tau + 1) \mathbb{G}_2 $	$n\ell + \mathbb{G}_1 + (n + 1) \mathbb{G}_T $

“**MPK**” and “**CT**” denote “Master Public Key” and “Ciphertext”, respectively; “**KeyDO**”/“**KeyDU**” denote the storage consumption of the key of a data owner/a data user, respectively; “**PreWork Cost**” denote the storage cost of the preparatory work for data sharing.

and an experimental simulation for our scheme to demonstrate its practicality.

A. Property & Functionality

As shown in TABLE II, several ABE-based data sharing schemes are compared with TD-DCSS of properties and functionality. We use $\checkmark(\mathbf{X})$ to denote that the scheme achieves (not achieves) this property or functionality. “N/A” means this property or functionality is not applicable in this scheme. For simplicity, the CR, TR, and ER represent collusion resistance, tamper resistance, and EDoS resistance, respectively.

Many data sharing schemes based on ABE [22], [23], [25], [27] have been proposed to realize fine-grained access control and privacy protection in data sharing. DYL+ [22] applies ABE and IBE to design a data sharing system that supports permission revocation. Furthermore, ZSL+ [23] introduces a two-factor mechanism to protect the security of users’ secret keys. It can revoke secret keys efficiently based on the proxy re-encryption [43] and key separation techniques. Moreover, when considering cloud-hosted data, the economics of data sharing is important. To resist EDoS attacks, a dual access control is proposed by NHS+ [25]. It achieves efficient access control over the download requests and protects sensitive data. To realize selective sharing and informed-consent based authorization, YSX+ [27] introduces a selective data sharing system based on ABE and SE. However, it neither realizes permission revocation nor resists collusion or EDoS attacks. Our TD-DCSS realizes the personal data sharing autonomy, including selective sharing, informed-consent based authorization, and permission revocation, with collusion, tamper, and EDoS resistance. Furthermore, since symmetric groups have serious security issues [28], our proposed TD-DCSS scheme is based on the asymmetric prime-order groups which support efficient hashing to \mathbb{G}_1 [29], [44].

In conclusion, our proposed TD-DCSS scheme has desirable properties and functionality superior to the state-of-the-art solutions. We give theoretical analysis via computation and storage cost comparisons to evaluate the efficiency of our scheme in the following.

B. Computation & Storage Cost

As shown in TABLE III and TABLE IV, we analyze those state-of-the-art solutions in terms of computation and storage overheads. In our comparisons, we are mainly considering the most time-consuming operations such as exponentiation and bilinear pairings. To show the results intuitively, we let n, n_1 be the number of the data granules that a PDO wants to share with an SP in the TaskIssue algorithm and the number of the rows of policy matrix \mathbb{M} , respectively. Let $|\mathcal{U}|, |\mathcal{S}|, I$ denote the number of the universe of attributes, the number of the attribute set of the data owner, and the number of attributes used in the decryption of the ABE scheme. In TABLE III, we let e_1, e_2, e_T, p be the overhead of a single exponentiation computation in \mathbb{G}_1 , a single exponentiation computation in \mathbb{G}_2 , a single exponentiation computation in \mathbb{G}_T and a pairing operation, respectively. In TABLE IV, we let $|\mathbb{Z}_p|, |\mathbb{G}_1|, |\mathbb{G}_2|, |\mathbb{G}_T|$ be the size of a single element in $\mathbb{Z}_p, \mathbb{G}_1, \mathbb{G}_2$ and \mathbb{G}_T , respectively. Let $|\ell|, |\mathbb{A}|$ denote the length of a single data granule and the size of an access policy, respectively.

As shown in TABLE III, we analyze the several phases: **Setup**, **KGDO**, **KGDU**, **Enc**, **PreWork**, **Dec**, **Revoke**. In detail, **KGDO**, **KGDU** means generating a secret key for data owners and generating a secret key for data users, respectively. The phase **PreWork** means the time cost of the preparatory work for data sharing, such as TaskIssue algorithm in TD-DCSS, data download in ZSL+ [23], NHS+ [25] and keyword test in YSX+ [27]. To better demonstrate the computation overhead borne by different sides at the same stage, we let the prefix indicate that the time cost is borne by this party. Especially, the prefix “S:” denotes the sum of all sides.

We can easily conclude that the computation cost of the **Setup** in NHS+ [25], YSX+ [27] and TD-DCSS is constant, while the cost of DYL+ [22] and ZSL+ [23] is growing with the $|\mathcal{U}|$ linearly. Our scheme has an additional phase **KGDO** but the cost is constant. The cost of **KGDU**, **Enc** in ZSL+ [23] follows a linear relationship with the $|\mathcal{U}|$ while other schemes follow a linear relationship with the $|\mathcal{S}|$ (in **KGDU**) or n_1 (in **Enc**) ($|\mathcal{S}| \leq |\mathcal{U}|$). Note that $\tau \leq I$ and we consider $\tau = 1$. We can observe that the cost of **PreWork** in TD-DCSS is growing with n . The computation overhead of **PreWork** in TD-DCSS is constant if $n = 1$, while other schemes follow a linear relationship with $|\mathcal{U}|$ or $|\mathcal{S}|$. We can also demonstrate that the cost of

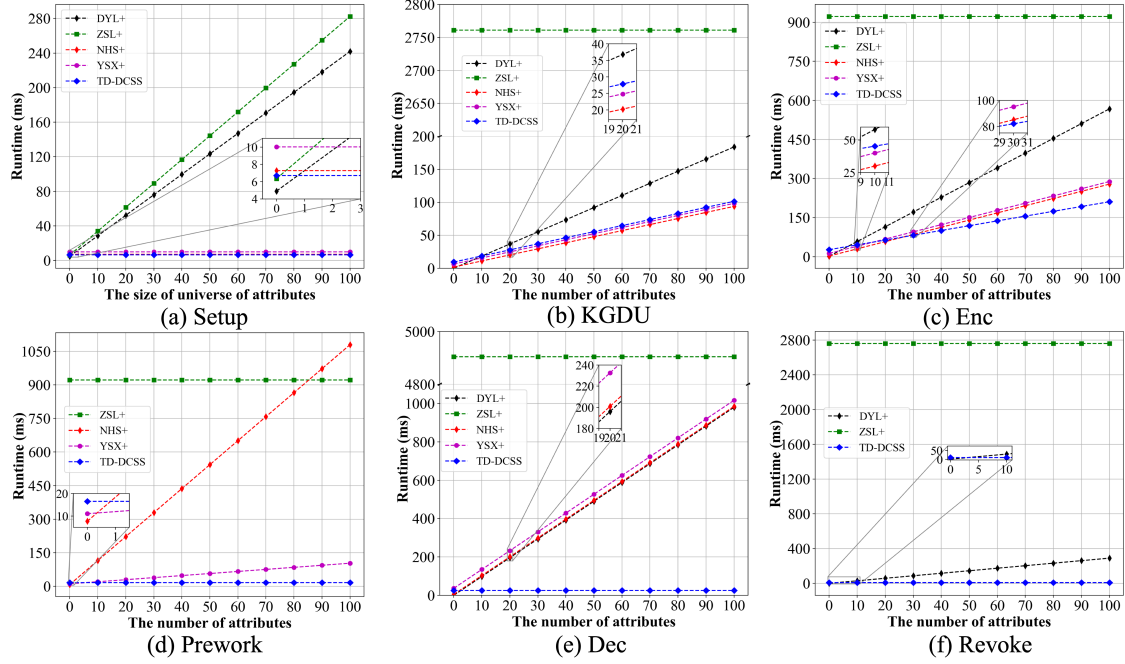


Fig. 6: Runtime evaluation of Setup, KGDU, Enc, PreWork, Dec and Revoke phases.

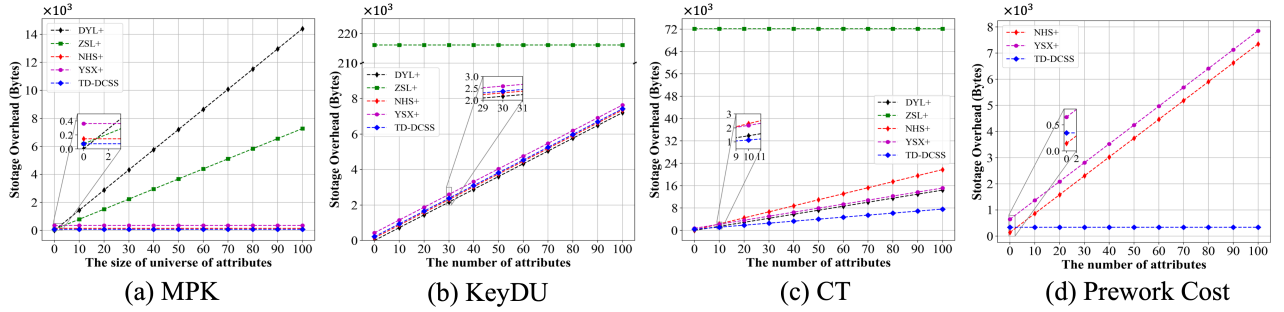


Fig. 7: Storage overhead of MPK, KeyDU, CT and PreWork Cost.

TD-DCSS is constant in **Dec** since we consider $\tau = 1$, while other schemes are growing with I or $|\mathcal{U}|$, where the cost at data owner side is constant in YSX+ [27]. The computation cost of **Revoke** in TD-DCSS is constant, while in DYL+ [22] and ZSL+ [23] is linear.

As depicted in TABLE IV, we mainly consider the most storage-consuming parts, such as master public key (**MPK**), secret key of data owner (**KeyDO**), secret key of data user (**KeyDU**), ciphertext (**CT**) and preparatory work cost (**PreWork**). We can find that the storage cost of **MPK** in NHS+ [25], YSX+ [27] and TD-DCSS is constant while it in DYL+ [22] and ZSL+ [23] is growing with $|\mathcal{U}|$. The storage cost of **KeyDO** in TD-DCSS is constant. It is easy to observe that the storage cost of **KeyDU**, **CT** in all schemes except for ZSL+ [23], has a linear relationship with $|\mathcal{S}|$ (**KeyDU**) and n_1 (**CT**), while ZSL+ [23] has a linear relationship with $|\mathcal{U}|$ of **KeyDU** and **CT**. When $n = 1$, the overhead of **PreWork** in TD-DCSS is constant while it in [25] and [27] is growing with $|\mathcal{S}|$.

C. Experimental Analysis

In order to demonstrate the practicality of our proposed scheme and the effectiveness of the task mechanism in protecting data privacy, we conducted an experimental simulation for our TD-DCSS scheme and other schemes, assessing their performance across various conditions. The experiments were conducted on a standard

personal computer running Windows 10, equipped with an Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz, and 16GB of RAM. Our code, implemented in Python 3.6.9, relies on the Charm 0.50 library [45], PBC-0.5.14 library, and operates within the Ubuntu 18.04.7 LTS environment (WSL). Our implementation based on the project available at <https://github.com/DoreenRiepel/FABEO>, adopting $\ell = 128$ and utilizing the MNT224 curve. The source code for our TD-DCSS, along with the corresponding evaluation code, has been made publicly accessible at <https://github.com/xiahezzz/TD-DCSS>. Since the fact that $\mathcal{S} \subseteq \mathcal{U}$ and $|\mathcal{S}| \ll |\mathcal{U}|$, we let $|\mathcal{U}|$ be equal to 1000. The outcomes of our experimental analysis are illustrated in Fig. 6 and Fig. 7.

Fig. 6 demonstrates the overhead of the **Setup**, **KGDU**, **Enc**, **Prework**, **Dec** and **Revoke** phases. As shown in Fig. 6a, it is evident that the runtime of **Setup** in NHS+ [25] and YSX+ [27]. In Fig. 6b and Fig. 6c, the runtime of **KGDU** and **Enc** of our TD-DCSS grows as the number of attributes in \mathcal{S} and \mathbb{A} grows, respectively. Moreover, the runtime of **Enc** in our TD-DCSS has a relatively constant and lower computational overhead compared to other schemes. Meanwhile, the runtime of the **KGDU** in our TD-DCSS consumes only a slightly more constant time than NHS+ [25], YSX+ [27]. Fig. 6d shows the relationship between the runtime of the **PreWork** and the number of attributes in \mathcal{S} (or the number of data granules in the task). For better comparisons, we set $|\mathcal{Z}| = |\mathcal{S}|$, and

$n = 1$ in TD-DCSS, as other schemes share one data granule. The results show that when $n = 1$, the runtime of the **PreWork** in our TD-DCSS remains constant, while in other schemes, it is linear. As shown in Fig. 6e and Fig. 6f, the runtime of the **Dec** and **Revoke** phases in our TD-DCSS is constant, taking relatively less computation overhead compared to other schemes. The efficient execution of the UpdateDC algorithm can be attributed to the parameters generated by the PDO, allowing the CS to readily update the data capsule without requiring an update to the ABE-related ciphertext.

In our simulation, the runtimes for the GenSeed, PKeyGenPDO, and SKeyGenPDO algorithms in our TD-DCSS are 7ms, 3.7e-04ms, and 6ms, respectively.

Fig. 7 depicts the storage overhead for **MPK**, **KeyDU**, **CT** and **PreWork Cost**. From Fig. 7a, we can find that the storage overhead of **MPK** in NHS+ [25], YSX+ [27], and our TD-DCSS is constant, while in DYL+ [22] and ZSL+ [23], it is linear with the $|U|$. As shown in Fig. 7b, the storage overhead of **KeyDU** in all works except ZSL+ [23] is linear with respect to \mathcal{S} . From Fig. 7c, it is easy to find that the storage overhead of **CT** in all works except for ZSL+ [23] is linear with the number of attributes in \mathbb{A} , and the **CT** in our TD-DCSS consumes less storage overhead. In Fig. 7d, we set $n = 1$, as other schemes also share one data granule, and then we can find that the storage overhead of **PreWork Cost** in our TD-DCSS remains constant, while in other schemes, it is linear with the $|\mathcal{S}|$.

These results demonstrate that the TD-DCSS scheme realizes personal data sharing for protecting owners' privacy and enabling fine-grained data sharing, with acceptable runtime and storage consumption.

VIII. CONCLUSION AND FUTURE WORK

In this paper, we first introduce the concept of personal data sharing autonomy, including selective sharing, informed-consent based authorization and permission revocation, focusing on data owners' privacy. To realize the personal data sharing autonomy, we then propose the task-driven data capsule data sharing system (TD-DCSS), which mainly includes the data capsule encapsulation method and the task-driven data sharing mechanism. Finally, a comprehensive security analysis, with a theoretical analysis and an experimental simulation are conducted. The result reveals that our scheme is correct, sound, and secure under the security model, and is proved to be an effective scheme compared with state-of-the-art schemes.

In our scheme, a task can be used to access the data capsule only once, which enables a data owner has the right of informed-consent based authorization and permission revocation, precisely. However, this may lead to inefficient data sharing in the case that the same data need to be accessed frequently. Therefore, a possible future direction could involve designing a new mechanism that allows one task to be used multiple times while keeping the informed-consent based authorization and permission revocation.

IX. ACKNOWLEDGEMENTS

We would like to thank Professor Jian Weng, Jinan University, China, for generously providing valuable comments and constructive feedback that substantially enriched the content and quality of this paper.

REFERENCES

- [1] W. E. Forum, "Rethinking personal data: Trust and context in user-centred data ecosystems." World Economic Forum Geneva, Switzerland, 2014.
- [2] R. Li, Z. Wang, L. Fang, C. Peng, W. Wang, and H. Xiong, "Efficient blockchain-assisted distributed identity-based signature scheme for integrating consumer electronics in metaverse," *IEEE Transactions on Consumer Electronics*, 2024.
- [3] C. Yan, C. Wang, J. Shen, K. Dev, M. Guizani, and W. Wang, "Edge-assisted hierarchical batch authentication scheme for vanets," *IEEE Transactions on Vehicular Technology*, 2023.
- [4] W. Wang, Z. Han, T. R. Gadekallu, S. Raza, J. Tanveer, and C. Su, "Lightweight blockchain-enhanced mutual authentication protocol for uavs," *IEEE Internet of Things Journal*, 2023.
- [5] K. U. Fallatah, M. Barhamgi, and C. Perera, "Personal data stores (pds): a review," *Sensors*, vol. 23, no. 3, p. 1477, 2023.
- [6] C. Mydex, "The case for personal information empowerment: The rise of the personal data store," *World*, pp. 1–44, 2010.
- [7] M. Van Kleek, D. A. Smith, N. Shadbolt *et al.*, "A decentralized architecture for consolidating personal information ecosystems: The webbox," 2012.
- [8] H. P. R. Team *et al.*, "Hat briefing paper 2: The hub-of-all-things (hat) economic model of the multisided market platform and ecosystem. wmg service systems research group working paper series (number 02/15), 2015," 2015.
- [9] M. Fernández, J. Jaimunk, and B. Thuraisingham, "A privacy-preserving architecture and data-sharing model for cloud-iot applications," *IEEE Transactions on Dependable and Secure Computing*, 2022.
- [10] Techcircle. (2023) Cloud attacks rise but most sensitive data remains unencrypted. [Online]. Available: <https://www.techcircle.in/2023/07/05/cloud-attacks-rise-but-most-sensitive-data-remains-unencrypted>
- [11] L. Wang, J. P. Near, N. Somani, P. Gao, A. Low, D. Dao, and D. Song, "Data capsule: A new paradigm for automatic compliance with data privacy regulations," in *Heterogeneous Data Management, Polystores, and Analytics for Healthcare: VLDB 2019 Workshops, Poly and DMAH, Los Angeles, CA, USA, August 30, 2019, Revised Selected Papers 5*. Springer, 2019, pp. 3–23.
- [12] G. Xu, S. Xu, J. Ma, J. Ning, and X. Huang, "An adaptively secure and efficient data sharing system for dynamic user groups in cloud," *IEEE Transactions on Information Forensics and Security*, 2023.
- [13] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (gdpr)," *A Practical Guide, 1st Ed.*, Cham: Springer International Publishing, vol. 10, no. 3152676, pp. 10–5555, 2017.
- [14] R. B. Ness, J. P. Committee *et al.*, "Influence of the hipaa privacy rule on health research," *Jama*, vol. 298, no. 18, pp. 2164–2170, 2007.
- [15] E. Goldman, "An introduction to the california consumer privacy act (ccpa)," *Santa Clara Univ. Legal Studies Research Paper*, 2020.
- [16] A. Tsesis, "The right to erasure: Privacy, data brokers, and the indefinite retention of data," *Wake Forest L. Rev.*, vol. 49, p. 433, 2014.
- [17] A. M. Abdullah *et al.*, "Advanced encryption standard (aes) algorithm to encrypt and decrypt data," *Cryptography and Network Security*, vol. 16, no. 1, p. 11, 2017.
- [18] R. Canetti, S. Halevi, and J. Katz, "A forward-secure public-key encryption scheme," in *Advances in Cryptology—EUROCRYPT 2003: International Conference on the Theory and Applications of Cryptographic Techniques, Warsaw, Poland, May 4–8, 2003 Proceedings 22*. Springer, 2003, pp. 255–271.
- [19] R. S. Sandhu, "Role-based access control," in *Advances in computers*. Elsevier, 1998, vol. 46, pp. 237–286.
- [20] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in *Annual international cryptology conference*. Springer, 2001, pp. 213–229.
- [21] X. Liang, Z. Cao, H. Lin, and J. Shao, "Attribute based proxy re-encryption with delegating capabilities," in *Proceedings of the 4th international symposium on information, computer, and communications security*, 2009, pp. 276–286.
- [22] X. Dong, J. Yu, Y. Luo, Y. Chen, G. Xue, and M. Li, "Achieving an effective, scalable and privacy-preserving data sharing service in cloud computing," *Computers & security*, vol. 42, pp. 151–164, 2014.
- [23] C. Zuo, J. Shao, J. K. Liu, G. Wei, and Y. Ling, "Fine-grained two-factor protection mechanism for data sharing in cloud storage," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 1, pp. 186–196, 2017.
- [24] Z. Song, H. Ma, R. Zhang, W. Xu, and J. Li, "Everything under control: Secure data sharing mechanism for cloud-edge computing," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2234–2249, 2023.
- [25] J. Ning, X. Huang, W. Susilo, K. Liang, X. Liu, and Y. Zhang, "Dual access control for cloud-based data storage and sharing," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 2, pp. 1036–1048, 2020.
- [26] S. Xu, J. Ning, Y. Li, Y. Zhang, G. Xu, X. Huang, and R. H. Deng, "A secure enr sharing system with tamper resistance and expressive access control," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 1, pp. 53–67, 2023.
- [27] K. Yang, J. Shu, and R. Xie, "Efficient and provably secure data selective sharing and acquisition in cloud-based systems," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 71–84, 2023.

- [28] S. Agrawal and M. Chase, "Fame: Fast attribute-based message encryption," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 665–682. [Online]. Available: <https://doi.org/10.1145/3133956.3134014>
- [29] D. Riepel and H. Wee, "Fabeo: Fast attribute-based encryption with optimal security," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, 2022, pp. 2491–2504.
- [30] Y. Zhang, R. H. Deng, S. Xu, J. Sun, Q. Li, and D. Zheng, "Attribute-based encryption for cloud computing access control: A survey," *ACM Computing Surveys (CSUR)*, vol. 53, no. 4, pp. 1–41, 2020.
- [31] C. Cocks, "An identity based encryption scheme based on quadratic residues," in *Cryptography and Coding: 8th IMA International Conference Cirencester, UK, December 17–19, 2001 Proceedings 8*. Springer, 2001, pp. 360–363.
- [32] M. Green and G. Ateniese, "Identity-based proxy re-encryption," in *Applied Cryptography and Network Security: 5th International Conference, ACNS 2007, Zhuhai, China, June 5-8, 2007. Proceedings 5*. Springer, 2007, pp. 288–306.
- [33] J. Sun, G. Xu, T. Zhang, X. Yang, M. Alazab, and R. H. Deng, "Verifiable, fair and privacy-preserving broadcast authorization for flexible data sharing in clouds," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 683–698, 2022.
- [34] J. Zhang, S. Su, H. Zhong, J. Cui, and D. He, "Identity-based broadcast proxy re-encryption for flexible data sharing in vanets," *IEEE Transactions on Information Forensics and Security*, 2023.
- [35] H. Li, D. Liu, Y. Dai, T. H. Luan, and X. S. Shen, "Enabling efficient multi-keyword ranked search over encrypted mobile cloud data through blind storage," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 127–138, 2014.
- [36] M. H. Sqalli, F. Al-Haidari, and K. Salah, "Edos-shield-a two-steps mitigation technique against edos attacks in cloud computing," in *2011 Fourth IEEE international conference on utility and cloud computing*. IEEE, 2011, pp. 49–56.
- [37] G. Somani, M. S. Gaur, and D. Sanghi, "Ddos/edos attack in cloud: affecting everyone out there!" in *Proceedings of the 8th International Conference on Security of Information and Networks*, 2015, pp. 169–176.
- [38] M. Gao, H. Dang, and E.-C. Chang, "Teekap: Self-expiring data capsule using trusted execution environment," in *Annual Computer Security Applications Conference*, 2021, pp. 235–247.
- [39] N. Attrapadung and H. Imai, "Attribute-based encryption supporting direct/indirect revocation modes," in *IMA international conference on cryptography and coding*. Springer, 2009, pp. 278–300.
- [40] J. Ning, Z. Cao, X. Dong, J. Gong, and J. Chen, "Traceable cp-abe with short ciphertexts: How to catch people selling decryption devices on ebay efficiently," in *Computer Security – ESORICS 2016*. Cham: Springer International Publishing, 2016, pp. 551–569.
- [41] Y. Zhang, D. Zheng, and R. H. Deng, "Security and privacy in smart health: Efficient policy-hiding attribute-based access control," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2130–2145, 2018.
- [42] A. Lewko and B. Waters, "Unbounded hibe and attribute-based encryption," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2011, pp. 547–567.
- [43] M. Blaze, G. Bleumer, and M. Strauss, "Divertible protocols and atomic proxy cryptography," in *International conference on the theory and applications of cryptographic techniques*. Springer, 1998, pp. 127–144.
- [44] A. Faz-Hernandez, S. Scott, N. Sullivan, R. S. Wahby, and C. A. Wood, "Hashing to Elliptic Curves," RFC 9380, Aug. 2023. [Online]. Available: <https://www.rfc-editor.org/info/rfc9380>
- [45] 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, pp. 111–128, 2013.



Qiuyun Lyu received the Ph.D. degree from the School of Computer Science and Technology, Hangzhou Dianzi University, China, in 2021. She is currently an associate professor with the School of Cyberspace, Hangzhou Dianzi University. She also works in Key Laboratory of Data Storage and Transmission Technology of Zhejiang Province and Pinghu Digital Technology Innovation Institute Co., Ltd, Hangzhou Dianzi University. Her research interests include data security, self-sovereign identity, and privacy-enhancing technology.



Yilong Zhou received his B.S. degree from the School of Cyberspace, Hangzhou Dianzi University, China, in 2022. He is currently working toward an M.S. degree in Cybersecurity with the School of Cyberspace, Hangzhou Dianzi University. He also works in Pinghu Digital Technology Innovation Institute Co., Ltd, Hangzhou Dianzi University. His research focuses on security and privacy in data sharing, and access control.



Yizhi Ren received his PhD in Computer software and theory from Dalian University of Technology, China in 2011. He is currently an professor with School of Cyberspace, Hangzhou Dianzi University, China. From 2008 to 2010, he was a research fellow at Kyushu University, Japan. His current research interests include: network security, complex network, and trust management. Dr. REN has published over 60 research papers in refereed journals and conferences. He won IEEE Trustcom 2018 Best Paper Award, CSS2009 Student Paper Award and

AINA2011 Best Student paper Award.



Zheng Wang received the Ph.D. degree in Software Engineering from Dalian University of Technology, China, in 2016. Now He is an associate professor with School of Cyberspace, and vice dean of ZhuoYue Honors College, Hangzhou Dianzi University, China. His research interests include complex networks, network security, and artificial intelligence.



Yunchuan Guo (M'14) received the B.S. and M.S. degrees in computing science and technology, Guilin, China, in 2000 and 2003, respectively, and the Ph.D. degree in computing science and technology from the Institute of Computing Technology, Chinese Academy of Science, Beijing, China, in 2011. He is currently a Professor with the Institute of Information Engineering, Chinese Academy of Sciences. His current research interests include network security and access control.