The 48th Symposium on Information Theory
and its Applications (SITA2025)
Koriyama, Fukushima, Japan, Nov. 25–28, 2025

On Correctness and Privacy in Private Information Delivery with Coded Storage[*]

Jun Kurihara[†‡] Koki Nakano[†] Toshiaki Tanaka[†]

Abstract— *Private information delivery* (PID) with coded storage enables push-based transmission from coded storage servers to a receiver without revealing which message is transmitted. Prior work on PID with coded storage assumed *bi-regular* data placement, limiting practical deployment. This paper reformulates the PID framework by focusing on *location privacy* of storage servers rather than message-index privacy, thereby allowing flexible placement. Within this generalized setting, we formalize *correctness* (exact message reconstruction) and *privacy* (concealing which servers store the delivered message). We further analyze schemes based on arbitrary linear codes, not restricted to MDS codes, and demonstrate that the conditions for both properties are characterized in terms of the minimum Hamming weight of its dual code.

Keywords— Private information delivery, anonymity, privacy, coded storage

1 Introduction

In the context of information and coding theory, numerous techniques ensuring security and privacy have been proposed. For *security*—that is, guaranteeing the confidentiality and integrity of sensitive information—*the wiretap channel II* [8] and *secret sharing schemes* [1,9] are classical and extensively studied examples. For *privacy*, various approaches have been explored, including *private information retrieval* (PIR) [2,11], which enables a user to obtain data from one or more servers without revealing which specific data is being requested in pull-based data transmission.

From the privacy perspective, *private information delivery* (PID) [7,10,12] was recently introduced as a method for *push-based* data transmission to remain the identifier of the delivered message confidential from the receiver. Specifically, we assume that a sender holds a collection of messages, each labeled by a unique identifier (index), and wishes to deliver one message without disclosing its index to the receiver. This problem is motivated by scenarios in which each index uniquely corresponds to an entity's identity—for instance, a patient identifier in a medical database—while the sender aims to deliver an associated record without revealing that identity. In the PID setting, multiple servers participate in delivery, and a subset of servers stores fragments (replicas or encoded symbols) of each message. We assume a fixed injective mapping that assigns each message index to a unique subset of servers storing its fragments, and that this mapping is known to the receiver; thus, if the receiver determines which servers hold the delivered message, the index is revealed, and vice versa. During delivery, all servers transmit in parallel either dummy data or randomized encodings of the desired message fragments, enabling the receiver to uniquely reconstruct the message by collecting all transmissions while learning no information about the message's index. We emphasize that this index privacy in PID is equivalent to the *location privacy* (or *location anonymity*) of the storage servers against the receiver, since the index-to-subset mapping is fixed and injective. For example, if a subset of hospitals is a specific patient's primary care provider storing the patient's medical record, the record can be delivered without revealing which hospitals—and hence their locations—are involved. Figure 1 illustrates this scenario from the viewpoint of location privacy.

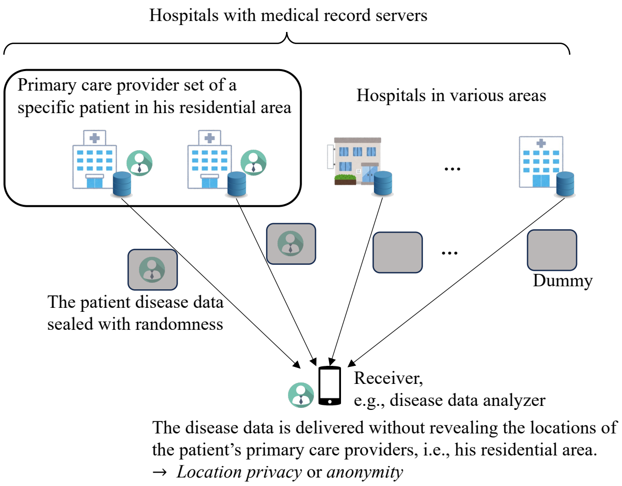
Hospitals with medical record servers
Primary care provider set of a specific patient in his residential area
Hospitals in various areas
...
...
Dummy
The patient disease data sealed with randomness
Receiver,
e.g., disease data analyzer
The disease data is delivered without revealing the locations of the patient's primary care providers, i.e., his residential area.
→ *Location privacy* or *anonymity*

Figure 1: Example of PID from the viewpoint of location privacy: a patient's disease data (with no identifying information) is delivered without revealing which hospitals store the record.

The above PID problem was first introduced by Sun et al. [10] with replicated, i.e., non-coded, storage storing copies of the messages. Subsequently, Vaidya et al. [12] extended the PID setting by allowing *coded* storage servers storing encoded message symbols to increase the transmission rate of the scheme, which is defined as the ratio of the size of the desired message to that of total communication. Vaidya et al. considered a special case of the data-placement structure on servers, referred to as the *bi-regular PID* setting, which imposes a regular and systematic placement pattern of encoded messages across servers. Furthermore, they presented an explicit PID scheme attaining the maximum transmission rate achievable under the setting of bi-regular PID by using a *maximum distance separable* (MDS) code for the coded storage.

For the PID with coded storage, here we point out the following problems:

(1) When anonymized message transmission originates

[*] The material of this paper will be partially published in IEICE Trans. Fundamentals [7]. A part of this work was done during the second author's master course program at University of Hyogo in 2023–2025.

[†] Graduate School of Information Science, University of Hyogo, 7-1-28 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. E-mail: `kurihara@ieee.org`

[‡] Zettant Inc. 1-8-1-8F BcH Nihonbashi-Kayabacho, Chuo-ku, Tokyo 103-0025, Japan.

from specific subsets of servers, biases and heterogeneity in data across servers should be taken into account. However, only a specific data-placement structure with regular patterns, i.e., the bi-regular PID, has been considered.

(2) The length of an MDS code is conjectured to be bounded by the size of the finite field [6], which consequently limits the maximum possible number of servers in the Vaidya et al.'s scheme. This restriction can cause inefficiencies and pose computational challenges, particularly for large-scale systems; hence PID schemes based on arbitrary linear codes should be explored. However, a characterization of schemes employing arbitrary linear codes remains open in the context of PID with coded storage.

For the problems posed above, contributions of this paper are summarized as follows:

- To address problem (1), we generalize the setting of Vaidya et al. and reformulate PID with coded storage without the bi-regular constraint, allowing flexible and more general data placement. In our reformulation, we focus on ensuring the privacy of server locations rather than the index privacy of messages.
- To address problem (2), we introduce two properties for PID schemes based on arbitrary linear codes in the reformulated setting: (a) *correctness*, whereby the receiver can accurately reconstruct the original message from the transmitted encoded symbols; and (b) *privacy*, whereby the receiver obtains no information about the subset of servers storing the delivered message. We then explicitly clarify that the conditions for attaining correctness and privacy are characterized in terms of the *minimum Hamming weight* of the dual of the linear code employed in the PID scheme.

By the above contributions, we can see that this paper gives a design principle of PID schemes suitable for various types of systems.

The remainder of this paper is organized as follows: Section 2 introduces the necessary notations and definitions used throughout the paper. Section 3 presents the reformulation of PID with coded storage, generalizing the existing framework to accommodate a wider range of data placement scenarios. Section 4 defines the correctness and privacy of PID with coded storage, and Section 5 characterizes the conditions for achieving these properties in the reformulated setting. Section 6 finally concludes the paper. Additionally, Appendices A and B respectively provide a detailed example of the PID scheme and discuss message secrecy beside the main scope of this paper.

2 Notations and Definitions

For random variables x and y, let $H(x)$ be Shannon's entropy of x, $H(x|y)$ be the conditional entropy of x given y, and $I(x;y)$ represent the mutual information between x and y [3]. Let $[n] \triangleq \{1,\dots,n\}$. Let $\mathbb{F}$ denote a finite field; $\mathbb{F}^n$ denotes the n-dimensional vector space over $\mathbb{F}$. Let $|\mathcal{X}|$ represent the cardinality of a set $\mathcal{X}$. The set difference of sets $\mathcal{X}$ and $\mathcal{Y}$ is denoted as $\mathcal{X} \setminus \mathcal{Y} \triangleq \{x \in \mathcal{X} : x \notin \mathcal{Y}\}$.

The *Hamming distance* between two vectors is defined as the number of positions at which the corresponding symbols are different, namely for $\vec{x} = [x_1,\dots,x_n] \in \mathbb{F}^n$ and $\vec{y} = [y_1,\dots,y_n] \in \mathbb{F}^n$, it is represented by

$$d(\vec{x},\vec{y}) \triangleq |\{i : x_i \neq y_i\}|.$$

For a linear subspace $C \subseteq \mathbb{F}^n$, the *minimum Hamming distance* or *minimum Hamming weight* of C is defined by

$$\begin{aligned} d(C) &\triangleq \min\{d(\vec{x},\vec{y}) : \vec{x},\vec{y} \in C, \vec{x} \neq \vec{y}\} \\ &= \min\left\{d(\vec{x},\vec{0}) : \vec{x} \in C \setminus \{\vec{0}\}\right\}. \end{aligned}$$

A linear code $C \subseteq \mathbb{F}^n$ is a linear subspace of the vector space $\mathbb{F}^n$. For a linear code $C \subseteq \mathbb{F}^n$, the Singleton bound [6] is expressed as

$$d(C) \leq n - \dim C + 1, \tag{1}$$

where $\dim C$ denotes the dimension of the code. A code is referred to as a *maximum distance separable (MDS) code* [6] if equality holds in Eq. (1). The *dual code* of a linear code $C \subseteq \mathbb{F}^n$ is defined by

$$C^\perp \triangleq \{\vec{x} \in \mathbb{F}^n : \vec{x} \cdot \vec{y} = 0, \forall \vec{y} \in C\},$$

where $\vec{x} \cdot \vec{y}$ represents the standard inner product of $\vec{x}$ and $\vec{y}$.

For a subset $\mathcal{J} \subseteq [n]$ and a vector $\vec{c} = [c_1,\dots,c_n] \in \mathbb{F}^n$, let $P_{\mathcal{J}}(\vec{c}) \in \mathbb{F}^n$ denote an n-dimensional vector such that the j-th component of $P_{\mathcal{J}}(\vec{c})$ is c_j if $j \in \mathcal{J}$ and 0 otherwise. For example, for $\mathcal{J} = \{1,3,5\}$ and $\vec{c} = [1,1,0,1,1]$ ($n = 5$), we have $P_{\mathcal{J}}(\vec{c}) = [1,0,0,0,1]$. The *punctured code* $P_{\mathcal{J}}(C)$ of a linear code $C \subseteq \mathbb{F}^n$ for $\mathcal{J} \subseteq [n]$ is defined as

$$P_{\mathcal{J}}(C) \triangleq \{P_{\mathcal{J}}(\vec{c}) : \vec{c} \in C\}.$$

The *shortened code* of C for $\mathcal{J}$ is defined as the set of all codewords whose components are all zero outside of $\mathcal{J}$, i.e.,

$$C_{\mathcal{J}} \triangleq \{\vec{c} = [c_1,\dots,c_n] \in C : c_j = 0 \text{ for } j \notin \mathcal{J}\} \subseteq \mathbb{F}^n.$$

3 Reformulation of PID with Coded Storage

In this section, we shall generalize the setting introduced in [12] and reformulate the problem of $\mathbb{F}$-linear *private information delivery (PID)* for servers' location privacy within the framework of coded storage systems.

Let $n > 0$ denote the number of storage servers, and let $l (< n)$ be the length of the original message. We denote by $\vec{w} \in \mathbb{F}^l$ the original message. Also let $\mathcal{F}$ be a family of all possible index sets, $\mathcal{J} \subsetneq [n]$ ($|\mathcal{J}| \geq l$), of servers storing the encoded version of $\vec{w}$. We assume that a message $\vec{w}$ is chosen uniformly at random from $\mathbb{F}^l$, and an index set $\mathcal{J} \in \mathcal{F}$ is chosen according to an arbitrary distribution. Also assume that $\mathcal{J}$ and $\vec{w}$ are mutually independent.

We illustrate the flow of the PID scheme with coded storage in Figure 2 for an example case of $\mathcal{J}$. We also present a detailed example in Appendix A. In a PID scheme with coded storage, the message $\vec{w}$ is stored on servers indexed by $\mathcal{J} \in \mathcal{F}$, and then $\vec{w}$ is delivered to a receiver without disclosing the subset of server indices $\mathcal{J}$. This process

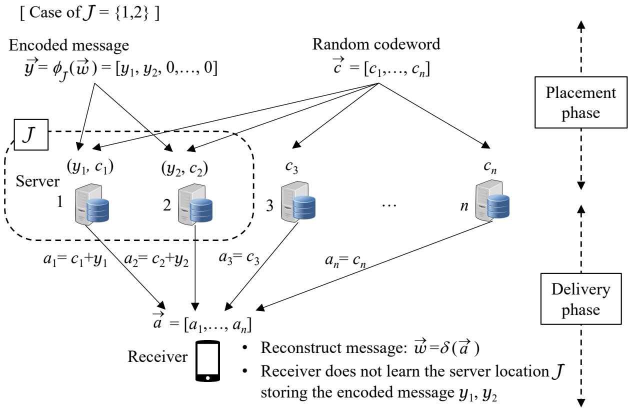
[Case of $\mathcal{J}$ = {1,2}]
Encoded message
$\vec{y} = \phi_{\mathcal{J}}(\vec{w}) = [y_1, y_2, 0, \ldots, 0]$
Random codeword
$\vec{c} = [c_1, \ldots, c_n]$
Placement phase
$\mathcal{J}$
(y_1, c_1)
(y_2, c_2)
c_3
c_n
Server
1
2
3
…
n
$a_1 = c_1 + y_1$
$a_2 = c_2 + y_2$
$a_3 = c_3$
$a_n = c_n$
Delivery phase
$\vec{a} = [a_1, \ldots, a_n]$
Receiver
• Reconstruct message: $\vec{w} = \delta(\vec{a})$
• Receiver does not learn the server location $\mathcal{J}$ storing the encoded message y_1, y_2

Figure 2: The flow of the PID scheme with coded storage for the case of $\mathcal{J} = \{1, 2\}$.

consists of the following two phases: the placement phase and the delivery phase.

Placement phase: In this phase, the message $\vec{w} \in \mathbb{F}^l$ is encoded into $\vec{y} \in \mathbb{F}^n$, satisfying $\vec{y} = P_{\mathcal{J}}(\vec{y})$, and each symbol in $\vec{y}$ is placed on the corresponding server. Specifically, the j-th server for $j \in \mathcal{J}$ stores the j-th symbol of $\vec{y}$, and nothing is placed on the j-th server for $j \notin \mathcal{J}$. To this end, we define $\phi_{\mathcal{J}} : \mathbb{F}^l \longrightarrow P_{\mathcal{J}}(\mathbb{F}^n)$ as an $\mathbb{F}$-linear injective mapping for $\mathcal{J} \in \mathcal{F}$. Then, the encoded message $\vec{y}$ is given as

$$\vec{y} \triangleq [y_1, \ldots, y_n] = \phi_{\mathcal{J}}(\vec{w}) \in P_{\mathcal{J}}(\mathbb{F}^n) \subseteq \mathbb{F}^n,$$

and each $y_j \in \mathbb{F}$ for $j \in \mathcal{J}$ is placed on the j-th server. Here we say that the j-th server for $j \in \mathcal{J}$ is associated with the message $\vec{w}$. We should note that the placement of $\vec{y}$ could be done only within the servers indexed by $\mathcal{J}$ and is not necessarily processed by a centralized entity.

In addition to the encoded message, randomness is introduced on every server. Let $C \subseteq \mathbb{F}^n$ be a linear code over $\mathbb{F}$ with $\dim C = n - l$, referred to as the *randomizing code*. A codeword $\vec{c} = [c_1, \ldots, c_n] \in C$ is chosen uniformly at random independently from $\vec{w}$ and $\mathcal{J}$, and $c_j \in \mathbb{F}$ is placed on the j-th server as randomness. Here we note that the placement of $\vec{c}$ must be done by a centralized trusted entity unlike the placement of $\vec{y}$.

Delivery phase: To deliver the message $\vec{w}$ while keeping the subset of server indices $\mathcal{J}$ confidential, the servers collectively generate $\vec{a} \triangleq [a_1, \ldots, a_n] \in \mathbb{F}^n$ based on their stored information, including $\vec{y}$ and $\vec{c}$. Specifically, the j-th server generates $a_j \in \mathbb{F}$ as follows.

$$a_j = \begin{cases} c_j + y_j, & j \in \mathcal{J}, \\ c_j, & j \notin \mathcal{J}. \end{cases}$$

Here, $y_j \in \mathbb{F}$ is the stored symbol of $\vec{y}$ on the j-th server. Each server transmits a_j to the receiver, enabling the receiver to construct $\vec{a} = \vec{c} + \vec{y} \in \mathbb{F}^n$. The receiver reconstructs $\vec{w} = \delta(\vec{a}) \in \mathbb{F}^l$ using an $\mathbb{F}$-linear surjective map $\delta : \mathbb{F}^n \longrightarrow \mathbb{F}^l$ in a *deterministic* manner, independently of $\mathcal{J}$.

In the PID scheme executed through the above process, we aim to ensure that the receiver cannot learn the subset $\mathcal{J}$ from the received symbols $\vec{a}$. This means that the locations of servers associated with $\vec{w}$ must remain confidential from the receiver, and the location privacy and the anonymity of the servers must be preserved.

Remark 1 (Relationship to the bi-regular PID setting [12])**.** Considering $K \triangleq |\mathcal{F}|$ distinct messages in the above framework, Vaidya et al. [12] introduced the *bi-regular PID* setting that imposes a strong regularity condition on the server-message associations. In the setting, each server is equally associated with the same number of messages, and has the same storage size defined as the number of its associated messages. More specifically, *every server index must be included in Kl/n sets in $\mathcal{F}$, where Kl must be divisible by n.* Under this regularity, all K messages are supposed to be placed on the servers in the placement phase in advance. Their original aim is to hide the message index $i \in [K]$ uniquely associated with a set in $\mathcal{F}$ by exploiting multiple servers for delivery, which implicitly hides the subset $\mathcal{J}$. As shown in the above, in this paper, we generalize and relax this setting by removing the restrictions on the structure of $\mathcal{F}$ and the total number of messages, and aim to directly hide the subset $\mathcal{J}$.

Remark 2 (Construction of Vaidya et al.'s scheme [12])**.** Vaidya et al. [12] employ an MDS code as the randomizing code $C \subseteq \mathbb{F}^n$ with the minimum Hamming weight $d(C) = l + 1$. They supposed that $\mathcal{F} = \mathcal{F}_{\mathrm{MDS}} \triangleq \{\mathcal{J} \subseteq [n] : |\mathcal{J}| = l\}$, meaning that each message is associated with a certain set consisting of l servers.[1] In their scheme, $\phi_{\mathcal{J}}$ is defined by the inverse of an $l \times l$ submatrix of the parity check matrix H of C, and δ is defined by H itself. Their scheme guarantees that $I(\vec{a}; \mathcal{J}) = 0$, which means the receiver cannot learn the indices of servers associated with $\vec{w}$. Also, it is guaranteed that every message $\vec{w}$ is uniquely determined from $\vec{a}$.

Provided that the message can be correctly determined by δ and $I(\vec{a}; \mathcal{J}) = 0$ holds, the performance of a PID scheme is measured by the ratio r_{PID} of the size of the message $\vec{w}$ to the total size of the transmitted symbols $\vec{a}$, where sizes are measured by the entropy function $H(\cdot)$ [12]. Namely, r_{PID} is defined as

$$r_{\mathrm{PID}} \triangleq \frac{H(\vec{w})}{\sum_j H(a_j)} = \frac{l}{\sum_j H(a_j)},$$

where $\vec{w}$ is chosen uniformly at random from $\mathbb{F}^l$ and hence $H(\vec{w}) = l$.

Remark 3. Under the bi-regular constraint, Vaidya et al. characterized the maximum achievable r_{PID}, referred to as the *capacity*, in terms of the message size l, the number of messages, and the storage size of each server. Note that in our generalized setting, we do not restrict the structure of $\mathcal{F}$, the total number of messages, or the storage size of each server. Hence we do not consider the characterization of the capacity in this paper.

[1] Note that as explained in Remark 1, Vaidya et al. have considered extra conditions on $\mathcal{F}_{\mathrm{MDS}}$ from the bi-regular PID setting.

4 Correctness and Privacy of PID with Coded Storage

In this section, we define two new information-theoretic properties, called *correctness* and *privacy*, for our generalized PID formulation, where the correctness is the condition under which a receiver can obtain the original message $\vec{w}$, and the privacy is the one that ensures the locations of servers associated with $\vec{w}$ remain hidden. For the sake of simplicity, we will henceforth identify random variables with their realizations.

Firstly, the correctness of PID with coded storage is given as follows.

Definition 1 (Correctness)**.** Consider a PID scheme with coded storage as defined in Section 3 with a linear code $C \subseteq \mathbb{F}^n$. Then, the scheme is said to achieve *correctness* if there exist mappings $\phi_{\mathcal{J}}$ and δ such that $H(\vec{w}|\vec{a}) = 0$ for every $\mathcal{J} \in \mathcal{F}$.

This definition means that the receiver can accurately reconstruct the original message $\vec{w}$ from the received vector $\vec{a}$, regardless of which subset $\mathcal{J}$ of servers stores the encoded pieces of the message.

Remark 4 (Correctness in the bi-regular PID setting)**.** The correctness in Definition 1 corresponds to Eq. (5) of [12, Section II] in the bi-regular PID setting.

Secondly, we define the privacy of PID with coded storage as follows.

Definition 2 (Privacy)**.** Consider the same setting as Definition 1. Then, the scheme is said to achieve *privacy* if $I(\vec{a}; \mathcal{J}) = 0$ holds for every $\mathcal{J} \in \mathcal{F}$.

This definition implies that the receiver obtains no information about which specific subset $\mathcal{J}$ of servers is storing the message, ensuring the servers' location privacy and anonymity.

Remark 5 (Privacy in the bi-regular PID setting)**.** The privacy condition in the bi-regular PID setting has been presented in Eq. (6) of [12, Section II] that is expressed not for the subset $\mathcal{J}$ of servers but for the message index, unlike Definition 2.

5 Characterization of Correctness and Privacy

This section shall characterize the correctness and privacy of our generalized PID formulation, and show that the conditions for correctness and privacy are expressed by the minimum Hamming weight of the code C.

We first present the following theorem regarding the conditions for the correctness of PID when using an arbitrary linear code as the randomizing code C.

Theorem 1. Consider a PID scheme with coded storage with a linear code $C \subseteq \mathbb{F}^n$ of $\dim C = n - l$. Let $\mathcal{F}$ be a family of all possible index sets for the scheme, given as

$$\mathcal{F} \subseteq \{\mathcal{J} \subsetneq [n] : |\mathcal{J}| \geq n - d(C^{\perp}) + 1\}.$$

Then the scheme achieves the correctness in Definition 1.

Proof. The correctness is attained when a coset $\vec{y} + C \in \mathbb{F}^n/C$ is uniquely determined from $\vec{a} \in \vec{y} + C$. This implies that by a parity check matrix $H \triangleq [h_1, \ldots, h_n] \in \mathbb{F}^{l \times n}$ of C, the message $\vec{w} \in \mathbb{F}^l$ can be determined as a *syndrome* in the context of error-correcting codes and coset coding [8]. From this observation, δ is given as H and we have a syndrome

$$\begin{aligned}\delta(\vec{a}) &= \vec{a}H^{\mathrm{T}} = \vec{c}H^{\mathrm{T}} + \vec{y}H^{\mathrm{T}} \\ &= \vec{y}H^{\mathrm{T}}.\end{aligned}$$

We thus see that $\vec{y}H^{\mathrm{T}}$ must associate with $\vec{w}$ uniquely. Here we recall that $P_{\mathcal{J}}(\vec{y}) = \vec{y}$ holds, that is, the j-th element for $j \notin \mathcal{J}$ is always zero. Thus, from now on, we will consider the submatrix $H_{\mathcal{J}}$ of H corresponding to $\mathcal{J}$, defined as $H_{\mathcal{J}} \triangleq [h_j : j \in \mathcal{J}] \in \mathbb{F}^{l \times |\mathcal{J}|}$. Then, we see that $\vec{w}$ can be uniquely determined when $H_{\mathcal{J}}$ is full-rank, i.e.,

$$\operatorname{rank} H_{\mathcal{J}} = \dim \left(P_{\mathcal{J}}(C^{\perp})\right) = l,$$

from $|\mathcal{J}| \geq l$. Here, by Forney's first duality lemma [4], letting $\bar{\mathcal{J}} \triangleq [n] \setminus \mathcal{J}$, we have

$$\begin{aligned}\dim (P_{\mathcal{J}}(C^{\perp})) &= \dim C^{\perp} - \dim (C^{\perp})_{\bar{\mathcal{J}}} \\ &= l - \dim(C^{\perp})_{\bar{\mathcal{J}}}.\end{aligned}$$

From this equation, the following equalities hold,

$$\begin{aligned}&\max_{\mathcal{J} \subsetneq [n]} \{|\mathcal{J}| : \dim \left(P_{\mathcal{J}}(C^{\perp})\right) = l - 1\} \\ &\quad = \max_{\mathcal{J} \subsetneq [n]} \{|\mathcal{J}| : \dim (C^{\perp})_{\bar{\mathcal{J}}} = 1\} \\ &\quad = n - \min_{\bar{\mathcal{J}} \subsetneq [n]} \{|\bar{\mathcal{J}}| : \dim (C^{\perp})_{\bar{\mathcal{J}}} = 1\} \\ &\quad = n - d(C^{\perp}),\end{aligned}$$

where the third equality follows from the definition of the shortened code. Therefore, if the encoded messages are placed on any subset $\mathcal{J}$ of at least $n - d(C^{\perp}) + 1$ servers, $\phi_{\mathcal{J}}$ attaining the theorem exists for $\mathcal{J}$. □

This theorem implies that the correctness of the PID scheme is guaranteed as long as the number of servers storing the encoded pieces of the message is at least $n - d(C^{\perp}) + 1$. On the other hand, by the proof of Theorem 1, we immediately have the following corollary that indicates the tightness of the condition shown in Theorem 1.

Corollary 1. Consider $\mathcal{F}$ containing subsets $\mathcal{J}$ of size $|\mathcal{J}| \geq n - d(C^{\perp})$. Then, there exists a choice of $\mathcal{F}$ such that the correctness is not satisfied.

Remark 6. As noted in Remark 2, the scheme proposed by Vaidya et al. can be regarded as a specific instance of our generalized framework. Specifically, the scheme of Vaidya et al. [12] is equivalent to the case of $\mathcal{F} = \mathcal{F}_{\mathrm{MDS}} = \{\mathcal{J} \subsetneq [n] : |\mathcal{J}| = l\}$ and an MDS code C with $d(C) = l + 1$. By $d(C^{\perp}) = n - l + 1$, we have

$$n - d(C^{\perp}) + 1 = n - (n - l + 1) + 1 = l.$$

Therefore, we see that their scheme achieves correctness by Theorem 1 for $\mathcal{F}_{\mathrm{MDS}} \subseteq \{\mathcal{J} \subsetneq [n] : |\mathcal{J}| \geq l\}$.

Next, we present the following theorem that clarifies the conditions required for the privacy of PID with coded storage using an arbitrary linear code C.

Theorem 2. Consider a PID scheme with coded storage using a linear code $C \subseteq \mathbb{F}^n$ of dimension $\dim C = n - l$. Then, if the correctness is satisfied, the scheme also ensures the privacy in Definition 2.

Proof. First, we compute the mutual information $I(\vec{a}; \mathcal{J})$ as

$$\begin{aligned} I(\vec{a}; \mathcal{J}) = & \underbrace{I(\vec{a}, \vec{y}; \mathcal{J})}_{=H(\vec{a},\vec{y})-H(\vec{a},\vec{y}|\mathcal{J})} - \underbrace{I(\vec{y}; \mathcal{J}|\vec{a})}_{=H(\vec{y}|\vec{a})-H(\vec{y}|\vec{a},\mathcal{J})} \\ = & \underbrace{H(\vec{a}, \vec{y})}_{=H(\vec{a})+H(\vec{y}|\vec{a})} - \underbrace{H(\vec{a}, \vec{y}|\mathcal{J})}_{=H(\vec{a},\vec{y},\vec{w}|\mathcal{J})-H(\vec{w}|\vec{a},\vec{y},\mathcal{J})} \\ & - H(\vec{y}|\vec{a}) + H(\vec{y}|\vec{a}, \mathcal{J}) \\ = & H(\vec{a}) - H(\vec{a}, \vec{y}, \vec{w}|\mathcal{J}) \\ & + H(\vec{w}|\vec{a}, \vec{y}, \mathcal{J}) + H(\vec{y}|\vec{a}, \mathcal{J}). \end{aligned} \tag{2}$$

Here, we have $H(\vec{a}, \vec{y}, \vec{w}|\mathcal{J}) = H(\vec{c}, \vec{y}, \vec{w}|\mathcal{J})$ since $\vec{a} = \vec{c} + \vec{y}$ holds. Also recall that $\vec{w}$ can be uniquely determined by $\vec{a}$ when the correctness is satisfied. Thus we have $H(\vec{w}|\vec{a}, \vec{y}, \mathcal{J}) = 0$ and $H(\vec{y}|\vec{a}, \mathcal{J}) = H(\vec{y}|\vec{a}, \vec{w}, \mathcal{J})$. Moreover, from $\mathcal{J}$ and $\vec{w}$, we can uniquely determine $\vec{y} = \phi_{\mathcal{J}}(\vec{w})$, and hence $H(\vec{y}|\vec{a}, \vec{w}, \mathcal{J}) = 0$ holds. Therefore Eq. (2) can be simplified as

$$I(\vec{a}; \mathcal{J}) = H(\vec{a}) - H(\vec{c}, \vec{y}, \vec{w}|\mathcal{J}). \tag{3}$$

Since $\vec{c} \in C$ is chosen uniformly at random and independently from $\vec{y}$ and $\vec{w}$, Eq. (3) can be rewritten as

$$\begin{aligned} I(\vec{a}; \mathcal{J}) &= H(\vec{a}) - H(\vec{c}|\mathcal{J}) - H(\vec{y}, \vec{w}|\mathcal{J}) \\ &= H(\vec{a}) - H(\vec{c}|\mathcal{J}) - H(\vec{w}|\mathcal{J}) - \underbrace{H(\vec{y}|\vec{w}, \mathcal{J})}_{=0}. \end{aligned} \tag{4}$$

Recall that $\vec{a} \in \mathbb{F}^n$, i.e., $H(\vec{a}) \leq n$. Furthermore, $H(\vec{c}|\mathcal{J}) = H(\vec{c}) = n - l$ because $\vec{c}$ and $\mathcal{J}$ are independent. Similarly, $H(\vec{w}|\mathcal{J}) = H(\vec{w}) = l$ because $\vec{w}$ and $\mathcal{J}$ are independent. Thus, Eq. (4) simplifies to

$$I(\vec{a}; \mathcal{J}) \leq n - (n - l) - l = 0.$$

By the non-negativity of the mutual information, we have $I(\vec{a}; \mathcal{J}) = 0$. This concludes the proof. □

Theorem 2 does not directly depend on the code parameter of the linear code C employed in the scheme unlike Theorem 1. However, since the correctness depends on $d(C^{\perp})$, we see that the privacy condition is indirectly dependent on $d(C^{\perp})$ through the correctness.

Remark 7. Consider the scheme proposed by Vaidya et al. [12]. Then, since the correctness holds, Theorem 2 also holds, and their scheme ensures privacy.

Remark 8 (Message secrecy via wiretap channel II [8]**).** Using the same randomizing code C, the message $\vec{w}$ is perfectly secret against the observation of any $\mu \leq d(C^{\perp}) - 1$ coordinates of $\vec{a}$ by an external eavesdropper, which is directly derived from the relationship to wiretap channel II [8]. See Appendix B and Proposition 1 for details. Note that this paper focuses on privacy/anonymity; message secrecy is beyond our main scope, and the formal treatment is deferred to Appendix B.

6 Conclusion

In this paper, we have generalized the existing framework to accommodate a wider range of data placement scenarios, and have reformulated the problem of private information delivery (PID) with coded storage by focusing on the location privacy (anonymity) of storage servers rather than the message-index privacy. We then introduced two new information-theoretic properties—correctness and privacy—for this generalized PID formulation. Allowing arbitrary linear codes (without being restricted to MDS codes considered in the prior work), we characterized explicit conditions under which both properties are achieved, expressed in terms of the minimum Hamming weight of the dual of the linear code used by the scheme. These results provide design principles for PID schemes across diverse system architectures, overcoming limitations of prior approaches constrained by bi-regular placement patterns and MDS requirements. A natural direction for future work is to investigate the capacity of PID with coded storage under several non-regular constraints on the server settings, e.g., the heterogeneous storage size and message placement scenarios.

Acknowledgments

This work was supported in part by Japan Society for the Promotion of Science, Japan, KAKENHI, under Grant JP22K11994 and JP25K03108.

References

[1] G. R. Blakley, "Safeguarding cryptographic keys," in *Proc. AFIPS 1979 National Computer Conf.*, Jun. 1979, pp. 313–317.

[2] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *Proc. FOCS 1995*, Oct. 1995, pp. 41–50.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, Jan. 2006.

[4] G. Forney, "Dimension/length profiles and trellis complexity of linear block codes," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1741–1752, Nov. 1994.

[5] J. Kurihara, T. Uyematsu, and R. Matsumoto, "Secret sharing schemes based on linear codes can be precisely characterized by the relative generalized hamming weight," *IEICE Trans. Fundamentals*, vol. 95, no. 11, pp. 2067–2075, Nov. 2012.

[6] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. North-Holland Mathematical Library, 1977.

[7] K. Nakano, J. Kurihara, and T. Tanaka, "Extensive study on the security of private information delivery

from coded storage," to appear in *IEICE Trans. Fundamentals*, vol. E109-A, no. 3, Mar. 2026.

[8] L. H. Ozarow and A. D. Wyner, "Wire-tap channel II," *AT&T Bell Lab. Tech. J.*, vol. 63, no. 10, pp. 2135–2157, Dec. 1984.

[9] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, Nov. 1979.

[10] H. Sun, "Private information delivery," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7672–7683, Dec. 2020.

[11] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4075–4088, Jul. 2017.

[12] K. Vaidya and B. S. Rajan, "Private information delivery with coded storage," in *Proc. IEEE ISIT 2022*, Aug. 2022, pp. 2011–2015.

A Example of PID with Coded Storage

Here we present a simple example of PID with coded storage to illustrate the concepts introduced in this paper.

Example 1. Let $\mathbb{F}$ be the finite field of order 5, $n = 3$, $l = 2$ and $\mathcal{F} = \{\{1,2\},\{2,3\}\}$. Here we choose $\mathcal{J} = \{1,2\} \in \mathcal{F}$ as an example. Let $C \triangleq \{\vec{x} \in \mathbb{F}^3 : \vec{x}H^{\mathrm{T}} = \vec{0}\} \subseteq \mathbb{F}^3$ be a Reed-Solomon code with $\dim C = 1$, defined by a parity check matrix

$$H \triangleq \left[\underbrace{\begin{matrix}1 & 1\\ 1 & 2\end{matrix}}_{=H_{\{1,2\}}}\ \begin{matrix}1\\ 3\end{matrix}\right] \in \mathbb{F}^{2\times 3}.$$

In the placement phase, for a message $\vec{w} = [w_1, w_2] \in \mathbb{F}^2$ and the index set $\mathcal{J} = \{1,2\}$, the encoded message $\vec{y} = [y_1, y_2, y_3]$ is generated, and y_1 and y_2 are respectively placed on the first and second servers. To this end, we set $\phi_{\mathcal{J}} : \mathbb{F}^2 \to P_{\mathcal{J}}(\mathbb{F}^3)$ as

$$\phi_{\mathcal{J}}(\vec{w}) \triangleq \vec{w}\left[\ (H_{\{1,2\}}^{-1})^{\mathrm{T}}\ \ \begin{matrix}0\\ 0\end{matrix}\ \right] = \vec{w}\begin{bmatrix}2 & 4 & 0\\ 4 & 1 & 0\end{bmatrix},$$

and obtain $y_1 = 2w_1 + 4w_2$ and $y_2 = 4w_1 + w_2$. Nothing is placed on the third server ($y_3 = 0$). We also choose $\vec{c} = [c_1, c_2, c_3] \in C$ uniformly at random, and place c_j on the j-th server for $j \in [3]$. Next, in the delivery phase, the j-th server transmits $a_j = c_j + y_j$ for $j \in \{1,2\}$ and $a_3 = c_3$ to the receiver. Then, the receiver reconstructs $\vec{w}$ from $\vec{a} = [a_1, a_2, a_3]$ by a deterministic map $\delta : \mathbb{F}^3 \to \mathbb{F}^2$ defined as $\delta(\vec{a}) \triangleq \vec{a}H^{\mathrm{T}}$. We can easily verify that $\vec{w}$ can be reconstructed as

$$\begin{aligned}\delta(\vec{a}) &= \vec{a}H^{\mathrm{T}}\\ &= [c_1 + 2w_1 + 4w_2, c_2 + 4w_1 + w_2, c_3]\begin{bmatrix}1 & 1\\ 1 & 2\\ 1 & 3\end{bmatrix}\\ &= [\underbrace{c_1 + c_2 + c_3}_{=0\ (\text{By } \vec{c}H^{\mathrm{T}} = \vec{0})} + w_1, \underbrace{c_1 + 2c_2 + 3c_3}_{=0\ (\text{By } \vec{c}H^{\mathrm{T}} = \vec{0})} + w_2].\end{aligned}$$

Note that even if a different set is chosen as $\mathcal{J}$ from $\mathcal{F}$, i.e., $\{2,3\}$ in this case, the receiver always uses the same decoding map δ, i.e., H.

B Message Secrecy via Wiretap Channel II

Here we explain the relation between the PID with coded storage and coding for the wiretap channel II [8]. We also show that the PID with coded storage has an additional security property for the message. Although this property has not been discussed in prior works [7, 12], it can be immediately deduced from the established results for the wiretap channel II.

Definition 3 (Wiretap channel II [8]). Consider that there exist n noiseless channels where each channel can transmit one symbol in $\mathbb{F}$ and an eavesdropper can observe any μ out of the n channels ($\mu < n$). In the problem of wiretap channel II, a sender aims to securely transmits a message $\vec{w} \in \mathbb{F}^l$ ($l < n$) to a receiver over the channels while keeping the message $\vec{w}$ secret from the eavesdropper.

To securely convey the message over the wiretap channel II, the sender employs a coding scheme called *coset coding* to encode $\vec{w}$ to an n-dimensional vector $\vec{b} \in \mathbb{F}^n$.

Definition 4 (Coset coding for wiretap channel II [5, 8]). Let $C \subsetneq \mathbb{F}^n$ be a linear code with $\dim C = n - l$. We first arbitrarily choose a subspace $\mathcal{S} \subseteq \mathbb{F}^n$ such that $\mathcal{S} + C = \mathbb{F}^n$ and $\mathcal{S} \cap C = \{\vec{0}\}$, and fix an isomorphism $f : \mathbb{F}^l \to \mathcal{S}$. Then, we choose $\vec{b} \in f(\vec{w}) + C$ uniformly at random, and output $\vec{b}$ as the encoded vector. Each symbol of the generated vector $\vec{b}$ is then transmitted over each channel instead of $\vec{w}$.

Note that $\mathcal{S}$ can be always chosen and $\dim \mathbb{F}^n/C = \dim \mathcal{S} = l$, and that $\vec{w}$ is always uniquely determined from $\vec{b}$. For the security in the wiretap channel II, the following theorem has been given in [5, 8][2].

Theorem 3 ([8, Lemma 4.1], [5, Theorem 9]). Consider the wiretap channel II and coset coding given above. For $\mathcal{E} \subseteq [n]$ with $|\mathcal{E}| = \mu$ and $\vec{b} = [b_1, \ldots, b_n] \in \mathbb{F}^n$, let $\vec{b}_{\mathcal{E}} \triangleq [b_j : j \in \mathcal{E}]$ be symbols observed by the eavesdropper. Then, $I(\vec{w}; \vec{b}_{\mathcal{E}}) = 0$ always holds for any $\mathcal{E}$ if $\mu < d(C^{\perp})$, i.e., no information about $\vec{w}$ is leaked to the eavesdropper.

Here we recall that for the PID with coded storage, the reconstructing the message $\vec{w}$ exactly corresponds to determining the coset $\phi_{\mathcal{J}}(\vec{w}) + C = \vec{y} + C \in \mathbb{F}^n/C$ as discussed in the proof of Theorem 1. We thus see that the decoding in the PID with coded storage is equivalent to that of the coset coding for the wiretap channel II. Therefore, by importing Theorem 3, we immediately obtain the following proposition.

Proposition 1. Consider a PID scheme with coded storage using a linear code $C \subseteq \mathbb{F}^n$. Suppose that the correctness given in Definition 1 is satisfied. For $\mathcal{E} \subseteq [n]$, we assume there exists an eavesdropper observing $|\mathcal{E}|$ symbols in the transmitted $\vec{a}$, represented as $\vec{a}_{\mathcal{E}} \triangleq [a_j : j \in \mathcal{E}]$. Then, we always have $I(\vec{w}; \vec{a}_{\mathcal{E}}) = 0$ for any $\mathcal{E}$ of size $|\mathcal{E}| < d(C^{\perp})$, i.e., no information about $\vec{w}$ is leaked to the eavesdropper.

Namely, in addition to the privacy in Definition 2, an additional security property for the message is automatically guaranteed in the PID with coded storage.

[2] In [5], the theorem is given in the context of secret sharing schemes [9], which can be viewed as a generalization of the wiretap channel II.