PAPER Special Section on Discrete Mathematics and Its Applications

# Private Information Retrieval from Coded Storage in the Presence of Omniscient and Limited-Knowledge Byzantine Adversaries*

**Jun KURIHARA**[†,††,†††a)], **Toru NAKAMURA**[†††,††††], ***and*** **Ryu WATANABE**[††††], ***Members***

**SUMMARY** This paper investigates an adversarial model in the scenario of *private information retrieval* (PIR) from $n$ coded storage servers, called *Byzantine adversary*. The Byzantine adversary is defined as the one altering $b$ server responses and erasing $u$ server responses to a user's query. In this paper, two types of Byzantine adversaries are considered; 1) the classic *omniscient* type that has the full knowledge on $n$ servers as considered in existing literature, and 2) the reasonable *limited-knowledge* type that has information on only $b+u$ servers, i.e., servers under the adversary's control. For these two types, this paper reveals that the resistance of a PIR scheme, i.e., the condition of $b$ and $u$ to correctly obtain the desired message, can be expressed in terms of a code parameter called the *coset distance* of linear codes employed in the scheme. For the omniscient type, the derived condition expressed by the coset distance is tighter and more precise than the estimation of the resistance by the minimum Hamming weight of the codes considered in existing researches. Furthermore, this paper also clarifies that if the adversary is limited-knowledge, the resistance of a PIR scheme could exceed that for the case of the omniscient type. Namely, PIR schemes can increase their resistance to Byzantine adversaries by allowing the limitation on adversary's knowledge.

***key words:*** *private information retrieval, coded storage, coset distance, Byzantine adversary*

## 1. Introduction

### 1.1 Background

*Private information retrieval* (PIR) [2] is a protocol allowing a user to retrieve a message from a set of $n$ storage (database) servers, without revealing any information about the identity of the user's desired message to each server or sets of colluded servers. Recently, PIR schemes for *distributed coded storage* have been actively studied, e.g., [4], [15]–[17]. In PIR schemes from coded storage, every original message is encoded into a codeword of an $[n, k]$ linear code $C$, and each coordinate of an encoded message is stored at each server. Also, queries from a user are also chosen from another linear code to privately retrieve a message.

In the context of PIR schemes from coded storage, several new adversarial settings have been introduced. *Byzantine adversary* [5], [16] is such one who intrudes a set of servers and corrupts server responses to a user's query. Consider a Byzantine adversary who pollutes $b$ responses and erases $u$ responses in $n$ server responses to a user's query. Then, provided $n > k + t + 2b + u - 1$, the explicit scheme of Tajeddine et al. [16] allows a user to correctly retrieve the desired message in the presence of the Byzantine adversary without revealing the user's demand to at most $t$ colluded servers. This scheme was proven to be optimal in the sense of the download efficiency under a specific condition of a PIR scheme [5], i.e., leveraging specific codes for storage construction and query generation. However, the resistance to the Byzantine adversary in PIR schemes based on *arbitrary linear codes* have not been characterized yet. Namely, the condition on the numbers of polluted and erased responses, $b$ and $u$, to correctly retrieve the desired message is still unknown for a PIR scheme employing arbitrary codes for storage and query. Moreover, considering the intrusion to storage servers, the adversarial model considered in existing researches is not practical since it has been assumed to have full-knowledge of *all* of $n$ servers about messages and queries for the corruption of a *subset* of $n$ responses.

### 1.2 Our Contribution

In this paper, we aim to express the resistance to Byzantine adversaries in arbitrary-code-based PIR schemes in terms of parameters of their employed codes. To this end, we first (re)formulate the process of PIR schemes from coded storage and redefine the decoding at the retrieval process as the *identification of a coset*. This decoding approach is completely different from that of Tajeddine et al. [16] based on the minimum Hamming distance [11], i.e., identification of a *codeword*. Secondly, we introduce two settings of Byzantine adversaries in PIR schemes from coded storage called *omniscient* and *limited-knowledge* ones, as in the scenarios of network coding [6], [19] and distributed storage coding [12]. Note that the omniscient adversary is a classical type of adversaries with *full-knowledge* of stored messages and queries of all $n$ servers, considered in existing PIR researches [5], [16]. On the other hand, the limited-knowledge adversary is more relaxed and reasonable than the omniscient one, which observes a subset of servers and controls responses only from the observed servers. For both adversary types, we reveal that the resistance to a Byzantine

Manuscript received September 18, 2020.
Manuscript revised December 11, 2020.
Manuscript publicized March 23, 2021.

[†]The author is with Graduate School of Applied Informatics, University of Hyogo, Kobe-shi, 650-0047 Japan.
[††]The author is with Zettant Inc., Tokyo, 103-0025 Japan.
[†††]The authors are with Advanced Telecommunication Research Institute International (ATR), Kyoto-fu, 619-0288 Japan.
[††††]The authors are with KDDI Research, Inc., Fujimino-shi, 356-8502 Japan.

*A preliminary version of the material in this paper was partially published at IEICE Communications Express (ComEx) [8].

a) E-mail: kurihara@ieee.org

DOI: 10.1587/transfun.2020DMP0018

adversary can be characterized a code parameter called the *coset distance* [3], [10][†] of the employed linear codes. Our contributions for each adversary type are described below in detail.

CONTRIBUTION FOR OMNISCIENT ADVERSARY: The omniscient Byzantine adversary pollutes $b$ responses and erases $u$ responses under full-knowledge of stored messages and queries at all of $n$ servers. As stated above, existing researches estimated the resistance to such an adversary, i.e., the condition on $b$ and $u$ for successful decoding, by the decoding approach via the minimum Hamming distance of employed codes. On the other hand, this paper expresses the resistance in terms of the coset distance, which is tighter and more precise than the existing approach. Moreover, we prove that the *capacity* introduced for a specific case in [16], i.e., the condition $n > k+t+2b+u-1$, can be easily explained by our condition based on the coset distance.

CONTRIBUTION FOR LIMITED-KNOWLEDGE ADVERSARY: On the other hand, the limited-knowledge Byzantine adversary intrudes $b + u$ servers, and controls responses from these servers with no knowledge about remaining non-intruded $n - b - u$ servers. For this setting, this paper derives the resistance of a PIR scheme expressed by the coset distance as well as the omniscient setting. We, furthermore, demonstrate that the resistance to the limited-knowledge adversary could exceed that to the omniscient one by allowing vanishing failure probability as [6], [12], [19]. In order to achieve this, we introduce an explicit scheme that employs pairwise *hashes* [12] generated from encoded messages with negligible overhead. Our scheme using the hashes allows the user to detect polluted responses and decode the original part of desired message by excluding polluted responses. We show that as a result, limiting the adversary's knowledge increases the resistance of a PIR scheme to Byzantine adversary.

We note that our characterizations summarized above provide analytical tools for PIR schemes using the coset distance via the generalized (re)formulation with arbitrary linear codes. In order to make this clear and gain readers better understanding, this paper re-explains and analyzes existing schemes, e.g., [16], as examples of specific codes and settings from our new viewpoint introduced in this paper. We believe that our characterizations are useful guides for future designer of PIR schemes with specialized codes.

### 1.3 Organization

The remainder of this paper is organized as follows. Section 2 introduces notations and definitions used in this paper. Section 3 reformulates the process of private information retrieval from coded storage, and introduces its relationship to existing definitions. Section 4 formally defines the omniscient Byzantine adversary and the limited-knowledge one considered in the scenario of PIR from coded storage, and introduces the resistance of PIR schemes to these adversaries. Then, Sect. 5 presents the condition to guarantee the resistance to the omniscient adversary, expressed in the coset distance. On the other hand, Sect. 6 starts from a toy example of our scheme for the limited-knowledge adversary in order to help readers gain better understanding, and then introduces the condition on the resistance to the limited-knowledge adversary. The detailed explanation of our scheme for general cases will be given in Appendix. Section 7 finally concludes this paper.

[†]a.k.a the *first relative generalized Hamming weight* [10]

## 2. Notations and Preliminaries

Let $[n] \triangleq \{1, \ldots, n\}$. Let $\mathbb{F}_q$ stand for a finite field containing $q$ elements, and $\mathbb{F}_{q^v}$ be a $v$-degree field extension of $\mathbb{F}_q$ ($v > 0$). We denote by $\mathbb{F}_q^n$ an $n$-dimensional (row) vector space over $\mathbb{F}_q$, and similarly by $\mathbb{F}_{q^v}^n$ an $n$-dimensional vector space over $\mathbb{F}_{q^v}$. We regard a $v$-dimensional vector space $\mathbb{F}_q^v$ as $\mathbb{F}_{q^v}$, and an element of $\mathbb{F}_{q^v}$ is identified as a vector in $\mathbb{F}_q^v$. To explain this, we explicitly define $\lambda : \mathbb{F}_{q^v} \to \mathbb{F}_q^v$ as an $\mathbb{F}_q$-linear isomorphism that expands an element of $\mathbb{F}_{q^v}$ to a vector over $\mathbb{F}_q$ with respect to some fixed basis for $\mathbb{F}_{q^v}$ over $\mathbb{F}_q$. Note that for $a_1, a_2 \in \mathbb{F}_q$ and $b_1, b_2 \in \mathbb{F}_{q^v}$, we have

$$\lambda(a_1 b_1 + a_2 b_2) = a_1 \lambda(b_1) + a_2 \lambda(b_2). \tag{1}$$

We also denote by $\Lambda : \mathbb{F}_{q^v}^n \to \mathbb{F}_q^{v \times n}$ an isomorphism defined as $\Lambda(w) \triangleq [\lambda(w_1)^{\mathrm{T}}, \ldots, \lambda(w_n)^{\mathrm{T}}] \in \mathbb{F}_q^{v \times n}$ for $w = [w_1, \ldots, w_n] \in \mathbb{F}_{q^v}^n$.

An $[n, k]$ linear code $C$ over $\mathbb{F}_{q^v}$ is a $k$-dimensional subspace of $\mathbb{F}_{q^v}^n$. Let $C^{\perp} \triangleq \{v \in \mathbb{F}_{q^v}^n : vw^{\mathrm{T}} = 0, w \in C\}$ be the dual code of a code $C$ [11, p. 26, Ch. 1]. A subspace of a code is called a *subcode*. The *minimum Hamming distance* [11] and the *coset distance* [3], [10] of linear codes are defined as follows.

**Definition 1.** Let $\mathbb{F}$ be a certain finite field. For a subspace $C \subseteq \mathbb{F}^n$, the *minimum Hamming distance* [11] of $C$ is defined as $d_{\min}(C) \triangleq \min\{d(v, w) : v, w \in C, v \neq w\}$, where $d(v, w) \triangleq |\{i : v_i \neq w_i\}|$ and $v = [v_1, \ldots, v_n], w = [w_1, \ldots, w_n]$. For a subspace $C \subseteq \mathbb{F}^n$ and its subcode $\mathcal{D} \subsetneq C$, the *coset distance* [3], [10] is defined by

$$\begin{aligned} &d_{\min}(C/\mathcal{D}) \\ &\triangleq \min\left\{\min\left\{d(x, y) : \begin{array}{l} x \in v + \mathcal{D}, \\ y \in w + \mathcal{D} \end{array}\right\} : \begin{array}{l} v, w \in C, \\ v - w \notin \mathcal{D} \end{array}\right\} \\ &= \min\{d(v, 0) : v \in C \setminus \mathcal{D}\}. \end{aligned} \tag{2}$$

Here we introduce the following lemma showing the error-correcting capability based on the coset distance.

**Lemma 2** ([3, Lemma 1.1])**.** For two subspaces $C \subseteq \mathbb{F}^n$ and $\mathcal{D} \subsetneq C$ over a certain vector space $\mathbb{F}^n$, suppose that $d_{\min}(C/\mathcal{D}) > 2t$ holds, and that $z \in \mathbb{F}^n$ satisfies $\min\{d(z, x) : x \in C\} \leq t$. Then there exists a unique coset $c + \mathcal{D} \in C/\mathcal{D}$ with $\min\{d(x, y) : x \in z + \mathcal{D}, y \in c + \mathcal{D}\} \leq t$.

For a subspace $C \subseteq \mathbb{F}_{q^v}^n$, we denote by $C|\mathbb{F}_q$ a *subfield subcode* $C \cap \mathbb{F}_q^n$ [11, p. 207, Ch. 7]. Observe that $\dim C$

means the dimension of $C$ as a vector space over $\mathbb{F}_{q^v}$ whereas $\dim_{\mathbb{F}_{q^v}} C|\mathbb{F}_q$ is the dimension of $C|\mathbb{F}_q$ over $\mathbb{F}_q$. Similarly, for an $\mathbb{F}_q$-linear subspace $\mathcal{D} \subseteq \mathbb{F}_q^n$, we denote by $\dim_q \mathcal{D}$ the dimension of $\mathcal{D}$ over $\mathbb{F}_q$. Unless otherwise stated, we consider subspaces, dimensions, etc., over $\mathbb{F}_{q^v}$ instead of $\mathbb{F}_q$.

For subfield subcodes, here we introduce their properties as follows.

**Lemma 3** ([14, Lemma 1])**.** For a subspace $C \subseteq \mathbb{F}_{q^v}^n$, there is a basis of $C$ consisting of vectors in $\mathbb{F}_q^n$ if and only if $\dim_{\mathbb{F}_{q^v}} C = \dim_{\mathbb{F}_{q^v}} C|\mathbb{F}_q$.

**Lemma 4** ([14, Corollary 1])**.** For a subspace $C \subseteq \mathbb{F}_{q^v}^n$ with $\dim_{\mathbb{F}_{q^v}} C|\mathbb{F}_q = \dim_{\mathbb{F}_{q^v}} C$, $d_{\min}(C) = d_{\min}(C|\mathbb{F}_q)$ holds.

For two row vectors, $v = [v_1, \ldots, v_n] \in \mathbb{F}_{q^v}^n$ and $w = [w_1, \ldots, w_n] \in \mathbb{F}_{q^v}^n$, we denote by $v \circ w \triangleq [v_1 w_1, \ldots, v_n w_n]$ their Hadamard product. We also define

$$\mathcal{V} \circ \mathcal{W} \triangleq \mathsf{span}\{v \circ w : v \in \mathcal{V}, w \in \mathcal{W}\} \subseteq \mathbb{F}_{q^v}^n,$$

as an $\mathbb{F}_{q^v}$-linear subspace spanned by Hadamard products of vectors in two sets $\mathcal{V}, \mathcal{W} \subseteq \mathbb{F}_{q^v}^n$. Also for sets $\mathcal{V}', \mathcal{W}' \subseteq \mathbb{F}_q^n$ containing elements of $\mathbb{F}_q^n$, we explicitly represent

$$(\mathcal{V}' \circ \mathcal{W}')_q \triangleq \mathsf{span}_{\mathbb{F}_q}\{v \circ w : v \in \mathcal{V}', w \in \mathcal{W}'\} \subseteq \mathbb{F}_q^n,$$

as an $\mathbb{F}_q$-linear subspace.

## 3. Private Information Retrieval from Coded Storage

In existing literature [4], [5], [9], [15], [16], original and encoded messages are assumed to be much larger than the query information from the observation of [1]. From this assumption, these papers only considered the download cost for message retrieval and omitted the upload cost for queries. We explicitly follow such supposition and shall present a reformulated definition of PIR scheme from coded storage in *packetized form* in this section, as *packetized communication* in network coding [7], [13]. In particular, message symbols stored in a server are defined as vectors over $\mathbb{F}_q$, i.e., elements of a field extension of $\mathbb{F}_q$. The reason why we consider the packetization is that it is necessary to introduce a protocol of Sect. 6 for the limited-knowledge adversary similarly to existing researches on distributed storage codes [12] and network coding [6], [19]. Yet we emphasize that this packetized form is more generalized than that considered in existing researches in term of the field on which the PIR scheme is employed. Hence it yields a more generalized result than our preliminary paper [8] (See Sect. 5).

### 3.1 Reformulated Definition

First we define $v > 0$ as the *packet size* of each message to be stored. Following [4], [5], [9], [15], [16], first let $C$ be a linear code $C \subseteq \mathbb{F}_{q^v}^n$ of $\dim_{\mathbb{F}_{q^v}} C = k$, called the *storage code*. Suppose that for $i \in [m]$ the $i$-th original message $x^i$ is defined as an element of $\mathbb{F}_{q^v}^k$. Given a certain $\mathbb{F}_{q^v}$-linear bijection $f : \mathbb{F}_{q^v}^k \to C$, we have an encoded message $y^i \triangleq f(x^i) \in C$. We then define the matrix of encoded messages as

$$Y \triangleq \begin{bmatrix} y^1 \\ \vdots \\ y^m \end{bmatrix} = \begin{bmatrix} y_1^1 & \cdots & y_n^1 \\ \vdots & \ddots & \vdots \\ y_1^m & \cdots & y_n^m \end{bmatrix} \in \mathbb{F}_{q^v}^{m \times n},$$

where every row of $Y$ forms a codeword of $C$. The $j$-th server ($j \in [n]$) stores its $j$-th column $Y_j \triangleq [y_j^1, \ldots, y_j^m]^{\mathrm{T}} \in \mathbb{F}_{q^v}^{m \times 1}$. Note that each symbol $y_j^i$ is stored in the form of a $v$-dimensional vector over $\mathbb{F}_q$. Under this storage setting, a user wishes to retrieve a message by a (linear) *PIR scheme*. Here we shall introduce the PIR scheme from coded storage that is a generalized version of existing ones, [4], [5], [15], [16] and [8].

**Definition 5** (PIR Scheme from Distributed Coded Storage)**.** Considering the setting described above, a linear PIR scheme executes the following steps.

1. First fix an *inner query code* $\mathcal{D}_{\mathsf{in}} \subseteq \mathbb{F}_q^n$ and an *outer query code* $\mathcal{D}_{\mathsf{out}} \supsetneq \mathcal{D}_{\mathsf{in}}$ as $\mathbb{F}_q$-linear subspaces. Also choose a subspace $\mathcal{E} \subsetneq \mathbb{F}_q^n$ such that $\mathcal{D}_{\mathsf{out}} = \mathcal{E} + \mathcal{D}_{\mathsf{in}}$, and choose a vector $\epsilon \in \mathcal{E}$. We emphasize here that $\mathcal{D}_{\mathsf{in}}$, $\mathcal{D}_{\mathsf{out}}$ and $\mathcal{E}$ are all defined as *subspaces over the subfield* $\mathbb{F}_q$ of $\mathbb{F}_{q^v}$.
2. For every $i \in [m]$, set a probability space $(\mathcal{G}[i], \mu[i])$ of queries. $\mathcal{G}[i]$ is a set of all possible query matrices for $i$, given as
$$\mathcal{G}[i] = \left\{ \begin{bmatrix} g^1 \\ \vdots \\ g^m \end{bmatrix} \in \mathbb{F}_q^{m \times n} : \begin{array}{l} g^j \in \mathcal{D}_{\mathsf{in}} \text{ for } j \neq i, \\ g^j \in \epsilon + \mathcal{D}_{\mathsf{in}} \text{ for } j = i \end{array} \right\}. \quad (3)$$
When a user wishes to download $x^i \in \mathbb{F}_{q^v}^k$, a *total query* $G \in \mathcal{G}[i]$ is selected randomly according to the probability measure $\mu[i]$, typically the uniform probability $\mu[i] = 1/q^{m \dim_q \mathcal{D}_{\mathsf{in}}}$. Each $G$ is also defined as a matrix $G \triangleq [G_1, \ldots, G_n] \in \mathbb{F}_q^{m \times n}$, where the $j$-th column $G_j \in \mathbb{F}_q^{m \times 1}$ of length $m$ over $\mathbb{F}_q$ is sent to the $j$-th server.
3. Let $G_j \triangleq [g_j^1, \ldots, g_j^m]^{\mathrm{T}}$. At the $j$-th server, its response
$$r_j \triangleq G_j^{\mathrm{T}} Y_j = \sum_{l=1}^{m} g_j^l y_j^l \in \mathbb{F}_{q^v}, \quad (4)$$
is computed and transmitted to the user. Hence Eq. (4) can be simply viewed as an $\mathbb{F}_q$-linear combination of $v$-dimensional vectors $y_j^1, \ldots, y_j^m \in \mathbb{F}_q^v$. We then set $r \triangleq [r_1, \ldots, r_n] \in \mathbb{F}_{q^v}^n$ as the *total response* to $G$.
4. The user and servers repeatedly execute steps 1–3 until the user retrieves sufficient data segments to obtain $y^i$, i.e., the original message $x^i$.

Notations used in Definition 5 are summarized in Table 1. In the definition, we explicitly defined the storage code $C$ over the field extension $\mathbb{F}_{q^v}$ while inner and outer query codes are defined over the subfield $\mathbb{F}_q$. Note that the ratio of the upload cost over the download cost is given as

**Table 1** Notations used to illustrate the PIR scheme.

| Notation | Explanation |
|---|---|
| $v$ | Packet size |
| $n$ | Number of servers |
| $m$ | Number of stored messages |
| $C$ | Storage code, $C \subseteq \mathbb{F}_{q^v}^n$ with $\dim C = k$ |
| $\mathcal{D}_{\mathsf{in}}$ | Inner query code, $\mathcal{D}_{\mathsf{in}} \subseteq \mathbb{F}_q^n$ |
| $\mathcal{D}_{\mathsf{out}}$ | Outer query code, $\mathcal{D}_{\mathsf{out}} \supseteq \mathcal{D}_{\mathsf{in}}$ |
| $\mathcal{E}$ | A subspace satisfying $\mathcal{D}_{\mathsf{out}} = \mathcal{E} + \mathcal{D}_{\mathsf{in}}$ |
| $Y_j$ | Encoded messages stored at $j$-th server, $Y_j \in \mathbb{F}_{q^v}^{m\times 1}$ |
| $Y$ | All encoded messages in $n$ servers, $Y = [Y_1, \ldots, Y_n]$ |
| $G_j$ | A query to $j$-th server, $G_j \in \mathbb{F}_q^{m\times 1}$ |
| $G$ | A total query to $n$ servers, $G = [G_1, \ldots, G_n]$ |
| $r_j$ | A response from $j$-th server, $r_j \in \mathbb{F}_{q^v}$ |
| $r$ | A total response from $n$ servers, $r = [r_1, \ldots, r_n]$ |

$\frac{mn\log q}{n\log q^v} = O(\frac{1}{v})$, which vanishes with increasing the packet size $v$. Thus the upload cost can be omitted for large $v$ as [1].

**Remark 6.** Let the number of colluded servers be $t \leq n$. Then in the scheme of Definition 5, no information on the user's demand $i$ is leaked to the adversary when $t \leq d_{\mathsf{min}}(\mathcal{D}_{\mathsf{in}}{}^{\perp_q}) - 1$ [4, Theorem 8], where $\mathcal{D}_{\mathsf{in}}{}^{\perp_q} \subseteq \mathbb{F}_q^n$ is a dual of $\mathcal{D}_{\mathsf{in}}{}^{\perp_q}$ over $\mathbb{F}_q^n$.

### 3.2 Relation to Existing Definition

Here we explain how the reformulated definition given above is related to existing ones of the PIR scheme from coded storage, from the perspectives of *parallelization* and *correctness*.

#### 3.2.1 Parallelization

Unlike Definition 5, existing schemes [4], [5], [8], [9], [15], [16] use a storage code defined over not $\mathbb{F}_{q^v}$ but $\mathbb{F}_q$, and implicitly assume that an encoded message consists of multiple distinct codewords of the storage code. To retrieve a message, they employ *parallelization* or *superposition* of multiple retrieval processes of codewords with the same query as described in [1, Remark 3]. By regarding the packet size $v$ as the degree of parallelization, Definition 5 coincides with these schemes if an additional condition on $C \subseteq \mathbb{F}_{q^v}^n$ is introduced. The condition is that it satisfies $\dim C = \dim C|\mathbb{F}_q$. This can be proved by introducing the following simple lemma.

**Lemma 7.** Assume that a subspace $C \subseteq \mathbb{F}_{q^v}^n$ satisfies $\dim C|\mathbb{F}_q = \dim C = k$. Then for a codeword $c \in C$, each row of its matrix form $\Lambda(c) \in \mathbb{F}_q^{v\times n}$ is a codeword of $C|\mathbb{F}_q$.

*Proof.* From Lemma 3, there is a basis $\{\beta_1, \ldots, \beta_k\}$ of $C$ consisting of vectors in $\mathbb{F}_q^n$ when $\dim C|\mathbb{F}_q = \dim C$. Then for $c \in C$, we have $c = \sum_{l=1}^k u_l\beta_l$ for some $u_1, \ldots, u_k \in \mathbb{F}_{q^v}$. Thus we obtain

$$\begin{aligned}\Lambda(c) &= \Lambda\left(\sum_{l=1}^k u_l\beta_l\right) = \sum_{l=1}^k \Lambda(u_l\beta_l)\\ &= \sum_{l=1}^k \left[\beta_{l,1}\lambda(u_l)^{\mathrm{T}}, \ldots, \beta_{l,n}\lambda(u_l)^{\mathrm{T}}\right] \quad \text{(by Eq. (1))}\\ &= \begin{bmatrix}\sum_{l=1}^k u_{l,1}\beta_l\\ \vdots\\ \sum_{l=1}^k u_{l,v}\beta_l\end{bmatrix},\end{aligned}$$

where $\beta_l = [\beta_{l,1}, \ldots, \beta_{l,n}] \in \mathbb{F}_q^n$ and $\lambda(u_l) = [u_{l,1}, \ldots, u_{l,v}] \in \mathbb{F}_q^v$. This yields the lemma. □

From Lemma 7, each row of the matrix form $\Lambda(y^i)$ of an encoded message $y^i \in C$ can be viewed as a distinct codeword of $C|\mathbb{F}_q$ under the restriction $\dim C = \dim C|\mathbb{F}_q$. Recall that the retrieval process of Definition 5 just computes an $\mathbb{F}_q$-linear combination of $y_j^1, \ldots, y_j^m \in \mathbb{F}_{q^v}$ at the $j$-th server. Thus by letting $\lambda(y_j^i)_l \in \mathbb{F}_q$ be the $l$-th coordinate of $\lambda(y_j^i) \in \mathbb{F}_q^v$, the retrieval process can be viewed as the one generating $v$ distinct $\mathbb{F}_q$-linear combinations of $\lambda(y_j^1)_l, \ldots, \lambda(y_j^m)_l$, $l \in [v]$. Therefore, it can be viewed as a parallelized operation of $v$ independent retrieval processes of *sub*-PIR schemes with $C|\mathbb{F}_q, \mathcal{D}_{\mathsf{in}}, \mathcal{D}_{\mathsf{out}}$ all over $\mathbb{F}_q$, where each sub-PIR scheme is simply regarded as the case of $v = 1$ and $C = C|\mathbb{F}_q$ of Definition 5. Namely, *decoding the desired message can be done independently for each sub-PIR scheme when* $\dim C = \dim C|\mathbb{F}_q$. We can also view that it reuses one query for all sub-PIR schemes, as exactly mentioned in [1, Remark 3]. Here we mention that this consideration on the parallelization yields some special cases for our characterization of the resistance to Byzantine adversary, which will be given in Sects. 5 and 6.

#### 3.2.2 Correctness

The scheme of Definition 5 does not guarantee the *correctness* meaning that the user can correctly retrieve some coordinates of $y^i$, by itself. To guarantee it, extra conditions on $\mathcal{D}_{\mathsf{out}}, \mathcal{D}_{\mathsf{in}}$ and $\mathcal{E}$ are required, as explained below.

For the correctness, in [4], [15], [16], $\mathcal{E}$ and its basis $\{e^1, \ldots, e^{\dim_q \mathcal{E}}\}$ are chosen such that $\dim_q \mathcal{E} = \dim_q((C|\mathbb{F}_q \circ \mathcal{E})_q) \leq d_{\mathsf{min}}((C|\mathbb{F}_q \circ \mathcal{D}_{\mathsf{in}})_q) - 1$ by setting its basis with $d(e^j, 0) = 1$ for $\forall j \in [\dim_q \mathcal{E}]$ and $e = \sum_{j\in[\dim_q \mathcal{E}]} e^j$. This condition was generalized as the *strong linearity* in [5]. On the other hand, under the setting of a repetition code $C$, it is assumed in [9] that $\mathcal{D}_{\mathsf{out}} \subseteq \mathbb{F}_q^n$ and $\mathcal{D}_{\mathsf{in}} \subsetneq \mathcal{D}_{\mathsf{out}}$ are both maximum distance separable (MDS) [11], and that $\mathcal{E} \cap \mathcal{D}_{\mathsf{in}} = \{0\}$ holds. We emphasize that, however, these specific conditions are not required to characterize the resistance to Byzantine adversaries. Thus, throughout this paper, we do not consider such extra conditions on $\mathcal{D}_{\mathsf{out}}, \mathcal{D}_{\mathsf{in}}$ and $\mathcal{E}$ in order to simplify the discussion and make our focus solid.

## 4. Resistance of PIR schemes to Byzantine Adversary

In this section, we first formally introduce definitions of two types of Byzantine adversary in the scenario of PIR from coded storage. We then define the *resistance* of PIR schemes to Byzantine adversary from the perspective of *identification*

*of coset* by analyzing the retrieval process of Definition 5.

### 4.1 Two Types of Byzantine Adversaries in PIR

In the scenario of PIR from coded storage, *Byzantine adversaries* are defined as ones maliciously corrupting and deleting some symbols in the total response $r$. Inspired by studies on network coding and distributed storage [6], [12], [19], here we define two types of Byzantine adversaries; *omniscient* and *limited-knowledge* adversaries.

#### 4.1.1 Omniscient Byzantine Adversary

The *omniscient* Byzantine adversary has complete knowledge about all encoded messages $Y$ stored at $n$ servers and the total query $G$ issued from the user. Moreover the adversary can control $b+u$ storage servers in total. This means that the adversary can directly alter parts of encoded messages and queries at servers under the adversary's control. As a result, up to $b+u$ symbols in $r$ are possibly corrupted, where $b$ of them are polluted/erroneous and $u$ of them are erased. Note that, as far as we know, existing researches on PIR schemes from coded storage have considered only this type of Byzantine adversary [5], [16].

#### 4.1.2 Limited-Knowledge Byzantine Adversary

This type of Byzantine adversary has *limited knowledge* about stored messages and queries in the scenario of PIR from coded storage. It is reasonable that in distributed storage systems, an adversary who observes stored data and communication at a server is also allowed to controls the server, as considered in [12]. We thus suppose that the adversary can intrude $b+u$ servers, observe their stored message and communication, and alter/erase responses only from these $b+u$ servers. Then, the pollution of $b$ responses and elimination of $u$ responses are done by using the observed information.

We should note that Wang et al. [18] considered a limited-knowledge adversary located between servers and the user in the symmetric PIR scenario from replicated, i.e., non-coded, storage. On the other hand, our two adversary types given above considers the adversary who can directly observe stored messages and queries, and control intruded servers. Hence we claim that our settings about the Byzantine adversary can be viewed as a stronger one than that in the existing research.

### 4.2 Definition of the Resistance of PIR Schemes to Byzantine Adversaries

Here we shall (re)define the decoding process in PIR schemes in terms of *identification of a coset*, and introduce the notion of its resistance to the Byzantine adversary polluting $b$ symbols and erasing $u$ symbols in $r$, called the $(b,u)$-*Byzantine resistance*. We also explain how this is different from the existing decoding approach in [4], [5], [15], [16]. Part of this subsection has been presented in our preliminary paper [8].

We first give the following analytic lemma characterizing the component of the response $r$. Before the lemma, we recall that for two sets $\mathcal{X} \subseteq \mathbb{F}_{q^v}^n$ and $\mathcal{Y} \subseteq \mathbb{F}_q^n (\subseteq \mathbb{F}_{q^v}^n)$, their Hadamard product $\mathcal{X} \circ \mathcal{Y}$ forms a subspace of $\mathbb{F}_{q^v}^n$. In the following, we shall denote Hadamard products of $\mathcal{C}, \mathcal{D}_{\mathsf{out}}, \mathcal{D}_{\mathsf{in}}$ by $\mathcal{M}_{\mathsf{out}} \triangleq \mathcal{C} \circ \mathcal{D}_{\mathsf{out}}$ and $\mathcal{M}_{\mathsf{in}} \triangleq \mathcal{C} \circ \mathcal{D}_{\mathsf{in}}$.

**Lemma 8.** Consider the scheme of Definition 5 in which the user wishes to retrieve the $i$-th message ($i \in [m]$). Then the total response $r$ to the total query $G \in \mathcal{G}[i]$ is represented by an element of a coset in the quotient space $\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}$ as

$$r \in z + \mathcal{M}_{\mathsf{in}} \in \mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}},$$

where $z \in \mathcal{Z}$ is an element of a subspace $\mathcal{Z} \subseteq \mathcal{C} \circ \mathcal{E}$ with $\mathcal{Z} + \mathcal{M}_{\mathsf{in}} = \mathcal{M}_{\mathsf{out}}$ and $\mathcal{Z} \cap \mathcal{M}_{\mathsf{in}} = \{0\}$.

*Proof.* Recall that for each row $g^j$ of the total query $G \in \mathcal{G}[i]$, we have $g^j \in \mathcal{D}_{\mathsf{in}}$ for $j \neq i$, or $g^j \in \mathcal{E} + \mathcal{D}_{\mathsf{in}} = \mathcal{D}_{\mathsf{out}}$ for $j = i$ from Eq. (3). We denote by $d^j \in \mathcal{D}_{\mathsf{in}}$ an element of $\mathcal{D}_{\mathsf{in}}$ chosen to generate $g^j$, that is, $g^j = d^j$ for $j \neq i$ and $g^j = \epsilon + d^j$ for $j = i$. Since $\mathcal{D}_{\mathsf{in}} \subsetneq \mathcal{D}_{\mathsf{out}}$, we see $g^j \in \mathcal{D}_{\mathsf{out}}$, and hence $y^j \circ g^j \in \mathcal{M}_{\mathsf{out}}$ for any $y^j \in \mathcal{C}$. From this observation, the total response $r$ to $G \in \mathcal{G}[i]$ can be rewritten as follows.

$$\begin{aligned} r &= \underbrace{\sum_{j\in[m]} y^j \circ g^j}_{\in \mathcal{M}_{\mathsf{out}}} = \sum_{j\in[m]\setminus\{i\}} y^j \circ d^j + \underbrace{y^i \circ (\epsilon + d^i)}_{= y^i \circ \epsilon + y^i \circ d^i} \\ &= \underbrace{y^i \circ \epsilon}_{\in \mathcal{C}\circ\mathcal{E}} + \underbrace{\sum_{j\in[m]} y^j \circ d^j}_{\in \mathcal{M}_{\mathsf{in}}}. \end{aligned} \tag{5}$$

Here we immediately see that $y^i \circ \epsilon \in \mathcal{C} \circ \mathcal{E}$ is decomposed as $y^i \circ \epsilon = z + w$ for a certain $w \in (\mathcal{C} \circ \mathcal{E}) \cap \mathcal{M}_{\mathsf{in}}$. This completes the lemma. □

**Remark 9.** From the viewpoint of Lemma 8, the schemes of [4], [5], [9], [15], [16] can be re-explained as special cases satisfying $\mathcal{C} \circ \mathcal{E} = \mathcal{Z}$ by choosing $\mathcal{E} \subseteq \mathbb{F}_q^n$ as the one satisfying $\mathcal{M}_{\mathsf{in}} \cap \mathcal{C} \circ \mathcal{E} = \{0\}$. That is, $z = y^i \circ \epsilon$ holds for them.

In existing PIR schemes [4], [5], [15], [16], desired parts in a total response are hidden by a 'randomness', and the user cancels the randomness to elicit the desired parts. In the context of Lemma 8, $\mathcal{M}_{\mathsf{in}}$ is served as the source of the randomness, and $z \notin \mathcal{M}_{\mathsf{in}}$ is exactly the desired parts in $r$. To obtain $z$, existing schemes take an *indirect* approach in which the randomness given as a codeword of $\mathcal{M}_{\mathsf{in}}$ is first identified, and the randomness is subtracted from $r$. However, Lemma 8 clarifies that eliciting $z$ from $r$ coincides with identifying the coset $z + \mathcal{M}_{\mathsf{in}} \ni r$. From this perspective, the decoding problem to elicit $z$ from the total response can be redefined as *a game to identify the unique coset* $z + \mathcal{M}_{\mathsf{in}} \ni r$, i.e., a *direct* approach unlike existing researches. Here we shall give a definition of the resistance to Byzantine adversaries in this sense.

**Definition 10** (The $(b,u)$-Byzantine resistance)**.** Suppose that a Byzantine adversary observes some storage servers, and that the adversary maliciously returns erroneous responses from $b$ of $n$ servers and returns nothing from $u$ servers. Also assume that a user has no knowledge on servers returning erroneous responses. Then, instead of the original error-free total response $r$, the user receives its polluted version $\hat{r}$ in which $b$ coordinates are altered from $r$ and $u$ coordinates are null. Under this setting, a PIR scheme is called the one attaining the $(b,u)$-*Byzantine resistance* if the user can correctly identify the original coset $\mathcal{W} \in \mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}, r \in \mathcal{W}$ from $\hat{r}$.

In existing schemes [4], [5], [9], [15], [16], by introducing additional conditions on $\mathcal{E}, \mathcal{D}_{\mathsf{in}}$ and $\mathcal{D}_{\mathsf{out}}$, it is guaranteed that some coordinates of $y^i$ (and hence $x^i$) is directly obtained from the identified $z$, where $r \in z + \mathcal{M}_{\mathsf{in}} \in \mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}$ in our definition. Since conditions on $\mathcal{E}, \mathcal{D}_{\mathsf{in}}$ and $\mathcal{D}_{\mathsf{out}}$ in Definitions 5 and 10 are more relaxed than those, such decodability on the identified $z$ is not always guaranteed.

In the following sections, we shall introduce and characterize the $(b,u)$-Byzantine resistance of PIR schemes for both cases of omniscient and limited-knowledge adversaries.

## 5. Resistance to Omniscient Adversary

In this section, we first introduce a condition to guarantee the $(b,u)$-Byzantine Resistance to the omniscient adversary, which is expressed in terms of coset distance. We then show that our condition is tighter and more precise than the one derived via the existing approach based on the minimum Hamming distance.

### 5.1 Characterization using the Coset Distance

The $(b,u)$-Byzantine resistance to omniscient adversary given in Sect. 4.1.1 was first characterized by our preliminary paper [8] using the coset distance [3], [10]. This subsection refines the result and gives its special case where the storage code $C \subseteq \mathbb{F}_{q^v}^n$ satisfies $\dim C = \dim C|\mathbb{F}_q$. We present the following proposition as a refined version of [8, Theorem 8].

**Proposition 11.** Consider the PIR scheme in Definition 5, and let $\mathcal{M}_{\mathsf{out}} \triangleq C \circ \mathcal{D}_{\mathsf{out}}$ and $\mathcal{M}_{\mathsf{in}} \triangleq C \circ \mathcal{D}_{\mathsf{in}}$. Then, if $d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}) > 2b + u$, the user can attain the $(b,u)$-Byzantine resistance to an omniscient Byzantine adversary.

*Proof.* From Lemma 8, the total response $r$ to the total query $G$ can be represented as an element of a coset $z + \mathcal{M}_{\mathsf{in}}$, where the coset leader $z \in \mathcal{Z}$ is an element of a subspace $\mathcal{Z}$ with $\mathcal{Z} + \mathcal{M}_{\mathsf{in}} = \mathcal{M}_{\mathsf{out}}$ and $\mathcal{Z} \cap \mathcal{M}_{\mathsf{in}} = \{0\}$. Thus from Eq. (2) and Lemma 2, when at most any $b$ coordinates of $r$ are altered, the coset $z + \mathcal{M}_{\mathsf{in}} \ni r$ can be identified if $d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}) > 2b$. Considering $u$ erasures in $n$ coordinates in addition to $b$ errors, we additionally require more than or equal to distance $u$ to identify the coset. Therefore the theorem holds. □

**Remark 12** (Correction of [8, Theorem 8])**.** The preliminary version of Proposition 11 [8, Theorem 8] claimed that the converse part is attained as well as the direct part, i.e., "if and only if" has been posed in the statement. However, this is not always satisfied by the following reason. The candidate of possible error-free $r \in z + \mathcal{M}_{\mathsf{in}}$ is actually distributed not over a coset $z + \mathcal{M}_{\mathsf{in}}$ but over a set $\{v \circ \epsilon + v \circ w : v \in C, w \in \mathcal{D}_{\mathsf{in}}\}$ as analyzed in Eq. (5). We then observe $z + \mathcal{M}_{\mathsf{in}} \supseteq \{v \circ \epsilon + v \circ w : v \in C, w \in \mathcal{D}_{\mathsf{in}}\}$. Thus there might be a case where the coset can be uniquely identified even when $d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}) \le 2b + u$.

Consider the case where $C$ is restricted to the one with $\dim C = \dim C|\mathbb{F}_q$. Then the retrieval process for each codeword of $C \subseteq \mathbb{F}_{q^v}^n$ can be viewed as a superposition of $v$-distinct sub-PIR schemes for codewords of $C|\mathbb{F}_q$ over $\mathbb{F}_q$ by Lemma 7 and the analysis in Sect. 3.2. Thus, considering each sub-PIR scheme over $\mathbb{F}_q$ in the superposition independently, Proposition 11 yields the following corollary.

**Corollary 13.** Consider the PIR scheme in Definition 5. Let $\mathcal{M}_{\mathsf{out},q} \triangleq (C|\mathbb{F}_q \circ \mathcal{D}_{\mathsf{out}})_q$ and $\mathcal{M}_{\mathsf{in},q} \triangleq (C|\mathbb{F}_q \circ \mathcal{D}_{\mathsf{in}})_q$ be subspaces of $\mathbb{F}_q^n$. Set $C$ be a storage code satisfying $\dim C = \dim C|\mathbb{F}_q$, and employ the decoding per each sub-PIR scheme in the $v$-degree parallelization described in Sect. 3.2.1. Then the $(b,u)$-Byzantine resistance to the omniscient Byzantine adversary is guaranteed if $d_{\min}\left(\mathcal{M}_{\mathsf{out},q}/\mathcal{M}_{\mathsf{in},q}\right) > 2b + u$.

We see that when the storage code $C$ of $\dim C = \dim C|\mathbb{F}_q$ is employed and $d_{\min}\left(\mathcal{M}_{\mathsf{out},q}/\mathcal{M}_{\mathsf{in},q}\right) \ge d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}})$ holds, Corollary 13 should be applied and the decoding per each sub-PIR scheme in the $v$-degree parallelization should be employed.

The following example of Corollary 13 reveals that we can easily re-explains the existing result for specific codes from the viewpoint of the characterization using the coset distance via the formulation.

**Example 14.** As an example, we analyze the scheme of [16]. In the scheme, the basis of the storage code $C \subseteq \mathbb{F}_{q^v}^n$ can be represented as that of a generalized Reed-Solomon (GRS) code over the subfield $\mathbb{F}_q$, i.e., the subfield subcode $C|\mathbb{F}_q$ is a GRS code and $\dim C = \dim C|\mathbb{F}_q$ by Lemma 3. The inner query code $\mathcal{D}_{\mathsf{in}}$ is chosen as a GRS code over $\mathbb{F}_q$ generated with the same sequence of distinct elements of $\mathbb{F}_q$ as $C|\mathbb{F}_q$. The outer query code $\mathcal{D}_{\mathsf{out}}$ is a GRS code chosen such that $\dim_q \mathcal{D}_{\mathsf{out}} = \dim_q \mathcal{D}_{\mathsf{in}} + 1$. These implies that $\mathcal{M}_{\mathsf{in},q} = (C|\mathbb{F}_q \circ \mathcal{D}_{\mathsf{in}})_q$ and $\mathcal{M}_{\mathsf{out},q} = (C|\mathbb{F}_q \circ \mathcal{D}_{\mathsf{out}})_q$ are also GRS codes with

$$\dim_q \mathcal{M}_{\mathsf{out},q} = \dim_q \mathcal{M}_{\mathsf{in},q} + 1 = \dim C|\mathbb{F}_q + \dim_q \mathcal{D}_{\mathsf{in}},$$

from [4, Proposition 3], and that we have

$$d_{\min}(\mathcal{M}_{\mathsf{out},q}/\mathcal{M}_{\mathsf{in},q}) = n - \dim_q \mathcal{M}_{\mathsf{out},q} + 1,$$

from [10, Corollary 2]. We thus have

$$d_{\min}(\mathcal{M}_{\mathsf{out},q}/\mathcal{M}_{\mathsf{in},q}) = n - \dim C|\mathbb{F}_q - \dim_q \mathcal{D}_{\mathsf{in}} + 1 = n - \dim C - \dim_q \mathcal{D}_{\mathsf{in}} + 1.$$

Therefore from Corollary 13, the scheme attains the $(b,u)$-Byzantine resistance to the omniscient adversary if $n - \dim C - \dim_q \mathcal{D}_{\mathsf{in}} - 1 > 2b + u$ is satisfied.

### 5.2 Tightness of Our Characterization

As the conclusion of this section, we prove that our result is always equal to or tighter than the analysis by the existing approach [16]. Although the existing approach assumes specific codes, i.e., GRS codes, the resistance has been computed by a classical approach of error-correction based on the *minimum Hamming distance* of $\mathcal{M}_{\mathsf{out}}$. In particular, since $r$ is a codeword of $\mathcal{M}_{\mathsf{out}}$ (a GRS code in [16]), the condition $d_{\min}(\mathcal{M}_{\mathsf{out}}) > 2b + u$ to correctly obtain $r$ itself from $\hat{r}$ has been regarded as the sufficient condition of the decodability of the desired message, unlike our approach of coset identification given as Definition 10.

Comparing the definition of the minimum Hamming distance and the coset distance given in Definition 1, we immediately see that for any subspaces $\mathcal{X} \subseteq \mathbb{F}_{q^v}^n$ and $\mathcal{Y} \subsetneq \mathcal{X}$, $d_{\min}(\mathcal{X}/\mathcal{Y}) \geq d_{\min}(\mathcal{X})$ always holds. Thus our characterization of Proposition 11 by the coset distance is always equal to or tighter than that given by the existing approach with the minimum Hamming distance. In Example 14, since the minimum Hamming distance equals the coset distance for GRS codes, the derived resistance coincides with that of [16].

## 6. Resistance to Limited-Knowledge Adversary

This section characterizes the $(b,u)$-Byzantine resistance to the *limited-knowledge* adversary defined in Sect. 4.1.2. Recall that the adversary can observe stored messages and received queries by intruding $b+u$ servers and control them much like an active adversary considered for distributed storage systems [12, Section VII]. Thus the adversary's knowledge about messages and queries is restricted to what it can obtain from the observed servers. In the following, we shall demonstrate that by this limitation of the adversary's knowledge, the $(b,u)$-Byzantine resistance can exceed that to the omniscient case given as Proposition 11. To this end, we introduce a new protocol that allows storage servers to securely store and compute supplementary *hash matrices* with negligible overhead. This is inspired by the work of Pawar et al. [12] for distributed storage system, while it aims to directly retrieve the stored data itself unlike PIR schemes. To help readers gain a better understanding, this section begins from a toy example illustrating the protocol.

### 6.1 Toy Example

Consider a distributed storage composed of $n = 4$ servers that employ the scheme of Tajeddine et al. [16] as in Example 14. We fix $\dim C = 2$ and $\dim_q \mathcal{D}_{\mathsf{in}} = 1$. We then see $d_{\min}(\mathcal{M}_{\mathsf{out},q}/\mathcal{M}_{\mathsf{in},q}) = 2$ by the analysis in Example 14, where $\mathcal{M}_{\mathsf{out},q} \triangleq (C|\mathbb{F}_q \circ \mathcal{D}_{\mathsf{out}})_q$ and $\mathcal{M}_{\mathsf{in},q} \triangleq (C|\mathbb{F}_q \circ \mathcal{D}_{\mathsf{in}})_q$ are subspaces of $\mathbb{F}_q^n$. Suppose there is a Byzantine adversary with $b = 1$ and $u = 0$. Then, if the adversary is omniscient, the user may not be able to correctly retrieve the desired message by Corollary 13. In contrast, the $(1,0)$-Byzantine resistance is always guaranteed with arbitrarily high probability if the adversary is a *limited-knowledge* one defined in Sect. 4.1.2. We shall demonstrate this in the followings.

From now on, for two elements $a, b \in \mathbb{F}_{q^v}$, we denote by $\langle a,b\rangle_q \triangleq \lambda(a)\lambda(b)^{\mathrm{T}} \in \mathbb{F}_q$ an inner product of vectors $\lambda(a)$ and $\lambda(b)$ over $\mathbb{F}_q$, where $\lambda$ was defined in Sect. 2. Also for the sake of simplicity, we assume that the first server of $n = 4$ servers is intruded and under the control of the adversary.

#### 6.1.1 Storage Construction

First suppose that two messages are served, i.e., $m = 2$. Recall that $C|\mathbb{F}_q$ is a GRS code, i.e., an MDS code, with $\dim C|\mathbb{F}_q = \dim C$. Then $C$ is still MDS from $d_{\min}(C|\mathbb{F}_q) = d_{\min}(C)$ by Lemma 4. Original messages are encoded into $Y \in \mathbb{F}_{q^v}^{2\times 4}$ given as

$$Y = [Y_1, Y_2, Y_3, Y_4] = \begin{bmatrix} y_1^1 & y_2^1 & y_3^1 & y_4^1 \\ y_1^2 & y_2^2 & y_3^2 & y_4^2 \end{bmatrix},$$

where each row $y^i \triangleq [y_1^i, y_2^i, y_3^i, y_4^i]$ for $i \in [2]$ is a codeword of $C$, and each column $Y_j$ for $j \in [4]$ is stored at the $j$-th server. Furthermore, as with $y_j^i$, a *hash matrix* $H_j^i \in \mathbb{F}_q^{2\times 4}$ is computed for $i \in [2]$ and $j \in [4]$, defined as

$$H_j^i = [H_{j,1}^i, H_{j,2}^i, H_{j,3}^i, H_{j,4}^i] = \begin{bmatrix} h_{j,1}^{1,i} & h_{j,2}^{1,i} & h_{j,3}^{1,i} & h_{j,4}^{1,i} \\ h_{j,1}^{2,i} & h_{j,2}^{2,i} & h_{j,3}^{2,i} & h_{j,4}^{2,i} \end{bmatrix},$$

where $h_{b,d}^{a,c} \triangleq \langle y_b^a, y_d^c\rangle_q \in \mathbb{F}_q$ is a *pairwise hash*.

For the sake of simplicity, this example assumes that every $H_j^i$ is securely stored at the $j$-th server in such a way that the adversary can neither observe nor corrupt them, e.g., in an HSM-like (hardware security module) device of the server. Note that this assumption is much like ones given in existing works [12], [18]. We also note that the adversary can reproduce a part of hash matrices without accessing an HSM if $2 > b + u$ or more servers are observed. In Sect. 6.3, we will explain how this assumption is achieved via secure and resilient schemes for network coding [19] and distributed storage [12] when an HSM-like device is unavailable.

#### 6.1.2 Query and Response

The user generates a total query $G \in \mathcal{G}[i]$ given as

$$G = [G_1, G_2, G_3, G_4] = \begin{bmatrix} g_1^1 & g_2^1 & g_3^1 & g_4^1 \\ g_1^2 & g_2^2 & g_3^2 & g_4^2 \end{bmatrix} \in \mathbb{F}_q^{2\times 4},$$

and then the user sends each column $G_j$ to the $j$-th server.

**Table 2** Comparison between $\hat{\tau}_{j_2}^{j_1}$ and $\tau_{j_2}^{j_1}$ ($j_1, j_2 \in [4]$) for an erroneous total response $\hat{r} = [r_1 + e_1, r_2, r_3, r_4]$, where $\tau_{j_2}^{j_1} = \langle r_{j_1}, r_{j_2} \rangle_q$, ✓means $\hat{\tau}_{j_2}^{j_1} = \tau_{j_2}^{j_1}$ and ✗means $\hat{\tau}_{j_2}^{j_1} \neq \tau_{j_2}^{j_1}$.

| Validation of $\hat{\tau}_{j_2}^{j_1} = \tau_{j_2}^{j_1}$ | $j_2 = 1$ | 2 | 3 | 4 |
|---|---|---|---|---|
| $j_1 = 1$ | ✓ | ✗ | ✗ | ✗ |
| 2 | ✗ | ✓ | ✓ | ✓ |
| 3 | ✗ | ✓ | ✓ | ✓ |
| 4 | ✗ | ✓ | ✓ | ✓ |

We note that each row $g^i \triangleq [g_1^i, g_2^i, g_3^i, g_4^i]$ is a codeword of $\mathcal{D}_{\mathsf{in}}$ or $\mathcal{D}_{\mathsf{out}}$, and that $\mathcal{D}_{\mathsf{in}}$ and $\mathcal{D}_{\mathsf{out}}$ are MDS from Example 14.

Since we have assumed that the first server is corrupted by the adversary, the total response can be represented as

$$\hat{r} = [\hat{r}_1, \hat{r}_2, \hat{r}_3, \hat{r}_4] = [e_1 + r_1, r_2, r_3, r_4] \in \mathbb{F}_{q^v}^4, \tag{6}$$

where $r_j = G_j^{\mathrm{T}} Y_j = \sum_{i \in [2]} g_j^i y_j^i$ is an error-free response from the $j$-th server and $e_1 \in \mathbb{F}_{q^v}$ is an additive error at the first server. Note that the response polluted not only at the server but also on the communication channel can be represented in the form of $\hat{r}$ given by Eq. (6).

As well as responses, the $j$-th server also computes the *auxiliary response* $\Omega_j$ with its stored two hash matrices ($H_j^i \in \mathbb{F}_q^{2\times 4} : i \in [2]$) and received query $G_j \in \mathbb{F}_q^{2\times 1}$. $\Omega_j$ is denoted as a 2-tuple of 4-dimensional vectors as follows.

$$\Omega_j = \left(\omega_j^i = G_j^{\mathrm{T}} H_j^i : i \in [2]\right),$$

where $\omega_j^i = [\omega_{j,1}^i, \ldots, \omega_{j,4}^i] \in \mathbb{F}_q^4$ and $\omega_{j,l}^i = G_j^{\mathrm{T}} H_{j,l}^i$ for $l \in [4]$. Similarly to the assumption on securely-stored hash matrices, we assume that auxiliary responses are securely computed against the adversary. Namely, the adversary can neither observe nor corrupt the process and result of computation of auxiliary responses, e.g., by a processor in an HSM. In other words, we assume that the adversary can observe and corrupt only $Y_1$ and $G_1$ at the first server. The justification of this assumption will be given in Sect. 6.3.

### 6.1.3 Decoding Logic

The user receives the (corrupted) total response $\hat{r} = [\hat{r}_1, \ldots, \hat{r}_4]$ and auxiliary responses $\Omega_1, \ldots, \Omega_4$ from four servers for the total query $G$. Recall that the response $\hat{r}_1$ was assumed to be corrupted, and that the user is unaware of the location of the corrupted server. In the decoding logic, the user first attempts to detect the location of the corrupted server, and then identifies the coset containing the error-free $r$ by the coset-distance-based erasure decoding from $[\mathsf{null}, \hat{r}_2, \hat{r}_3, \hat{r}_4]$ excluding the corrupted $\hat{r}_1$. To these ends, the user computes *response hashes* $\hat{\tau}_{j_2}^{j_1} = \langle \hat{r}_{j_1}, \hat{r}_{j_2} \rangle_q \in \mathbb{F}_q$ for $j_1, j_2 \in [4]$ from the received $\hat{r}$ that may contain errors. On the other hand, the user also generates *error-free* response hashes $\tau_{j_2}^{j_1}$ from auxiliary responses $\Omega_1, \ldots, \Omega_4$ and the total query $G = [G_1, \ldots, G_4] \in \mathbb{F}_q^{2\times 4}$ by

$$\tau_{j_2}^{j_1} = G_{j_2}^{\mathrm{T}} \begin{bmatrix} \omega_{j_1,j_2}^1 \\ \omega_{j_1,j_2}^2 \end{bmatrix}.$$

Note that $G$ has been issued by the user itself. We see that an error-free $\tau_{j_2}^{j_1}$ is generated from the query $G_{j_2}$ to the $j_2$-th server and the auxiliary response $\Omega_{j_1}$ from $j_1$-th server. By comparing $\tau_{j_2}^{j_1}$'s and $\hat{\tau}_{j_2}^{j_1}$'s, the user composes a $4 \times 4$ table of comparison results, as presented by Table 2. It is easy to verify $\tau_{j_2}^{j_1} = \hat{\tau}_{j_2}^{j_1}$ when $r_{j_1} = \hat{r}_{j_1}$ and $r_{j_2} = \hat{r}_{j_2}$, which will be proven in Appendix for a general case. In Table 2, every response hash generated with $\hat{r}_1$ does not pass the comparison test except for that of $j_1 = j_2 = 1$. Note that $\hat{\tau}_1^1 = \tau_1^1$ holds (as the worst case) when $e_1$ satisfying $\langle e_1, r_1 \rangle_q = \langle e_1, e_1 \rangle_q = 0$ is chosen by the adversary.

The user selects a *trusted subset of response symbol indices* $\mathcal{T} \subset [4]$ from the obtained comparison table, i.e., Table 2. Recall that this toy example corresponds to Example 14 of $v$-degree parallelization. Thus we shall consider decoding in each sub-PIR scheme by letting $r[i]$ be the $i$-th row of the matrix form $\Lambda(r)$ of $r$ as explained in Sect. 3.2. Then, the user tries to identify the original coset $\mathcal{W}[i] \in \mathcal{M}_{\mathsf{out},q}/\mathcal{M}_{\mathsf{out},q}, r[i] \in \mathcal{W}[i], \forall i \in [v]$ from the subvector $[\hat{r}_j : j \in \mathcal{T}]$ with the coset distance. Here,

$$|\mathcal{T}| \geq n - d_{\mathsf{min}}(\mathcal{M}_{\mathsf{out},q}/\mathcal{M}_{\mathsf{in},q}) + 1 = 3, \tag{7}$$

must hold for successful decoding from Definition 1. Hence the user chooses $\mathcal{T}$ of size 3 from $\hat{r}_1, \ldots, \hat{r}_4$ that yields a 3×3 submatrix of Table 2 where all entries are "✓", i.e., $\hat{r}_2, \hat{r}_3, \hat{r}_4$. Then for every row of the matrix form of $[\mathsf{null}, \hat{r}_2, \hat{r}_3, \hat{r}_4]$, the coset-distance-based erasure decoding is employed over $\mathcal{M}_{\mathsf{out},q}/\mathcal{M}_{\mathsf{in},q}$, and the original coset containing $r[i]$ is eventually obtained for all $i \in [v]$. Therefore $(b, u)$-Byzantine resistance is attained for $b = 1$ and $u = 0$.

**Remark 15.** In this example, we have focused on the superposition of $v$ distinct decoding over $\mathcal{M}_{\mathsf{out},q}/\mathcal{M}_{\mathsf{out},q}$ similarly to Example 14. In a general case, i.e., $\dim C > \dim C|\mathbb{F}_q$, the user aims to identify the coset $\mathcal{W} \in \mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{out}}, r \in \mathcal{W}$ as described in Definition 10, and hence Eq. (7) is given with $d_{\mathsf{min}}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{out}})$ in such cases. Then we directly identify the coset from the subvector $[\hat{r}_j : j \in \mathcal{T}]$ over $\mathbb{F}_{q^v}$.

### 6.1.4 Error Analysis

As we assumed that the original response $r_1$ is altered by the adversary, we can write the polluted symbol as $\hat{r}_1 = r_1 + e_1$ for some error $e_1 \in \mathbb{F}_{q^v}$. Observe that a set $\hat{r}_{\mathcal{T}} \triangleq \{\hat{r}_j : j \in \mathcal{T}\}$ with any chosen $\mathcal{T}$ always contains at least two error-free symbols that are not observed by the adversary. This is because the adversary can observe and corrupt only $b + u = 1$ servers, and $|\mathcal{T}| = 3$ in this example. Considering the case that $\hat{r}_{\mathcal{T}}$ contains altered $\hat{r}_1$ and error-free $r_2, r_3$, it has to generate a consistent subtable of comparison test of size $3 \times 3$. This means that in order to pass the validation using the subtable generated from the auxiliary responses $\Omega_j$, the adversary must pick the error $e_1$ satisfying $\langle \hat{r}_1, r_2 \rangle_q = \langle r_1, r_2 \rangle_q$ and $\langle \hat{r}_1, r_3 \rangle_q = \langle r_1, r_3 \rangle_q$, i.e., $\langle e_1, r_2 \rangle_q = \langle e_1, r_3 \rangle_q = 0$ from $\langle a + b, c \rangle_q = \langle a, c \rangle_q + \langle b, c \rangle_q$ for $a, b, c \in \mathbb{F}_{q^v}$.

Also observe that the pollution of $r_1$ by the adversary

is independent of $r_2$. This is because $Y_2$ is independent from the observed $G_1$ and $Y_1$ due to the properties of $[4, 2]$ MDS codes $C$ and $C|\mathbb{F}_q$. Note that while $G_2$ can be determined from $G_1$, $r_2$ is still independent from the adversary's corruption since $r_2$ is an $\mathbb{F}_q$-linear combination of completely unknown symbols in $Y_2$. These mean that for the adversary, the number of candidate of $r_2$ cannot be decreased even if $Y_1$ and $G_1$ are revealed. Hence for any $e_1 \in \mathbb{F}_{q^v}$ chosen by the adversary, there exist $q^v$ equally likely choices of $r_2$. In other words, the knowledge on $Y_1, G_1$ is not useful at all to choose $e_1$ satisfying $\langle e_1, r_2\rangle_q = 0$ and $\langle e_1, r_3\rangle_q = 0$. Note that $q^{v-1}$ out of $q^v$ choices of $e_1$ are orthogonal to chosen $e_1$. This implies that two validations of $\tau_2^1 = \hat{\tau}_2^1$ and $\tau_3^1 = \hat{\tau}_3^1$ simultaneously succeed with probability

$$\begin{aligned}
&\Pr[\langle e_1, r_2\rangle_q = 0, \langle e_1, r_3\rangle_q = 0 | Y_1, G_1, e_1] \\
&= \Pr[\langle e_1, r_3\rangle_q = 0 | Y_1, G_1, e_1, \langle e_1, r_2\rangle_q = 0] \\
&\quad \Pr[\langle e_1, r_2\rangle_q = 0 | Y_1, G_1, e_1] \\
&\le \Pr[\langle e_1, r_2\rangle_q = 0 | Y_1, G_1, e_1] = \frac{q^{v-1}}{q^v} = \frac{1}{q}.
\end{aligned}$$

We thus conclude that the failure probability to detect polluted symbols is upper-bounded by $1/q$ that vanishes with increasing the field size $q$.

#### 6.1.5 Overhead Analysis

In this exemplary scenario, the user downloads auxiliary responses $\Omega_j$ consisting of $m = 2$ vectors in $\mathbb{F}_q^4$ for $j \in [4]$ in addition to the total response $r \in \mathbb{F}_{q^v}^4$. Hence the additional overhead of auxiliary responses is $\frac{32}{4v} = O(\frac{1}{v})$ per response symbol that goes to zero with increasing the degree $v$ of the field extension $\mathbb{F}_{q^v}$, i.e., the packet size. This is exactly the same as the upload cost, i.e., query size, ignored in most of modern PIR researches [4], [5], [9], [15], [16].

### 6.2 Characterization of the $(b, u)$-Byzantine Resistance

This subsection summarizes our result on the $(b, u)$-Byzantine resistance to the limited-knowledge adversary, which is guaranteed by our protocol illustrated by an example of the previous subsection. The detailed description for a general case will be presented in Appendix.

The following is the main theorem introducing the condition for the $(b, u)$-Byzantine resistance to the limited-knowledge adversary.

**Theorem 16.** Consider the PIR scheme in Definition 5, and let $\mathcal{M}_{\mathsf{out}} \triangleq C \circ \mathcal{D}_{\mathsf{out}}$ and $\mathcal{M}_{\mathsf{in}} \triangleq C \circ \mathcal{D}_{\mathsf{in}}$. Then the $(b, u)$-Byzantine resistance to the limited-knowledge adversary is guaranteed with arbitrarily high probability if

$$n - d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}) + 1 > b, \text{ and} \tag{8}$$

$$\min\left\{d_{\min}(C^\perp) - 1, d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}})\right\} > b + u. \tag{9}$$

*Proof Sketch.* Firstly, Eq. (8) is the condition to choose the trusted index set $\mathcal{T}$ always containing at least one index of an error-free response that is not observed by the adversary.

Secondly, observe that any $t$ columns of the generator matrix of $C$ are linearly independent if and only if $t \le d_{\min}(C^\perp) - 1$ from the property of dual codes [11]. Thus for $c = [c_1, \ldots, c_n] \in C$ and $\mathcal{I} \subseteq [n]$ of $|\mathcal{I}| < d_{\min}(C^\perp) - 1$, any $c_j$ for $j \in [n] \setminus \mathcal{I}$ is uniformly distributed over $\mathbb{F}_{q^v}$ conditioned on $(c_i : i \in \mathcal{I})$. This implies that a response $r_j$ is uniformly distributed over $\mathbb{F}_{q^v}$ conditioned on $(Y_i : i \in \mathcal{I})$ when $j \in [n] \setminus \mathcal{I}$ and $|\mathcal{I}| < d_{\min}(C^\perp) - 1$ even if all $G_i$'s are revealed to the adversary. Then there exist $q^{v-1}$ candidates in $\mathbb{F}_{q^v}$ that are orthogonal to $r_j$. Hence, the adversary can inject errors in $b$ responses that are consistent in pairwise hashes with a response from any non-intruded server only with probability $\frac{q^{v-1}}{q^v} = \frac{1}{q}$, which goes to zero with increasing $q$. Thus $b$ polluted responses can be detected by the approach of pairwise hashing as exemplarily illustrated in Sect. 6.1 if $b+u < d_{\min}(C^\perp)-1$, provided the trusted index set $\mathcal{T}$ contains at least one error-free response.

Finally we consider the decoding from the remaining $n - b - u$ symbols in $\hat{r}$, i.e., identifying the original coset $\mathcal{W} \in \mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}, r \in \mathcal{W}$. We immediately see that $\mathcal{W}$ is uniquely identified from error-free $n - b - u$ symbols in $\hat{r}$ when $d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}) - b - u > 0$ from Definition 1. Therefore, by taking the minimum of $d_{\min}(C^\perp) - 1$ given above and $d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}})$, we obtain Eq. (9). □

The detailed proof is given in Appendix via the description of our protocol for the general case.

Note that the $(b, u)$-Byzantine resistance is breached when the adversary is allowed to obtain some part of encoded messages stored in non-intruded servers, while queries to non-intruded servers is allowed to be revealed. Recall that each symbol $r_j$ in $r$ is $\mathbb{F}_q$-linear combination of $\mathbb{F}_{q^v}$ symbols, where coefficients in $\mathbb{F}_q$ is symbols in a query $G_j \in \mathbb{F}_q^{m\times 1}$. Assume that the adversary knows $Y_j = [y_j^1, \ldots, y_j^m]^{\mathrm{T}} \in \mathbb{F}_{q^v}^{m\times 1}$ stored at a non-intruded $j$-th server. Then, without knowing the query $G_j = [g_j^1, \ldots, g_j^m]^{\mathrm{T}}$ to the $j$-th server, the adversary can compute $e \in \mathbb{F}_{q^v}$ that satisfies $\langle e, r_j\rangle_q = \langle e, G_j^{\mathrm{T}} Y_j\rangle_q = 0$ since we have

$$\langle e, G_j^{\mathrm{T}} Y_j\rangle_q = \langle e, \sum_{l\in[m]} g_j^l y_j^l\rangle_q = \sum_{l\in[m]} g_j^l \underbrace{\langle e, y_j^l\rangle_q}_{\text{Adversary can control}},$$

and $e$, s.t., $\langle e, y_j^l\rangle_q = 0, \forall l \in [m]$, could be chosen. Therefore we emphasize that the privacy on the query in PIR schemes is independent of the $(b, u)$-Byzantine resistance.

As the conclusion of this section, we finally introduce the following corollary of Theorem 16 for the special case where $\dim C = \dim C|\mathbb{F}_q$, i.e., $v$ distinct PIR processes are parallelized much like Corollary 13.

**Corollary 17.** Consider the PIR scheme in Definition 5. Set the storage code $C$ of $\dim C = \dim C|\mathbb{F}_q$, and employ the decoding per each sub-PIR scheme in the $v$-degree parallelization described in Sect. 3.2.1. Let $\mathcal{M}_{\mathsf{out},q} \triangleq (C|\mathbb{F}_q \circ \mathcal{D}_{\mathsf{out}})_q$ and $\mathcal{M}_{\mathsf{in},q} \triangleq (C|\mathbb{F}_q \circ \mathcal{D}_{\mathsf{in}})_q$ be subspaces of $\mathbb{F}_q^n$. Then the $(b, u)$-Byzantine resistance to the limited-knowledge adversary is

guaranteed with arbitrarily high probability if

$$n - d_{\min}(\mathcal{M}_{\mathrm{out},q}/\mathcal{M}_{\mathrm{in},q}) + 1 > b, \text{ and}$$
$$\min\left\{d_{\min}(C^{\perp}) - 1, d_{\min}(\mathcal{M}_{\mathrm{out},q}/\mathcal{M}_{\mathrm{in},q})\right\} > b + u.$$

### 6.3 Secure Environment for Storage and Computation

As the conclusion of this section, we justify the assumption that some extra information of fixed size, i.e., hash matrices and auxiliary responses, are stored, computed and retrieved securely from the adversary's observation and pollution. In Sect. 6.1 (and later in Appendix for a general case), we have assumed that secure channels and secure computing processors, e.g., HSM-like devices, are equipped to all servers and that the user is allowed to obtain the non-observed and error-free extra information. This assumption, however, might be unachievable, i.e., a part of the information may be leaked and the user might receive polluted information as discussed in [12].

In such an adversarial environment, we can employ the encoding method of *reliable and secure transmission/storage* [19], [12, Appendix C2] to store and download hashes themselves. Namely, all hash matrices are encoded and stored with the method of [12], [19] in storage servers. The user obtains all original hash matrices instead of auxiliary responses, and computes auxiliary responses by the user itself on behalf of servers. Then, under the condition of $n > 2(b + u)$, the user can prevent the adversary intruding $b + u$ servers from obtaining any information about hash matrices from the analysis in [12], [19]. Simultaneously, the user can obtain the complete and error-free hash matrices by allowing vanishing error probability that goes to zero with increasing $q$. Thus the extra information can still be securely stored, computed and retrieved.

For the privacy, downloading *all* of hash matrices obviously leaks no information on the user's demand. Also for the size of downloaded data, the total size of encoded hash matrices in the storage [12], [19] is always constant for fixed $m$ and $n$ and independent from the packet size $v$. Therefore much like auxiliary responses, the overhead of retrieving hash matrices to the download cost can still be vanishing with increasing $v$.

## 7. Concluding Remarks

This paper investigated the resistance of PIR schemes to Byzantine adversaries in the retrieval scenario from $n$ coded storage servers. We considered two types of Byzantine adversaries polluting $b$ responses and erasing $u$ responses to a user's query. One is the classic omniscient type that has the full knowledge on $n$ servers, as considered in existing researches [5], [16]. The other is newly introduced type in this paper, called the limited-knowledge adversary, that has the information of only $b + u$ servers controlled by the adversary itself. By redefining the decoding in PIR schemes as the identification of a coset, we revealed that for both two types, the Byzantine resistance of PIR schemes based on arbitrary linear codes is expressed in terms of the coset distance of employed codes. Furthermore, we demonstrated that by limiting the Byzantine adversary's knowledge, the resistance of a PIR scheme could be improved much like the network coding [6], [19] and distributed storage coding [12]. We believe that this paper provided useful analytical tools for existing and future PIR schemes via the reformulation.

A natural avenue of this work is to investigate a 'tighter' condition on the resistance than Proposition 11 and Theorem 16. As mentioned in Remark 12, Since $|\{v \circ w : v \in \mathcal{X}, w \in \mathcal{Y}\}| \le |\mathcal{X} \circ \mathcal{Y}|$ for subspaces $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{F}_{q^v}^n$, the domain considered in the characterization is larger than the one where the actual total response exists. Hence pruning such extra space in the Hadamard product of subspaces could be an interesting research topic to obtain a tighter condition.

## Acknowledgments

The authors would like to thank Dr. Daishi Kondo at Osaka Prefecture University for valuable discussions. This work was supported in part by University of Hyogo special research grant for young researchers, JSPS KAKENHI Grant Number JP20K23329, and KDDI Research collaborative research project.

### References

[1] T.H. Chan, S.-W. Ho, and H. Yamamoto, "Private information retrieval for coded storage," Proc. IEEE ISIT 2015, Hong Kong, China, pp.2842–2846, June 2015.
[2] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," Proc. FOCS 1995, Milwaukee, WI, USA, pp.41–50, Oct. 1995.
[3] I.M. Duursma and S. Park, "Coset bounds for algebraic geometric codes," Finite Fields Th. App., vol.16, no.1, pp.36–55, Jan. 2010.
[4] R. Freij-Hollanti, O.W. Gnilke, C. Hollanti, and D.A. Karpuk, "Private information retrieval from coded databases with colluding servers," SIAM J. Appl. Algebra Geometry, vol.1, no.1, pp.647–664, Jan. 2017.
[5] L. Holzbaur, R. Freij-Hollanti, and C. Hollanti, "Towards the capacity of private information retrieval from coded and colluding servers," https://arxiv.org/abs/1903.12552, March 2020.
[6] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, M. Médard, and M. Effros, "Resilient network coding in the presence of Byzantine adversaries," IEEE Trans. Inf. Theory, vol.54, no.6, pp.2596–2603, June 2008.
[7] R. Koetter and F.R. Kschischang, "Coding for errors and erasures in random network coding," IEEE Trans. Inf. Theory, vol.54, no.8, pp.3579–3591, Aug. 2008.
[8] J. Kurihara and T. Nakamura, "On the resistance to Byzantine and unresponsive servers in code-based PIR schemes," IEICE ComEx, vol.9, no.7, pp.342–347, July 2020.
[9] L. Li, M. Militzer, and A. Datta, "rPIR: Ramp secret sharing-based communication-efficient private information retrieval," Int. J. Inf. Secur., vol.16, no.6, pp.603–625, Nov. 2017.
[10] Y. Luo, C. Mitrpant, A.J. Han Vinck, and K. Chen, "Some new characters on the wire-tap channel of type II," IEEE Trans. Inf. Theory, vol.51, no.3, pp.1222–1229, March 2005.
[11] F.J. MacWilliams and N.J.A. Sloane, The Theory of Error-Correcting Codes, North-Holland Mathematical Library, 1977.

[12] S. Pawar, S.Y. El Rouayheb, and K. Ramchandran, "Securing dynamic distributed storage systems against eavesdropping and adversarial attacks," IEEE Trans. Inf. Theory, vol.57, no.10, pp.6734–6753, Sept. 2010.
[13] D. Silva, F.R. Kschischang, and R. Koetter, "A rank-metric approach to error control in random network coding," IEEE Trans. Inf. Theory, vol.54, no.9, pp.3951–3967, Sept. 2008.
[14] H. Stichtenoth, "On the dimension of subfield subcodes," IEEE Trans. Inf. Theory, vol.36, no.1, pp.90–93, Jan. 1990.
[15] R. Tajeddine, O.W. Gnilke, and S. El Rouayheb, "Private information retrieval from MDS coded data in distributed storage systems," IEEE Trans. Inf. Theory, vol.64, no.11, pp.7081–7093, Nov. 2018.
[16] R. Tajeddine, O.W. Gnilke, D. Karpuk, R. Freij-Hollanti, and C. Hollanti, "Private information retrieval from coded storage systems with colluding, Byzantine, and unresponsive servers," IEEE Trans. Inf. Theory, vol.65, no.6, pp.3898–3906, June 2019.
[17] Q. Wang and M. Skoglund, "Symmetric private information retrieval from MDS coded distributed storage with non-colluding and colluding servers," IEEE Trans. Inf. Theory, vol.65, no.8, pp.5160–5175, Aug. 2019.
[18] Q. Wang, H. Sun, and M. Skoglund, "The $\epsilon$-error capacity of symmetric PIR with Byzantine adversaries," Proc. IEEE ITW 2018, Guangzhou, China, Nov. 2019.
[19] H. Yao, D. Silva, S. Jaggi, and M. Langberg, "Network codes resilient to jamming and eavesdropping," IEEE/ACM Trans. Netw., vol.22, no.6, pp.1978–1987, Dec. 2014.

## Appendix: Proof of Theorem 16

Here we prove Theorem 16 via explaining the protocol generalized from the toy example in Sect. 6.1. Suppose that the storage code $C \subseteq \mathbb{F}_{q^v}^n$ is employed, and that $n$ servers store $m$ encoded messages as given in Sect. 5. Also suppose that there exists a limited-knowledge Byzantine adversary intruding $b+u$ of $n$ servers. We denote by $\mathcal{I} \subseteq [n]$, $|\mathcal{I}| = b+u$ as the index set of intruded servers. Define subsets $\mathcal{I}_b \subseteq \mathcal{I}$ of $|\mathcal{I}_b| = b$ and $\mathcal{I}_u \subseteq \mathcal{I}$ of $|\mathcal{I}_u| = u$ with $\mathcal{I}_b \cap \mathcal{I}_u = \{\}$. Let the $j$-th server for $j \in \mathcal{I}_b$ return an erroneous response, and the one for $j \in \mathcal{I}_u$ return nothing. In the following, we denote Hadamard product of $C, \mathcal{D}_{\mathsf{in}}\mathcal{D}_{\mathsf{out}}$ by $\mathcal{M}_{\mathsf{out}} \triangleq C \circ \mathcal{D}_{\mathsf{out}}$ and $\mathcal{M}_{\mathsf{in}} \triangleq C \circ \mathcal{D}_{\mathsf{in}}$.

### A.1 Storage Construction

For symbols $y_b^a, y_d^c \in \mathbb{F}_{q^v}$ of codewords in $C$, let $h_{b,d}^{a,c} \triangleq \langle y_b^a, y_d^c \rangle_q \in \mathbb{F}_q$ be a pairwise hash. The user computes hash matrices $H_j^i$ in addition to encoded messages, given as

$$H_j^i \triangleq [H_{j,1}^i, \ldots, H_{j,n}^i] = \begin{bmatrix} h_{j,1}^{1,i} & h_{j,2}^{1,i} & \ldots & h_{j,n}^{1,i} \\ h_{j,1}^{2,i} & h_{j,2}^{2,i} & \ldots & h_{j,n}^{2,i} \\ \vdots & \vdots & \ddots & \vdots \\ h_{j,1}^{m,i} & h_{j,2}^{m,i} & \ldots & h_{j,n}^{m,i} \end{bmatrix} \in \mathbb{F}_q^{m \times n}, \quad (\mathrm{A}\cdot 1)$$

for $i \in [m]$ and $j \in [n]$. We assume that $H_j^i$ is securely stored at the $j$-th server, e.g., in an HSM-like device of a server, and that the adversary can neither observe nor corrupt any hash matrix. We have justified this assumption in Sect. 6.3.

### A.2 Query and Response

The user generates a total query $G \in \mathbb{F}_q^{m \times n}$ and sends its $j$-th column $G_j \in \mathbb{F}_q^{m \times 1}$ as a query to the $j$-th server, as given in Sect. 3. Then the $j$-th server that is non-intruded, i.e., $j \in [n] \setminus \mathcal{I}$, responds to $G_j$ with a response $r_j = G_j^{\mathrm{T}} Y_j \in \mathbb{F}_{q^v}$. The total response received by the user can be written by

$$\hat{r} = \begin{bmatrix} & \hat{r}_j = r_j \text{ for } j \in [n] \setminus \mathcal{I}, \\ \hat{r}_j : & \hat{r}_j = e_j + r_j, e_j \in \mathbb{F}_{q^v} \text{ for } j \in \mathcal{I}_b, \\ & \hat{r}_j = \mathsf{null} \text{ for } j \in \mathcal{I}_u, \end{bmatrix},$$

where $e_j$ is an additive error. In addition to responses, the user obtains an auxiliary response $\Omega_j$ from the $j$-th server, defined as an $m$-tuple:

$$\Omega_j \triangleq \left( \omega_j^i \triangleq G_j^{\mathrm{T}} H_j^i \in \mathbb{F}_q^n : i \in [m] \right), \quad (\mathrm{A}\cdot 2)$$

where $\omega_j^i = [\omega_{j,1}^i, \ldots, \omega_{j,n}^i]$ and $\omega_{j,l}^i = G_j^{\mathrm{T}} H_{j,l}^i$ for $l \in [n]$ by Eq. (A· 1). For the sake of simplicity, we assume that $\Omega_j$ is computed at the $j$-th server and correctly downloaded by the user via a secure channel that the adversary cannot observe, unlike $G_j$, $Y_j$ and $\hat{r}_j$. We have justified this assumption as well as the secure storage of hash matrices in Sect. 6.3.

### A.3 Decoding Logic

The user receives $\hat{r}$ and $\Omega_j$ for $\forall j \in [n]$. We should note that the user has no knowledge on locations of intruded servers, i.e., $\mathcal{I}$, and cannot recognize altered responses $\hat{r}_j$'s ($j \in \mathcal{I}_b$) only by checking the corrupted total response $\hat{r}$. Thus the user leverages the auxiliary responses $\Omega_j$'s defined as Eq. (A· 2) in order to pick up error-free responses $\hat{r}_j$'s $j \in [n] \setminus \mathcal{I}$ in $\hat{r}$. The user first computes response hashes $\hat{\tau}_{j_2}^{j_1} \triangleq \langle \hat{r}_{j_1}, \hat{r}_{j_2} \rangle_q \in \mathbb{F}_q$ for $j_1, j_2 \in [n] \setminus \mathcal{I}_u$, where the user, of course, is aware of locations of erasure symbols. The user also generates the error-free version of response hashes $\tau_{j_2}^{j_1} \in \mathbb{F}_q$ from auxiliary responses $\Omega_j$'s and queries $G_j$'s issued by the user itself as follows.

$$\tau_{j_2}^{j_1} \triangleq G_{j_2}^{\mathrm{T}} \begin{bmatrix} \omega_{j_1,j_2}^1 \\ \vdots \\ \omega_{j_1,j_2}^m \end{bmatrix} = \sum_{l \in [m]} g_{j_2}^l \omega_{j_1,j_2}^l, \quad (\mathrm{A}\cdot 3)$$

for $j_1, j_2 \in [n] \setminus \mathcal{I}_u$, where $\omega_{j_1,j_2}^i \in \mathbb{F}_q$ is the $j_2$-th element of $\omega_{j_1}^i$ in the $m$-tuple $\Omega_{j_1}$ defined by Eq. (A· 2). We can easily verify $\tau_{j_2}^{j_1} = \hat{\tau}_{j_2}^{j_1}$ when $r_{j_1} = \hat{r}_{j_1}$ and $r_{j_2} = \hat{r}_{j_2}$ as follows.

$$\begin{aligned} &\langle r_{j_1}, r_{j_2} \rangle_q \\ &= \lambda(r_{j_1}) \lambda(r_{j_2})^{\mathrm{T}} = \lambda(G_{j_1}^{\mathrm{T}} Y_{j_1}) \lambda(G_{j_2}^{\mathrm{T}} Y_{j_2})^{\mathrm{T}} \\ &= \underbrace{\lambda\left( \sum_{i \in [m]} g_{j_1}^i y_{j_1}^i \right)}_{= \sum_{i \in [m]} g_{j_1}^i \lambda(y_{j_1}^i) \text{ by Eq. (1)}} \underbrace{\lambda\left( \sum_{i \in [m]} g_{j_1}^i y_{j_1}^i \right)}_{= \sum_{i \in [m]} g_{j_2}^i \lambda(y_{j_2}^i) \text{ by Eq. (1)}}^{\mathrm{T}} \end{aligned}$$

$$= \sum_{s,t\in[m]} g^s_{j_1} g^t_{j_2} \underbrace{\lambda(y^s_{j_1})\lambda(y^t_{j_2})^{\mathrm{T}}}_{=\langle y^s_{j_1}, y^t_{j_2}\rangle_q = h^{s,t}_{j_1,j_2}}$$

$$= \sum_{t\in[m]} g^t_{j_2} \underbrace{\sum_{s\in[m]} g^s_{j_1} h^{s,t}_{j_1,j_2}}_{=G^{\mathrm{T}}_{j_1} H^t_{j_1,j_2} = \omega^t_{j_1,j_2} \text{ by Eq. (A·2)}} = \tau^{j_1}_{j_2},$$

where the final equality is given by Eq. (A·3). Hence by checking whether $\tau^{j_1}_{j_2} = \hat{\tau}^{j_1}_{j_2}$ holds or not, the user will identify the locations of erroneous responses.

In order to find $b$ erroneous symbols in $\hat{r}$, the user secondly composes the $(n-u)\times(n-u)$ comparison table verifying if $\tau^{j_1}_{j_2} = \hat{\tau}^{j_1}_{j_2}$, indexed with $j_1, j_2 \in [n] \setminus \mathcal{I}_u$, similarly to Table 2 in Sect. 6.1. The user selects the trusted subset of response symbol indices $\mathcal{T} \subset [n] \setminus \mathcal{I}_u$ in such a way that $|\mathcal{T}|\times|\mathcal{T}|$ subtable of the comparison table with all its entries pass the validation test of $\tau^{j_1}_{j_2} = \hat{\tau}^{j_1}_{j_2}$. Then supposing that $\hat{r}_j$ is error-free for $\forall j \in \mathcal{T}$, i.e., $\hat{r}_j = r_j$, the user executes the coset-distance-based decoding on $[\hat{r}_j : j \in \mathcal{T}]$ in order to uniquely determine the original coset $\mathcal{W} \in \mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}$, $r \in \mathcal{W}$. To achieve this, the cardinality of $\mathcal{T}$ must be

$$|\mathcal{T}| \geq n - d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}) + 1, \tag{A·4}$$

for successful decoding by Definition 1. Thus in Theorem 16,

$$d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}) \geq n - |\mathcal{T}| + 1 > b + u, \tag{A·5}$$

is a sufficient condition of the existence of such $\mathcal{T} \subset [n]$ of $\hat{r}_j = r_j, \forall j \in \mathcal{T}$. The next subsection discusses the failure probability of this decoding, i.e., the probability of picking up indices of erroneous symbols in $\mathcal{T}$, and also shows that it vanishes with increasing the field size $q$.

### A.4 Error Analysis

The user may fail to decode the total response only if the selected $\mathcal{T}$ contains at least one index of an erroneous symbol. On the other hand, we see that by the lower bound of $|\mathcal{T}|$ given as Eq. (A·4), $n - d_{\min}(\mathcal{M}_{\mathsf{out}}/\mathcal{M}_{\mathsf{in}}) + 1 > b$ is the sufficient condition that the chosen $\mathcal{T}$ always contains at least one index of an error-free symbol, posed as Eq. (8). We then consider the worst case scenario where only one index $j \in [n]$ of an error-free symbol $\hat{r}_j$ is contained in the chosen $\mathcal{T}$, i.e., $\hat{r}_j = r_j$, and others in $\mathcal{T}$ are erroneous. For this $\mathcal{T}$ to be chosen, it has to generate a consistent comparison subtable of size $|\mathcal{T}|\times|\mathcal{T}|$, which means the adversary has to generate $\hat{r}_i$'s $\forall i \in \mathcal{T} \setminus \{j\}$ satisfying $\langle \hat{r}_i, r_j\rangle_q = \langle r_i, r_j\rangle_q$, i.e, $\hat{\tau}^i_j = \tau^i_j$ from $\hat{\tau}^i_j = \langle \hat{r}_i, r_j\rangle_q$ and $\tau^i_j = \langle r_i, r_j\rangle_q$ as shown in the previous subsection. Letting $\hat{r}_i = r_i + e_i$ with $e_i \in \mathbb{F}_{q^v}$, this means that $e_i$ must be $\langle e_i, r_j\rangle_q = 0$ for $\forall i \in \mathcal{T} \setminus \{j\}$.

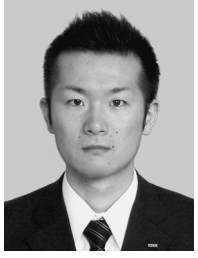
Next we compute the probability of such event. Recall that any $t$ columns of the generator matrix of $C$ are linearly independent if and only if $t \leq d_{\min}(C^\perp) - 1$ from the property of dual codes [11]. Thus for $c = [c_1, \ldots, c_n] \in C \subseteq \mathbb{F}^n_{q^v}$ and the index set of intruded servers $\mathcal{I} \supseteq \mathcal{T} \setminus \{j\}$ of $|\mathcal{I}| < d_{\min}(C^\perp) - 1$, a symbol $c_j$ is uniformly distributed over $\mathbb{F}_{q^v}$ conditioned on $c_{\mathcal{I}} = [c_i : i \in \mathcal{I}]$. We then also see that for $c = [c_1, \ldots, c_n], d = [d_1, \ldots, d_n] \in C$, their linear combination $\alpha c_j + \beta d_j$ is uniformly distributed over $\mathbb{F}_{q^v}$ conditioned on observed symbols $((c_i, d_i) : i \in \mathcal{I})$ and coefficients $\alpha, \beta \in \mathbb{F}_q$. These imply that when $|\mathcal{I}| < d_{\min}(C^\perp) - 1$, the error-free response $r_j \in \mathbb{F}_{q^v}$ is uniformly distributed over $\mathbb{F}_{q^v}$ conditioned on $(Y_i : i \in \mathcal{I})$ even if all queries $G_i$'s, i.e., the total query $G$, are revealed to the adversary. Then there exist $q^{v-1}$ candidates of $e_i$'s in $\mathbb{F}_{q^v}$ that are orthogonal to $r_j$, i.e., $\langle e_i, r_j\rangle_q = 0$. Hence, the adversary can inject errors in $b$ responses that are consistent in pairwise hashes with an error-free response $r_j$ only with probability $\frac{q^{v-1}}{q^v} = \frac{1}{q}$, which goes to zero with increasing $q$. Thus, when the number of intruded servers $|\mathcal{I}|$ satisfies

$$d_{\min}(C^\perp) - 1 > |\mathcal{I}| = b + u, \tag{A·6}$$

$b$ corrupted responses can always be detected with an arbitrary high probability as long as Eq. (8) holds. Therefore we finally obtain the condition of Eq. (9) in Theorem 16 by combining Eq. (A·5) and Eq. (A·6).

### A.5 Overhead Analysis

The user downloads auxiliary responses $\Omega_j$'s for $j \in [n]$ in addition to the total response $r \in \mathbb{F}^n_{q^v}$, where $\Omega_j$ consists of $m$ vectors in $\mathbb{F}^n_q$. Hence, considering that $m$ and $n$ are constant, the additional overhead due to auxiliary responses is $\frac{mn^2}{vn} = O(\frac{1}{v})$ per response symbol that goes to zero with increasing the degree $v$ of the field extension $\mathbb{F}_{q^v}$, i.e., the packet size. Therefore asymptotically in the packet size, the protocol satisfies Theorem 16.



**Jun Kurihara** received degrees of the B.E., M.E. and Ph.D. in Engineering from Tokyo Institute of Technology in 2004, 2006 and 2012, respectively. From 2006 to 2017, he was with KDDI Corp. and KDDI R&D Labs., Inc. as a strategic planner and a researcher. He is currently with Graduate School of Applied Informatics, University of Hyogo as an associate professor, and with Zettant Inc. as a principal researcher. He was a visiting researcher at Palo Alto Research Center (PARC), CA, USA from 2013 to 2014. His research interests include coding theory and networking architecture. He received the Best Paper Award from IEICE in 2014.

**Toru Nakamura** received degrees of the B.E., M.E. and Ph.D. in Engineering from Kyushu University, in 2006, 2008 and 2011, respectively. In 2011, he joined KDDI and in the same year he moved to KDDI R&D Laboratories, Inc. (currently renamed KDDI Research, Inc.). In 2018, he moved to Advanced Telecommunications Research Institute International (ATR). Since 2020, he is a researcher in KDDI Research, Inc. again. He received CSS2016 SPT Best Paper Award. His current research interests include security, privacy, and trust, especially privacy enhanced technology and analysis of privacy attitudes. He is a member of IEICE and IPSJ.

**Ryu Watanabe** received the B.E. and M.E degrees from University of Tokyo in 1997 and 1999, respectively. He joined KDD Corp. (currently KDDI Corp.) and moved to KDD R&D Laboratories, Inc. (currently KDDI Research, Inc.) in 1999. He is currently a research manager at Cyber Security Laboratory in KDDI Research, Inc. His research interests include security, privacy and identity management in networking. He received DICOMO Paper Award from IPSJ in 2011. He is a member of IEICE and IPSJ.