# Two-Level Private Information Retrieval

Ruida Zhou*, Chao Tian*, Hua Sun†, James S. Plank‡
*Department of Electrical and Computer Engineering, Texas A&M University, College Station TX
†Department of Electrical Engineering, University of North Texas, Denton TX
‡Department of Electrical Engineering and Computer Science, University of Tennessee Knoxville, Knoxville TN

***Abstract*—In the conventional robust $T$-colluding private information retrieval (PIR) system, the user needs to retrieve one of the possible messages while keeping the identity of the requested message private from any $T$ colluding servers. Motivated by the possible heterogeneous privacy requirements for different messages, we consider the $(N, T_1 : K_1, T_2 : K_2)$ two-level PIR system, where $K_1$ messages need to be retrieved privately against $T_1$ colluding servers, and all the messages need to be retrieved privately against $T_2$ colluding servers where $T_2 \leq T_1$. We obtain a lower bound to the capacity by proposing a novel coding scheme, namely the non-uniform successive cancellation scheme. A capacity upper bound is also derived. The gap between the upper bound and the lower bound is analyzed, and shown to vanish when $T_1 = T_2$.**

## I. Introduction

Capacity characterizations of the canonical private information retrieval (PIR) system and its variants have drawn considerable attention recently in the information and coding theory community, for which novel code constructions and impossibility results have been discovered.

In the canonical PIR model, user privacy needs to be preserved during message retrieval from replicated servers, i.e., the identity of the desired message should not be revealed to any single server. Specifically, the user is required to retrieve one of the $K$ messages from $N$ servers, each of which stores a copy of $K$ messages, such that the identity of the desired message is not revealed to any single server. In the PIR capacity characterization problem, the goal is to identify the minimum download cost, i.e., the minimum amount of download per-bit of the desired message, the inverse of which is referred to as the capacity of PIR. The PIR capacity was characterized in [1]. The extension to provide privacy against any $T$-colluding servers was considered in [2]. Various generalizations to allow homogeneity among users [3], [4], constraints on the message length [5], [6], and to use coded or uncoded storage [7]–[19] have been considered.

The main motivation of this work is a crucial aspect that has not been previously addressed in the literature - the heterogeneity of the privacy constraints on the messages. That is, in all existing works, each message is required to be equally private in the sense that any single server [1], or any colluding set of $T$ servers [2], is completely ignorant of the desired message identity. However, the sensitivities of different types of information are commonly different in practice. For example, politically biased messages, which may reflect the user's political tendency, can be more sensitive than messages of cooking recipes. The user may ask for a "higher" privacy level when requesting such politically biased messages than those of cooking recipes. Therefore, a new and more general PIR system is duly needed to take into account this consideration.

We formulate the problem of multilevel private information retrieval problem. Specifically, the *privacy level* of a message set is defined as the maximum allowed number of colluding servers that the identity of a desired message is kept private among that message set. We focus on the two-level PIR system, where some $K_1$ messages out of the $K$ messages have a higher privacy level of $T_1$, i.e., any colluding set of $T_1$ servers do not learn anything about which one of the $K_1$ messages is desired, while all the $K$ messages together have a lower privacy level of $T_2$, i.e., any colluding set of $T_2$ servers do not learn anything about which one of the $K$ messages is desired.

A naive approach which can be used as a baseline, is to treat the system as if it were a homogeneous $T_1$-colluding private information retrieval system. However, the crux of the two-level PIR hinges on how to leverage the less stringent privacy requirement for some messages. Towards this end, we propose a general scheme that can outperform the naive baseline scheme by treating the two sets of messages with distinct privacy levels differently. A capacity lower bound is also derived, and the gap between the lower bound and the upper bound is analyzed.

**Notations:** We adopt the notation $i : j \triangleq \{i, i+1, \ldots, j-1, j\}$. Denote vector $a_{\mathcal{N}} \triangleq (a_i)_{i\in\mathcal{N}}$ for any sequence $(a_1, a_2, \ldots)$ and $\mathcal{N} \subset \mathbb{N}$. We use $X \sim Y$ to indicate the random variables $X$ and $Y$ following an identical distribution. For any matrix $A[:, :]$, the first coordinate is for row indices and the second coordinate is for column indices.

## II. Problem Formulation

There are a total of $K$ mutually independent messages $W_{1:K} = (W_1, W_2, \ldots, W_K)$ in the system. Each message is uniformly distributed over $\mathbb{F}_q^L$, where $\mathbb{F}_q$ is a large enough finite field and $L$ is the number of $q$-ary symbols in the message (i.e., the message length). This is equivalent to

$$H(W_1) = H(W_2) = \cdots = H(W_K) = L, \tag{1}$$

$$H(W_{1:K}) = KL, \tag{2}$$

where (and in the rest of this work) we take base-$q$ logarithm for simplicity. There are $N$ servers in the system, each of which stores a copy of all the $K$ messages. Let $k^* \in 1 : K$

be the identity of the desired message. The process to retrieve message $W_{k^*}$, for any $k^* \in 1:K$, involves three steps:

1. (Query) The user sends a randomized query $Q_n^{[k^*]}$ to server $n$ for each $n \in 1:N$;
2. (Answer) Each server $n$, where $n \in 1:N$, returns an answer $A_n^{[k^*]}$ to the user;
3. (Recovery) The user recovers the message as $\hat{W}_{k^*}$, using the queries $Q_{1:N}^{[k^*]}$ to all the servers and the answers $A_{1:N}^{[k^*]}$ from all the servers.

Denote the set of all possible queries sent to server $n$ as $\mathcal{Q}_n$. $Q_n^{[k^*]} \in \mathcal{Q}_n$ is a random variable, whose superscript $[k^*]$ indicates that the query is for retrieving message $W_{k^*}$. The user has no knowledge of $W_{1:K}$, and thus the queries are independent of the messages

$$I(Q_{1:N}^{[k^*]}; W_{1:K}) = 0, \quad \forall k^* \in 1:K. \tag{3}$$

Each symbol of the answer $A_n^{[k^*]}$, the answer to the query $Q_n^{[k^*]}$, is a sequence of symbols in $\mathbb{F}_q$; denote the number of symbols of $A_n^{[k^*]}$ as $\ell_n^{[k^*]}$. The answer $A_n^{[k^*]}$ is a deterministic function of the query $Q_n^{[k^*]}$ and the messages $W_{1:K}$, that is

$$H(A_n^{[k^*]} | Q_n^{[k^*]}, W_{1:K}) = 0, \quad \forall k^* \in 1:K,\ n \in 1:N. \tag{4}$$

The recovered message $\hat{W}_{k^*}$ depends on the queries $Q_{1:N}^{[k^*]}$ as well as the answers $A_{1:N}^{[k^*]}$, that is

$$H(\hat{W}_{k^*} | A_{1:N}^{[k^*]}, Q_{1:N}^{[k^*]}) = 0, \quad \forall k^* \in 1:K. \tag{5}$$

The message should be retrieved correctly, i.e., $W_{k^*} = \hat{W}_{k^*}$ for all $k^* \in 1:K$. Additionally, the system has certain privacy requirements. To measure user privacy when querying for any message in a certain set of messages, we first introduce the definition of *privacy level*.

**Definition 1** (Privacy level)**.** *The queries of a scheme have privacy level $T$ for a set of messages $W_{\mathcal{S}}$, where $\mathcal{S} \subseteq 1:K$, if for any $\mathcal{T} \subseteq 1:N$ with $|\mathcal{T}| = T$, for retrieving any message in $W_{\mathcal{S}}$, the queries to the servers in $\mathcal{T}$ have the same joint distribution, i.e.,*

$$Q_{\mathcal{T}}^{[k]} \sim Q_{\mathcal{T}}^{[k']}, \qquad \forall k, k' \in \mathcal{S}. \tag{6}$$

The notion of privacy level has the following operational meaning: if $W_{\mathcal{S}}$ has privacy level $T$, then when one of the messages in $W_{\mathcal{S}}$ is retrieved, even if any $T$ of the $N$ servers collude, the identity of the requested message in $W_{\mathcal{S}}$ remains private, however these colluding servers may be able to infer that the requested message is in the set $W_{\mathcal{S}}$. It is straightforward to verify that the set of messages with higher privacy level automatically has lower privacy levels. In addition, when the set $\mathcal{S}$ is a singleton, if $T$ servers can infer the desired message is in $W_{\mathcal{S}}$, the identity of the desired message is known. Thus it is not meaningful to study the privacy level of $W_{\mathcal{S}}$ for singleton $\mathcal{S}$, though we will still allow it for notational convenience.

In this work, we consider the two-level PIR system. The system parameters in such a system are $(N, T_1 : K_1, T_2 : K)$ with $T_1 \geq T_2 \geq 1$ and $1 \leq K_1 \leq K$. All the messages $W_{1:K}$ have the default weaker privacy level $T_2$, but the first $K_1$ messages $W_{1:K_1}$ have an enhanced privacy level $T_1$. We are interested in the retrieval rate (or simply rate) which is the number of useful message symbols retrieved per unit download

$$R \triangleq \frac{L}{\sum_{n=1}^{N} \mathbb{E}[\ell_n^{[k^*]}]}. \tag{7}$$

The download cost $D$ is defined as the inverse of $R$, i.e., $D \triangleq R^{-1}$. Schemes with higher achievable rates are preferred, and the supremum of the achievable rates among all possible schemes is called the capacity of the system, denoted as $C$.

## III. Main Result

To introduce the main result, we first provide some new notation. Denote the number of messages having the weaker privacy level $T_2$ but not the higher privacy level $T_1$ as

$$K_2 \triangleq K - K_1. \tag{8}$$

Define the function $D_N^*(K, T)$ as follows

$$D_N^*(K,T) \triangleq 1 + \frac{T}{N} + \cdots + \left(\frac{T}{N}\right)^{K-1}, \quad \forall T, K, N \in \mathbb{N}, \tag{9}$$

whose inverse is the capacity of the $T$-colluding PIR system with $N$ servers and $K$ messages (sometimes simply referred to as a $T$-private system). The main result of this work is summarized in the theorem below.

**Theorem 1.** *The capacity $C$ of the $(N, T_1 : K_1, T_2 : K)$ two-level PIR system satisfies*

$$R_{\mathrm{NS}} \leq C \leq \bar{R}, \tag{10}$$

*where*

$$R_{\mathrm{NS}} = \left( D_N^*(K_1, T_1) + \left(\frac{T_1}{N}\right)^{K_1} D_N^*(K_2, T_2) \right)^{-1}, \tag{11}$$

*and*

$$\overline{R} = \left( D_N^*(K_1, T_1) + \frac{T_2}{N} \left(\frac{T_1}{N}\right)^{K_1 - 1} D_N^*(K_2, T_2) \right)^{-1}. \tag{12}$$

The bound $R_{\mathrm{NS}}$ is obtained by a new coding scheme, which we refer to as the Non-uniform Successive-cancellation (NS) coding scheme. The proof for the bound $\bar{R}$ can be found in [20]. To further understand these bounds in Theorem 1, define

$$\underline{D} = \overline{R}^{-1}, \quad D_{\mathrm{NS}} = R_{\mathrm{NS}}^{-1}.$$

The difference between $\underline{D}$ and $D_{\mathrm{NS}}$ is

$$D_{\mathrm{NS}} - \underline{D} = \frac{T_1 - T_2}{N} \left(\frac{T_1}{N}\right)^{K_1 - 1} D^*(K_2, T_2). \tag{13}$$

It is seen that this gap diminishes geometrically as $K_1$ grows, and also vanishes when $T_1 = T_2$ as expected. Any $(N, T_1 : K, T_2 : K)$ code, i.e., a $T_1$-private code with $N$ servers and $K$

TABLE I
NS SCHEME IN $(N, T_1 : K_1, T_2 : K) = (4, 2 : 2, 1 : 4)$ FOR RETRIEVING $W_1$

| Coding group | Server-1 | Server-2 | Server-3 | Server-4 |
|---|---|---|---|---|
| $a$: (64, 64) | $a_1, a_2, a_3$ | $a_4, a_5, a_6$ | $a_7, a_8, a_9$ | $a_{10}, a_{11}, a_{12}$ |
| $b$: (24, 12) | $b_1, b_2, b_3$ | $b_4, b_5, b_6$ | $b_7, b_8, b_9$ | $b_{10}, b_{11}, b_{12}$ |
| $c$: (8, 4) | $c_1$ | $c_2$ | $c_3$ | $c_4$ |
| $d$: (8, 4) | $d_1$ | $d_2$ | $d_3$ | $d_4$ |
| | $a_{13} + b_{13}$ | $a_{14} + b_{14}$ | $a_{15} + b_{15}$ | $a_{16} + b_{16}$ |
| | $a_{17} + b_{17}$ | $a_{18} + b_{18}$ | $a_{19} + b_{19}$ | $a_{20} + b_{20}$ |
| | $a_{21} + b_{21}$ | $a_{22} + b_{22}$ | $a_{23} + b_{23}$ | $a_{24} + b_{24}$ |
| | $a_{25} + c_5$ | $a_{26} + c_6$ | $a_{27} + c_7$ | $a_{28} + c_8$ |
| | $a_{29} + d_5$ | $a_{30} + d_6$ | $a_{31} + d_7$ | $a_{32} + d_8$ |
| $b, c$: (8,4) | $b_{25} + c_9$ | $b_{26} + c_{10}$ | $b_{27} + c_{11}$ | $b_{28} + c_{12}$ |
| $b, d$: (8,4) | $b_{29} + d_9$ | $b_{30} + d_{10}$ | $b_{31} + d_{11}$ | $b_{32} + d_{12}$ |
| $c, d$: (24, 12) | $c_{13} + d_{13}$ | $c_{16} + d_{16}$ | $c_{19} + d_{19}$ | $c_{22} + d_{22}$ |
| | $c_{14} + d_{14}$ | $c_{17} + d_{17}$ | $c_{20} + d_{20}$ | $c_{23} + d_{23}$ |
| | $c_{15} + d_{15}$ | $c_{18} + d_{18}$ | $c_{21} + d_{21}$ | $c_{24} + d_{24}$ |
| | $a_{33} + b_{33} + c_{25}$ | $a_{34} + b_{34} + c_{26}$ | $a_{35} + b_{35} + c_{27}$ | $a_{36} + b_{36} + c_{28}$ |
| | $a_{37} + b_{37} + d_{25}$ | $a_{38} + b_{38} + d_{26}$ | $a_{39} + b_{39} + d_{27}$ | $a_{40} + b_{40} + d_{28}$ |
| | $a_{41} + c_{29} + d_{29}$ | $a_{42} + c_{30} + d_{30}$ | $a_{43} + c_{31} + d_{31}$ | $a_{44} + c_{32} + d_{32}$ |
| | $a_{45} + c_{33} + d_{33}$ | $a_{46} + c_{34} + d_{34}$ | $a_{47} + c_{35} + d_{35}$ | $a_{48} + c_{36} + d_{36}$ |
| | $a_{49} + c_{37} + d_{38}$ | $a_{50} + c_{38} + d_{38}$ | $a_{51} + c_{39} + d_{39}$ | $a_{52} + c_{40} + d_{40}$ |
| $b, c, d$: (24, 12) | $b_{41} + c_{41} + d_{41}$ | $b_{42} + c_{42} + d_{42}$ | $b_{43} + c_{43} + d_{43}$ | $b_{44} + c_{44} + d_{44}$ |
| | $b_{45} + c_{45} + d_{45}$ | $b_{46} + c_{46} + d_{46}$ | $b_{47} + c_{47} + d_{47}$ | $b_{48} + c_{48} + d_{48}$ |
| | $b_{49} + c_{49} + d_{49}$ | $b_{50} + c_{50} + d_{50}$ | $b_{51} + c_{51} + d_{51}$ | $b_{52} + c_{52} + d_{52}$ |
| | $a_{53} + b_{53} + c_{53} + d_{53}$ | $a_{54} + b_{54} + c_{54} + d_{54}$ | $a_{55} + b_{55} + c_{55} + d_{55}$ | $a_{56} + b_{56} + c_{56} + d_{56}$ |
| | $a_{57} + b_{57} + c_{57} + d_{57}$ | $a_{58} + b_{58} + c_{58} + d_{58}$ | $a_{59} + b_{59} + c_{59} + d_{59}$ | $a_{60} + b_{60} + c_{60} + d_{60}$ |
| | $a_{61} + b_{61} + c_{61} + d_{61}$ | $a_{62} + b_{62} + c_{62} + d_{62}$ | $a_{63} + b_{63} + c_{63} + d_{63}$ | $a_{64} + b_{64} + c_{64} + d_{64}$ |

messages, is valid for the $(N, T_1 : K_1, T_2 : K)$ PIR system. The optimal download cost of the former is exactly given by $D_{\text{T-PIR}} = D_N^*(K, T_1)$. Comparing with this naive approach, the coding gain of the proposed NS scheme is thus

$$D_{\text{T-PIR}} - D_{\text{NS}} = \left(\frac{T_1}{N}\right)^{K_1} \left(D_N^*(K_2, T_1) - D_N^*(K_2, T_2)\right), \tag{14}$$

which is always non-negative, and strictly positive if and only if $K_2 \geq 2$. Note that the strategy of using an $(N, T_1 : K, T_2 : K)$ code when a message in $W_{\mathcal{S}}$ is requested, and using an $(N, T_1 : 1, T_2 : K)$ code for the other messages is not valid, since this would lead to privacy leakage in the latter case, i.e., the information that the requested message is not in $\mathcal{S}$.

## IV. The Non-uniform Successive Cancellation Scheme

We next provide an example to illustrate the proposed NS coding scheme, and the details for the general code construction can be found in [20]. In this example, the two-level PIR system is specified by the parameters $(N, T_1 : K_1, T_2 : K) = (4, 2 : 2, 1 : 4)$, i.e., there are 4 servers and 4 messages $W_{1:4}$, and messages $W_{1:2}$ have privacy level $T_1 = 2$, while all messages $W_{1:4}$ have privacy level $T_2 = 1$. The length of each message is $L = 64$ here.

**Encoding:** To retrieve a message, the answers are formed in three steps, and the queries are simply the encoding matrix for these answers. Assume for each $(n, k)$ pair where $n \geq k$, an MDS code in $\mathbb{F}_q$ is given and fixed, and we refer to it as the $(n, k)$ MDS code. The coding structure is illustrated in Table I and Table II, for the retrieval of $W_1$ and $W_4$, respectively. The coding steps can be understood as follows:

1) *Precoding:* Let $S_1, S_2, S_3$, and $S_4$ be four random matrices, which are independently and uniformly drawn from the set of all $64 \times 64$ full rank matrices over $\mathbb{F}_q$; these matrices are known only to the user. The precoded messages $W_{1:4}^*$ are

$$W_1^* = S_1 W_1; \quad W_2^* = S_2 W_2; \\ W_3^* = S_3 W_3; \quad W_4^* = S_4 W_4. \tag{15}$$

2) *Group-wise MDS coding:* The precoded messages are partitioned into non-overlapping segments, and each segment is MDS-coded under certain $(n, k)$ parameters, the result of which is referred to as a coding group. These MDS-coded symbols for the four messages are denoted as

TABLE II
NS SCHEME IN $(N, T_1 : K_1, T_2 : K) = (4, 2 : 2, 1 : 4)$ FOR RETRIEVING $W_4$

| Coding group | Server-1 | Server-2 | Server-3 | Server-4 |
|---|---|---|---|---|
| $d$: (64, 64) | $d_1$ | $d_2$ | $d_3$ | $d_4$ |
| $a$: (16, 4)<br>$b$: (16, 4)<br>$c$: (16, 4) | $a_1, a_2, a_3$<br>$b_1, b_2, b_3$<br>$c_1$ | $a_4, a_5, a_6$<br>$b_4, b_5, b_6$<br>$c_2$ | $a_7, a_8, a_9$<br>$b_7, b_8, b_9$<br>$c_3$ | $a_{10}, a_{11}, a_{12}$<br>$b_{10}, b_{11}, b_{12}$<br>$c_4$ |
| | $a_{29} + d_5$<br>$b_{29} + d_9$<br>$c_{13} + d_{13}$<br>$c_{14} + d_{14}$<br>$c_{15} + d_{15}$ | $a_{30} + d_6$<br>$b_{30} + d_{10}$<br>$c_{16} + d_{16}$<br>$c_{17} + d_{17}$<br>$c_{18} + d_{18}$ | $a_{31} + d_7$<br>$b_{31} + d_{11}$<br>$c_{19} + d_{19}$<br>$c_{20} + d_{20}$<br>$c_{21} + d_{21}$ | $a_{32} + d_8$<br>$b_{32} + d_{12}$<br>$c_{22} + d_{22}$<br>$c_{23} + d_{23}$<br>$c_{24} + d_{24}$ |
| $a, b$ : (16, 4)<br><br><br>$a, c$: (16,4)<br>$b, c$: (16,4) | $a_{13} + b_{13}$<br>$a_{17} + b_{17}$<br>$a_{21} + b_{21}$<br>$a_{25} + c_5$<br>$b_{25} + c_9$ | $a_{14} + b_{14}$<br>$a_{18} + b_{18}$<br>$a_{22} + b_{22}$<br>$a_{26} + c_6$<br>$b_{26} + c_{10}$ | $a_{15} + b_{15}$<br>$a_{19} + b_{19}$<br>$a_{23} + b_{23}$<br>$a_{27} + c_7$<br>$b_{27} + c_{11}$ | $a_{16} + b_{16}$<br>$a_{20} + b_{20}$<br>$a_{24} + b_{24}$<br>$a_{28} + c_8$<br>$b_{28} + c_{12}$ |
| | $a_{37} + b_{37} + d_{25}$<br>$a_{41} + c_{29} + d_{29}$<br>$a_{45} + c_{33} + d_{33}$<br>$a_{49} + c_{37} + d_{38}$<br>$b_{41} + c_{41} + d_{41}$<br>$b_{45} + c_{45} + d_{45}$<br>$b_{49} + c_{49} + d_{49}$ | $a_{38} + b_{38} + d_{26}$<br>$a_{42} + c_{30} + d_{30}$<br>$a_{46} + c_{34} + d_{34}$<br>$a_{50} + c_{38} + d_{38}$<br>$b_{42} + c_{42} + d_{42}$<br>$b_{46} + c_{46} + d_{46}$<br>$b_{50} + c_{50} + d_{50}$ | $a_{39} + b_{39} + d_{27}$<br>$a_{43} + c_{31} + d_{31}$<br>$a_{47} + c_{35} + d_{35}$<br>$a_{51} + c_{39} + d_{39}$<br>$b_{43} + c_{43} + d_{43}$<br>$b_{47} + c_{47} + d_{47}$<br>$b_{51} + c_{51} + d_{51}$ | $a_{40} + b_{40} + d_{28}$<br>$a_{44} + c_{32} + d_{32}$<br>$a_{48} + c_{36} + d_{36}$<br>$a_{52} + c_{40} + d_{40}$<br>$b_{44} + c_{44} + d_{44}$<br>$b_{48} + c_{48} + d_{48}$<br>$b_{52} + c_{52} + d_{52}$ |
| $a, b, c$: (16,4) | $a_{33} + b_{33} + c_{25}$ | $a_{34} + b_{34} + c_{26}$ | $a_{35} + b_{35} + c_{27}$ | $a_{36} + b_{36} + c_{28}$ |
| | $a_{53} + b_{53} + c_{53} + d_{53}$<br>$a_{57} + b_{57} + c_{57} + d_{57}$<br>$a_{61} + b_{61} + c_{61} + d_{61}$ | $a_{54} + b_{54} + c_{54} + d_{54}$<br>$a_{58} + b_{58} + c_{58} + d_{58}$<br>$a_{62} + b_{62} + c_{62} + d_{62}$ | $a_{55} + b_{55} + c_{55} + d_{55}$<br>$a_{59} + b_{59} + c_{59} + d_{59}$<br>$a_{63} + b_{63} + c_{63} + d_{63}$ | $a_{56} + b_{56} + c_{56} + d_{56}$<br>$a_{60} + b_{60} + c_{60} + d_{60}$<br>$a_{64} + b_{64} + c_{64} + d_{64}$ |

$a_{1:64}, b_{1:64}, c_{1:64}, d_{1:64}$, respectively. In the tables, these coding groups are distinguished using different colors, with the corresponding MDS parameters given in the first column. For example, the red coding groups in Table I for both $b_{25:28,33:36}$ and $c_{9:12,25:28}$ are obtained by encoding 4 pre-coded symbols in $W_2^*$ and $W_4^*$, respectively. In each coding group, the coded symbols are ordered and sequentially placed in the tables, indicated by their subscripts.

3) *Forming pre-coded message sums:* The summations of the MDS-coded messages are formed accordingly, which can be seen clearly from Table I and Table II.

**Decoding and correctness:** The coding structure is layered, where in each layer the number of summands in each downloaded symbol is the same. From top to bottom, the number of summands increases from 1 to 4. The symbols of interference messages in each coding group are placed in two adjacent layers, where the signals (i.e., the summation of the symbols of interference messages) in the top layer can decode the interference signals in lower layer due to the common linear MDS code.

In Table I, for each coding group, the total number of interference signals placed in two adjacent layers and the top layer follow the ratio $(2 : 1) = (8 : 4) = (24 : 12)$. For example, 8 interference signals in the red coding group are placed in the second and third layers, where 4 downloaded symbols $b_{25:28} + c_{9:12}$ in the second layer can decode $b_{33:36} + c_{25:28}$ in the third layer, because $b, c$ are encoded by the same linear $(8, 4)$ MDS code. Consequently, $a_{33:36}$ can be recovered. It can be verified that $a_{1:64}$ can all be recovered either directly or in this fashion. By symmetry, $W_2$ can be retrieved similarly.

In Table II, for each coding group, the numbers of interference signals of each coding group placed in two adjacent layers and the top layer have the ratio $4 : 1$. For example, 16 interference signals in red coding groups are placed in the second and third layers, where any 4 of the 12 downloaded symbols $a_{13:24} + b_{13:24}$ in the second layer can decode $a_{37:40} + b_{37:40}$ in the third layer because $a, b$ are encoded by the same linear $(16, 4)$ MDS code. Consequently, $d_{25:28}$ can be recovered. It can be verified that $d_{1:64}$ can all be recovered either directly, or in this fashion. By symmetry, $W_3$ can be retrieved similarly.

**Privacy:** The coding pattern, i.e., the manner of forming pre-coded message sums, is the same for the retrieval of any message in $W_{1:4}$. Since it is a linear code, the coded symbols can be generated by the corresponding coding matrices. From

Table I, it is seen that the coding matrix of the coded symbols of any message from any two servers has full row-rank. For examples, the coded symbols $a$'s in server-1 and server-2 can be generated by a full row rank coding matrix using the message $W_1$, due to the pre-coding and the group-wise MDS coding. By applying Lemma 1 below, the messages $W_{1:2}$ thus have privacy level 2. The 1-privacy for all the messages can be seen in a similar manner.

**Lemma 1** (Statistical effect of full rank matrices [2])**.** *Let $S_1, S_2, \ldots, S_K \in \mathbb{F}_q^{\alpha\times\alpha}$ be $K$ random matrices, drawn independently and uniformly from all $\alpha\times\alpha$ full-rank matrices over $\mathbb{F}_q$. Let $G_1, G_2, \ldots, G_K \in \mathbb{F}_q^{\beta\times\beta}$ be $K$ invertible square matrices of dimension $\beta\times\beta$ over $\mathbb{F}_q$. Let $\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_K \in \mathbb{N}^{\beta\times 1}$ be $K$ index vectors, each containing $\beta$ distinct indices from $[1:\alpha]$. Then*

$$(G_1S_1[\mathcal{I}_1,:], G_2S_2[\mathcal{I}_2,:], \ldots, G_KS_K[\mathcal{I}_K,:]) \sim (S_1[1:\beta,:], S_2[1:\beta,:], \ldots, S_K[1:\beta,:]), \quad (16)$$

*where the notation $S[\mathcal{I},:]$ is used to indicate the submatrix of $S$ by taking its rows in $\mathcal{I}$.*

**Performance:** The total number of downloaded symbols is 116 and the message length is 64. Thus the rate is $R_{\rm NS} = \frac{64}{116} = \frac{16}{29}$. The optimal scheme for 2-private systems has rate $\frac{8}{15} < R_{\rm NS}$.

*Remark:* The construction resembles the scheme in [2], but it allows non-uniform coding structure to leverage the requirements of two levels of privacy. Due to the homogeneity of the privacy requirements for all the messages in $T$-private systems, the MDS coding parameters for each coding group are chosen to be $(N, T)$. In the proposed scheme for the $(N, T_1 : K_1, T_2 : K)$ system, there is symmetry among servers, and also symmetries among $W_{1:K_1}$ and among $W_{K_1+1:K}$ but not across all the messages. Thus when retrieving message $W_{k^*}$ with $k^* \in 1 : K_1$, the ratio of the MDS parameters $(n, k)$ in each coding group of the undesired messages need to be chosen as $(N, T_1)$, while as for message $W_{k^*}$ with $k^* \in K_1 + 1 : K$, the MDS coding parameters in each coding group would be $(N, T_2)$. However, since $N/T_1 < N/T_2$, with the same retrieval pattern, there exists certain slack in the placement pattern when retrieving $W_{k^*}$ with $k^* \in K_1+1 : K$. For example, the red coding group in Table II only needs 4 symbols in layer-2 to decode the remaining symbols in both layer-2 and layer-3, yet 12 symbols are retrieved and available directly in layer-2.

## V. Concluding Remarks

We considered two-level private information retrieval systems, and provided a capacity lower bound by proposing a novel code construction and a capacity upper bound. In the extended version [20] of this work, we provide another coding scheme which is referred to the non-uniform block cancellation scheme (NB). The NB scheme is able to outperform the NS scheme in certain $(N, T_1 : K_1, T_2 : K)$ regimes. However, the lower bounds and the upper bound do not match in general. Moreover, in the extended version we show that the upper bound $\bar{R}$ is not tight in general by providing a stronger upper bound for the special case of $(3, 2 : 2, 1 : 3)$.

We suspect the proposed code constructions can be further improved to yield better lower bounds, which we leave as a future work. Other possible future work includes extensions to mutlilevel PIR with more than two privacy levels, and the joint design of multilevel PIR and the storage codes.

## Acknowledgment

The work of R.-D Zhou and C. Tian was supported in part by the National Science Foundation under Grants CCF-1816546 and CCF-2007067. The work of H. Sun was supported in part by the National Science Foundation under Grant CCF-2007108. The work of J. Plank was upported in part by the National Science Foundation under Grant CCF-1816518.

## References

[1] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4075–4088, 2017.

[2] ——, "The capacity of robust private information retrieval with colluding databases," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2361–2370, 2017.

[3] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, C. Hollanti, and S. El Rouayheb, "Private information retrieval schemes for coded data with arbitrary collusion patterns," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 1908–1912.

[4] X. Yao, N. Liu, and W. Kang, "The capacity of private information retrieval under arbitrary collusion patterns," *arXiv preprint arXiv:2001.03843*, 2020.

[5] C. Tian, H. Sun, and J. Chen, "Capacity-achieving private information retrieval codes with optimal message size and upload cost," *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7613–7627, 2019.

[6] H. Sun and S. A. Jafar, "Optimal download cost of private information retrieval for arbitrary message length," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 2920–2932, 2017.

[7] R. Zhou, C. Tian, H. Sun, and T. Liu, "Capacity-achieving private information retrieval codes from mds-coded databases with minimum message size," *IEEE Transactions on Information Theory*, 2020.

[8] K. Banawan and S. Ulukus, "Asymmetry hurts: Private information retrieval under asymmetric traffic constraints," *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7628–7645, 2019.

[9] H.-Y. Lin, S. Kumar, E. Rosnes, and A. G. i Amat, "Asymmetry helps: Improved private information retrieval protocols for distributed storage," in *2018 IEEE Information Theory Workshop (ITW)*. IEEE, 2018, pp. 1–5.

[10] K. Banawan and S. Ulukus, "The capacity of private information retrieval from coded databases," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1945–1956, 2018.

[11] R. Freij-Hollanti, O. W. Gnilke, C. Hollanti, and D. A. Karpuk, "Private information retrieval from coded databases with colluding servers," *SIAM Journal on Applied Algebra and Geometry*, vol. 1, no. 1, pp. 647–664, 2017.

[12] S. Kumar, H.-Y. Lin, E. Rosnes, and A. G. i Amat, "Achieving maximum distance separable private information retrieval capacity with linear codes," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4243–4273, 2019.

[13] M. A. Attia, D. Kumar, and R. Tandon, "The capacity of private information retrieval from uncoded storage constrained databases," *arXiv preprint arXiv:1805.04104*, 2018.

[14] K. Banawan, B. Arasli, Y.-P. Wei, and S. Ulukus, "The capacity of private information retrieval from heterogeneous uncoded caching databases," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3407–3416, 2020.

[15] N. Woolsey, R.-R. Chen, and M. Ji, "Uncoded placement with linear sub-messages for private information retrieval from storage constrained databases," *IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6039–6053, 2020.

[16] C. Tian, H. Sun, and J. Chen, "A shannon-theoretic approach to the storage-retrieval tradeoff in pir systems," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 1904–1908.

[17] H. Sun and C. Tian, "Breaking the mds-pir capacity barrier via joint storage coding," *Information*, vol. 10, no. 9, p. 265, 2019.

[18] C. Tian, "On the storage cost of private information retrieval," *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7539–7549, 2020.

[19] T. Guo, R. Zhou, and C. Tian, "New results on the storage-retrieval tradeoff in private information retrieval systems," *arXiv preprint arXiv:2008.00960*, 2020.

[20] R. Zhou, C. Tian, H. Sun, and J. Plank "Two-level private information retrieval," *arXiv preprint arXiv:2101.04821*, 2020.