Effect of Full Common Randomness Replication in Symmetric PIR on Graph-Based Replicated Systems

Shreya Meel Sennur Ulukus
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
smeel@umd.edu *ulukus@umd.edu*

***Abstract*—We revisit the problem of symmetric private information retrieval (SPIR) in settings where the database replication is modeled by a simple graph. Here, each vertex corresponds to a server, and a message is replicated on two servers if and only if there is an edge between them. To satisfy the requirement of database privacy, we let all the servers share some common randomness, independent of the messages. We aim to quantify the improvement in SPIR capacity, i.e., the maximum ratio of the number of desired and downloaded symbols, compared to the setting with graph-replicated common randomness. Towards this, we develop an algorithm to convert a class of PIR schemes into the corresponding SPIR schemes, thereby establishing a capacity lower bound on graphs for which such schemes exist. This includes the class of path and cyclic graphs for which we derive capacity upper bounds that are tighter than the trivial bounds given by the respective PIR capacities. For the special case of path graph with three vertices, we identify the SPIR capacity to be $\frac{1}{2}$.**

I. Introduction

Private information retrieval (PIR) [1], is a cryptographic primitive where a user wishes to download their desired message in a database, potentially replicated in multiple non-colluding servers, without revealing the index of the message to any server. Over the past decade, PIR has been heavily studied from an information-theoretic perspective, with the primary metric of interest being the PIR capacity [2], defined as the maximum number of message symbols retrieved per downloaded symbol. Beyond the basic setting [2], characterizing the PIR capacity under various configurations has been the focus of multiple works (see e.g., [3]–[6] and [7] for a survey). One assumption in PIR is that, the privacy comes at the cost of the user learning a part of the messages they did not require. This is detrimental if the database contains sensitive information, and no database information beyond the desired message should be revealed to the user. To address this issue, symmetric PIR (SPIR) problem was formulated in [8], introducing symmetric privacy requirements for both the user and the database.

To make SPIR feasible, the servers share some common randomness variable [8] prior to the communication protocol. The SPIR capacity and the minimum amount of randomness required to achieve the capacity was characterized in [9] in the basic fully-replicated database setting. This study was extended to other settings; e.g., SPIR with MDS coded messages [10], [11], SPIR with resilience against passive and active adversaries [12], SPIR for multiple messages [13], SPIR with side information [14]. All these works assume that, the entire database is available at all servers in a coded or an uncoded form. In practice, it may be judicious to constrain the replication of the databases at all servers, subject to cost and security reasons. Moreover, a database, though replicated, may be partially accessible to the user, subject to access control constraints [15], [16]. Such scenarios motivate the study of PIR [17]–[20] and SPIR [21] on databases whose replication pattern is described by a graph or hypergraph.

In the graph-replicated database model, each vertex corresponds to a server, and each edge represents a message stored on them. Different from the SPIR problem formulation in [21], we do not restrict the availability of common randomness to the servers that share a message. Instead, we let the common randomness to be shared by all servers, independent of the graph modeling the database replication. Our goal in this paper is to study the improvement in capacity through full common randomness replication. For instance, we show that the capacity of a path graph on 3 vertices improves from $\frac{1}{3}$ in [21] to $\frac{1}{2}$ in our setting. In general, we characterize a capacity lower bound for all graphs whose PIR scheme admits a symmetric structure, followed by an upper bound for path and cyclic graphs. Similar to PIR capacity, our bounds on SPIR capacity depend on the explicit graph structure.

II. System Model

Consider a database of $K \geq 2$ independent messages given by the set $\mathcal{W} := \{W_1, \ldots, W_K\}$. Each message comprises L symbols selected uniformly at random from a finite field $\mathbb{F}_q$,

$$H(\mathcal{W}) = H(W_1) + \ldots + H(W_K) \tag{1}$$
$$= KL, \quad \text{in } q\text{-ary units.} \tag{2}$$

The messages are stored on $N \geq 2$ non-colluding servers, denoted by the set $\mathcal{S} = [N]$. Each message $W_k \in \mathcal{W}$ is replicated exactly twice and stored on two distinct servers in $\mathcal{S}$. We represent the database system by a simple, undirected graph $G = (V, E)$, with vertex set $V = \mathcal{S}$, and edge set $E = \mathcal{W}$. Hence, each vertex represents a server and each edge incident with a pair of servers, represents the message replicated at those servers. Further, we assume that G is connected, i.e., there exists a path between each pair of vertices.

In PIR, a user chooses an index $\theta \in [K]$, and wishes to privately retrieve the message $W_\theta \in \mathcal{W}$, without revealing θ to any individual server. This is known as the user privacy. As

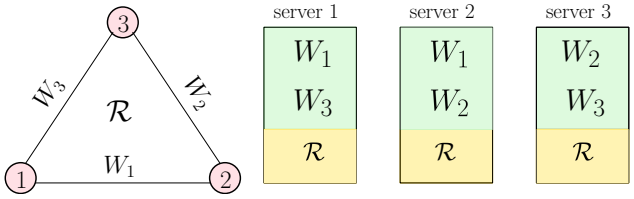

Fig. 1: System model for $\mathbb{C}_3$.

an additional constraint in SPIR, the servers want to keep the messages other than W_θ private from the user. This is known as the database privacy. It is known [8] that SPIR is feasible only if the servers share some common randomness $\mathcal{R}$. The symbols of $\mathcal{R}$ are picked uniformly at random from $\mathbb{F}_q$, and are shared apriori by the servers, independent of $\mathcal{W}$. The SPIR storage, modeled by the cyclic graph on three vertices is illustrated in Fig. 1. Note that, different from [21], the common randomness $\mathcal{R}$ is available to all servers, independent of the structure of G.

Let $\mathcal{Q}$ represent the randomness in the schemes followed by the user to retrieve the messages. Since $\mathcal{Q}$ is decided prior to choosing the message index, it is independent of θ. Further, θ and $\mathcal{Q}$ are independent of $\mathcal{W}$ and $\mathcal{R}$, since the user has no information of the content stored at the servers.

To retrieve W_k, the user generates N queries $Q_1^{[k]}, \ldots, Q_N^{[k]}$ employing $\mathcal{Q}$,

$$H(Q_1^{[k]}, \ldots, Q_N^{[k]} | \mathcal{Q}) = 0, \tag{3}$$

and sends $Q_i^{[k]}$ to server i. Upon receiving the query, server i responds with an answer $A_i^{[k]}$. Let $\mathcal{W}_i$ denote the set of messages stored at server i. Then,

$$H(A_i^{[k]} | Q_i^{[k]}, \mathcal{W}_i, \mathcal{R}) = 0. \tag{4}$$

The answers and queries should satisfy the requirements of user privacy, reliability and database privacy, as defined next. For user privacy, the query and answer for each server should be identically distributed, irrespective of θ, i.e., for every $i \in [N]$ and any index k,

$$(Q_i^{[k]}, A_i^{[k]}, \mathcal{W}_i, \mathcal{R}) \sim (Q_i^{[1]}, A_i^{[1]}, \mathcal{W}_i, \mathcal{R}). \tag{5}$$

For reliability, the user should be able to exactly recover their requested message W_k, using the answers from all the servers,

$$H(W_k | A_1^{[k]}, \ldots, A_N^{[k]}, \mathcal{Q}) = 0. \tag{6}$$

For database privacy, no information on the desired message complement $W_{\overline{k}} := \mathcal{W} \setminus \{W_k\}$ should be revealed to the user,

$$H(W_{\overline{k}}; A_1^{[k]}, \ldots, A_N^{[k]}, Q_1^{[k]}, \ldots, Q_N^{[k]} | \mathcal{Q}) = 0. \tag{7}$$

If an SPIR scheme with fully replicated randomness, simultaneously satisfies (5), (6), and (7), it is said to be achievable. The rate $R_{\text{F-R}}$ of an SPIR scheme on G is defined as the ratio between the number of desired message symbols and the total number of downloaded symbols,

$$R_{\text{F-R}}(G) \triangleq \frac{L}{\sum_{i=1}^{N} H(A_i^{[k]})}. \tag{8}$$

To distinguish this from the setting of [21], we include the subscript F-R, to imply fully-replicated. The capacity $\mathscr{C}_{\text{F-R}}(G)$ is defined as the supremum over all achievable rates. To characterize the amount of common randomness for an SPIR scheme, we define the total randomness ratio,

$$\rho_{total}(G) \triangleq \frac{H(\mathcal{R})}{L}, \tag{9}$$

as the size of total common randomness $\mathcal{R}$, required to be shared by all servers, relative to the message size.

It is intuitive that, replicating the common randomness in all servers leads to better interference alignment, implying savings in the download cost. Therefore, for any graph G, the capacity $\mathscr{C}_{\text{F-R}}(G)$ is greater than $\mathscr{C}(G)$, and the inequality

$$\mathscr{C}(G) < \mathscr{C}_{\text{F-R}}(G) < \mathscr{C}_{PIR}(G) \tag{10}$$

holds, where $\mathscr{C}_{PIR}(G)$ denotes the capacity of PIR on G, i.e., without the database privacy constraint. Therefore, $\mathscr{C}_{PIR}(G)$ is a trivial upper bound on $\mathscr{C}_{\text{F-R}}(G)$.

III. Main Results

Our first result characterizes a lower bound on $\mathscr{C}_{\text{F-R}}(G)$ by constructing an SPIR scheme from a PIR scheme on G. We consider the class of schemes wherein the PIR answer from every server is a set of t-sums (sum of t symbols from distinct messages stored in the server) where $t \geq 1$. Additionally, the PIR scheme should satisfy the symmetric retrieval property [20], as defined next.

Definition 1 (Symmetric Retrieval Property (SRP)) *A graph-based PIR scheme is said to satisfy the symmetric retrieval property if, for any realization k of θ, the number of symbols of W_k retrieved from each of the servers storing W_k, is equal. Equivalently, $H(A_i^{[k]} | Q_i^{[k]}, \mathcal{W}_i \setminus \{W_k\}) = H(A_j^{[k]} | Q_j^{[k]}, \mathcal{W}_j \setminus \{W_k\}) = \frac{H(W_k)}{2}$,if W_k is replicated on servers i and j.*

Theorem 1 *Given a PIR scheme on G with N vertices and K edges, that satisfies SRP, there exists an SPIR scheme with fully-replicated randomness on G. Let L' be the number of symbols per message (note that SRP forces the PIR scheme to have even L'), and D' be the total number of downloaded symbols of the PIR scheme, then the following SPIR rate,*

$$R_{F\text{-}R}(G) = \frac{L'x}{D'x + Ny} \leq \mathscr{C}_{F\text{-}R}(G) \tag{11}$$

is achievable, where $x = \frac{\ell}{L'/2}$, $y = \frac{\ell}{N-1}$, $\ell = lcm(L'/2, N-1)$ and lcm denotes the least common multiple. The corresponding randomness ratio is,

$$\rho_{total}(G) = \frac{K-1}{2} + \frac{Ny}{L'x}. \tag{12}$$

The algorithm in Section IV-B yields Theorem 1.

Corollary 1 *The capacity-achieving schemes of path [22] and cyclic [18] graphs satisfy the SRP. In fact, if G is a path graph $\mathbb{P}_N$ or a cyclic graph $\mathbb{C}_N$,* (11) *can be reduced to*

$$R_{F\text{-}R}(\mathbb{P}_N) = \frac{2}{N + \frac{N}{N-1}}, \tag{13}$$

$$R_{F\text{-}R}(\mathbb{C}_N) = \frac{2}{N + 1 + \frac{N}{N-1}}. \tag{14}$$

This strictly improves upon $\mathscr{C}(\mathbb{P}_N) = \mathscr{C}(\mathbb{C}_N) = \frac{1}{N}$ in [21] and is related to the respective PIR capacities by

$$\frac{1}{R_{F\text{-}R}(G)} = \frac{1}{\mathscr{C}_{PIR}(G)} + \frac{N}{2(N-1)}. \tag{15}$$

Further, (12) *simplifies to $\rho_{total}(G) = R_{F\text{-}R}(G)^{-1} - 1$.*

Theorem 2 *The SPIR capacity of $\mathbb{P}_3$ (see Fig. 2) is $\mathscr{C}_{F\text{-}R}(\mathbb{P}_3) = \frac{1}{2}$ with the optimal $\rho_{total}(\mathbb{P}_3) = 1$.*

We prove this through an achievable scheme (Section IV-C) and a matching upper bound (Section V-A). Next, we generalize this upper bound to $\mathbb{P}_N$ and $\mathbb{C}_N$. Both these bounds improve upon the trivial bounds $\mathscr{C}_{PIR}(\mathbb{P}_N)$ and $\mathscr{C}_{PIR}(\mathbb{C}_N)$.

Theorem 3 *The $\mathscr{C}_{F\text{-}R}(G)$ of path and cyclic graphs satisfy*

$$\mathscr{C}_{F\text{-}R}(\mathbb{P}_N) \leq \frac{2}{N + \frac{2}{N-1}}, \tag{16}$$

$$\mathscr{C}_{F\text{-}R}(\mathbb{C}_N) \leq \frac{2}{N + 1 + \frac{1}{N-1}}. \tag{17}$$

The proof of Theorem 3 is provided in Section V-B.

IV. SPIR from PIR Schemes

In this section, we describe the algorithm for Theorem 1, followed by the achievability proof of Theorem 2. Before the algorithm description, we present its idea with two examples.

A. Motivating Examples

Example 1 (Path graph) *Consider the SPIR system for $\mathbb{P}_3$ as shown in Fig. 2. Suppose, $L = 4$ and $H(\mathcal{R}) = 5$. Let the message and randomness symbols, after private and independent permutations by the user be denoted by $W_1 = (a_1, \ldots, a_4)$, $W_2 = (b_1, \ldots, b_4)$ and $\mathcal{R} = (s_1, \ldots, s_5)$. The answers of the SPIR scheme are shown in Table I. First, the user downloads a unique randomness symbol from each server. Next, the user performs two repetitions of the PIR scheme on $\mathbb{P}_3$, with the modification that each downloaded answer is mixed with randomness symbols from $\mathcal{R}$. For instance, if $\theta = 1$, we add a new randomness symbol s_3, s_5 to b_2 and b_4, respectively. In the first repetition, we add s_1 (downloaded from server 2) to a_1, and s_2 (downloaded from server 1) to a_2. In the second repetition, we add s_4 (downloaded from server 3) to a_3 and a_4. This gives $R_{F\text{-}R}(\mathbb{P}_3) = \frac{4}{9} = \frac{2}{3+\frac{3}{2}}$ and $\rho_{total}(\mathbb{P}_N) = \frac{5}{4} = \frac{9}{4} - 1$.*

Example 2 (Cyclic graph) *For the SPIR scheme on $\mathbb{C}_N$ with $N = 3$, we build upon the scheme in [18]. The answer table when W_1 is desired is shown in Table II. Corresponding to each undesired symbol, we assign a new distinct randomness symbol to be added to the answer. For each desired symbol downloaded from server 1 (server 2), a randomness symbol downloaded from server 2 (server 1) and server 3 is assigned. This gives $R_{F\text{-}R}(\mathbb{C}_N) = \frac{4}{11} = \frac{2}{3+1+\frac{3}{2}}$ and $\rho_{total}(\mathbb{C}_N) = \frac{7}{4} = \frac{11}{4} - 1$.*

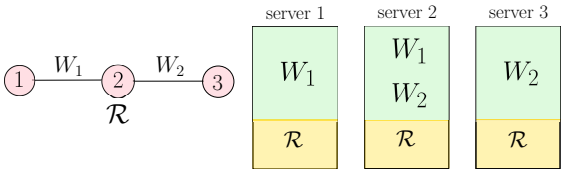

Fig. 2: SPIR model for $\mathbb{P}_3$.

$\theta = 1$	database 1	database 2	database 3
	s_2	s_1	s_4
rep. 1	$a_1 + s_1$	$a_2 + b_2 + s_2 + s_3$	$b_2 + s_3$
rep. 2	$a_3 + s_4$	$a_4 + b_4 + s_4 + s_5$	$b_4 + s_5$
$\theta = 2$	database 1	database 2	database 3
	s_5	s_3	s_1
rep. 1	$a_1 + s_1$	$a_1 + b_1 + s_1 + s_2$	$b_2 + s_3$
rep. 2	$a_3 + s_4$	$a_3 + b_3 + s_4 + s_5$	$b_4 + s_5$

TABLE I: Answer table for $\mathbb{P}_3$.

B. Algorithm Description

It is clear from the examples that our scheme exhibits two key properties. 1) The user downloads an equal number y of uncoded randomness symbols from every server, and 2) if the randomness symbols are removed from the answers, our scheme reduces to x repetitions of the starting PIR scheme.

Let W_k be the desired message, replicated on servers $i, j \in [N]$. To determine x and y, for a given PIR scheme, observe that, in each repetition x, the user retrieves $xL'/2$ symbols of W_k, masked by a symbol of $\mathcal{R}$, from each server i and j. For decoding the $xL'/2$ symbols of W_k retrieved from each of server i and j, the user downloads the corresponding $xL'/2$ randomness symbols from the remaining servers, i.e., $[N] \setminus \{i\}$ and $[N] \setminus \{j\}$, respectively. Therefore, by property 1, this implies, x and y are least positive integers satisfying

$$x\frac{L'}{2} = (N-1)y. \tag{18}$$

Clearly, (18) is equal to $\text{lcm}(L'/2, N-1)$ by the definition of least common multiple. Now, we describe our algorithm steps.

a) Index permutation: For the SPIR scheme, let $L = xL'$ and $H(\mathcal{R}) = Ny + \left(\frac{K-1}{2}\right) L$. The user applies private and independent permutations to the symbols of the K messages, with $W_\ell = (w_\ell(1), \ldots, w_\ell(L))$ and $\mathcal{R} = (s_1, \ldots, s_{Ny+L(K-1)/2})$.

b) Query generation: The user follows the query structure of the starting PIR scheme. Queries sent to the servers correspond to x repetitions of the scheme, where in each repetition, the user queries for L' new symbols of W_k. Corresponding to each queried symbol, the user sends the following common randomness assignment to each server.

	database 1	database 2	database 3
	s_7, s_9, s_{11}	s_1, s_3, s_5	s_{13}, s_{15}, s_{17}
rep. 1	$a_1 + s_1, c_1 + s_2$	$a_4 + s_7, b_1 + s_8$	$b_2 + s_{10}, c_2 + s_4$
	$a_2 + c_2 + s_3 + s_4$	$a_5 + b_2 + s_9 + s_{10}$	$b_1 + c_3 + s_8 + s_6$
	$a_3 + c_3 + s_5 + s_6$	$a_6 + b_3 + s_{11} + s_{12}$	$b_3 + c_1 + s_{12} + s_2$
rep. 2	$a_7 + s_{13}, c_7 + s_{14}$	$a_{10} + s_{13}, b_7 + s_{19}$	$b_8 + s_{20}, c_8 + s_{16}$
	$a_8 + c_8 + s_{15} + s_{16}$	$a_{11} + b_8 + s_{15} + s_{20}$	$b_7 + c_9 + s_{19} + s_{18}$
	$a_9 + c_9 + s_{17} + s_{18}$	$a_{12} + b_9 + s_{17} + s_{21}$	$b_9 + c_7 + s_{14} + s_{21}$

TABLE II: Answer table for $\mathbb{C}_3$ when $\theta = 1$.

c) Common randomness assignment: The user assigns a unique randomness symbol to each queried message symbol $W_\ell \in W_{\overline{k}}$ across all repetitions. Note that, because of SRP, to maintain user privacy, we query $L'/2$ symbols of W_ℓ from every server in the PIR scheme. For reliability of the PIR scheme through interference cancellation, we require the same $L'/2$ symbols to be queried from both servers where W_ℓ is replicated. This assigns the $(K-1)L/2$ symbols of $\mathcal{R}$. To the first $y = \frac{L}{2(N-1)}$ symbols of W_k queried from server i and j, the user assigns distinct randomness symbols $s_{j_1}, \ldots, s_{j_y}$ and $s_{i_1}, \ldots, s_{i_y}$, respectively. To each of the remaining $L/2 - y = (N-2)y$ queried symbols of W_k from each of server i and j, the user assigns the same randomness symbol from the remaining set $\{s_{n_1}, \ldots, s_{n_y}, n \in [N] \setminus \{i, j\}\}$ of $\mathcal{R}$.

d) Answer formation: The user downloads the randomness symbols $s_{n_1}, \ldots, s_{n_y}$ from server $n \in [N]$, followed by x repetitions of PIR answers. To each answer, the servers add randomness symbols, with each message symbol in the answer added to the assigned randomness symbol. That is, a t-sum in the answer is padded with t symbols from $\mathcal{R}$.

This completes the algorithm, with $R_{\text{F-R}}$ and ρ_{total} derived by direct computation. Next, we will show that our algorithm outputs a feasible SPIR scheme.

Reliability: By reliability of the PIR scheme, in repetition $u \in [x]$, the user retrieves $w_k((u-1)L'+1), \ldots, w_k(uL')$ added to an already downloaded symbol of $\mathcal{R}$. Using this, they recover $L = xL'$ symbols of W_k. Specifically, $s_{j_1}, \ldots, s_{j_y}$ and $s_{i_1}, \ldots, s_{i_y}$ cancel out the randomness symbols to retrieve the first y symbols of W_k from server i and j, respectively. Further, $s_{n_1}, \ldots, s_{n_y}$ downloaded from servers $n \in [N] \setminus \{i, j\}$ are used as common side information to retrieve the remaining $L - 2y$ symbols of W_k from both servers i and j.

User privacy: From the perspective of each server, the queried message symbols are independent of θ, due to the PIR scheme. Further, the randomness symbols directly downloaded by the user, and those added to the PIR answers, appear to be uniformly chosen from $\mathcal{R}$, irrespective of θ, since the private permutation on $\mathcal{R}$ hides the common randomness assignment.

Database privacy: No information on $W_{\overline{k}}$ is revealed to the user, since every non-desired message symbol is protected by a unique symbol of $\mathcal{R}$, which is distinct from those downloaded by the user in the scheme.

C. Achievability for $\mathbb{P}_3$

We demonstrate a simple scheme on $\mathbb{P}_3$ that achieves a higher rate of $R_{\text{F-R}}(\mathbb{P}_3) = \frac{1}{2}$ than $\frac{4}{9}$ in Example 1 with $\rho_{total}(\mathbb{P}_3) = 1 < \frac{5}{4}$. With $L = L' = 2$, the answers are shown in Table III, which in the absence of s_1, s_2, reduce to the PIR scheme on $\mathbb{P}_3$. The improvement comes from adding a single randomness symbol in the answer of server 2, and downloading a randomness symbol from server 2 alone. Notice that $\mathbb{P}_3$ is identical to a star graph, which enables this, unlike $\mathbb{P}_N, N \geq 4$.

	database 1	database 2	database 3
$\theta = 1$	$a_1 + s_1$	$s_1, a_2 + b_2 + s_2$	$b_2 + s_2$
$\theta = 2$	$a_1 + s_1$	$s_2, a_1 + b_1 + s_1$	$b_2 + s_2$

TABLE III: SPIR answer table for $\mathbb{P}_3$ for the improved scheme.

V. Upper Bound on SPIR Capacity

We have that any feasible SPIR scheme with fully-replicated randomness, $\mathcal{R}$ satisfies

$$\begin{aligned} H(\mathcal{R}) =& H(\mathcal{R}|\mathcal{W}, \mathcal{Q}) + H(A_{1:N}^{[k]}|\mathcal{W}, \mathcal{R}, \mathcal{Q}) && (19)\\ =& H(A_{1:N}^{[k]}, \mathcal{R}|\mathcal{W}, \mathcal{Q}) && (20)\\ \geq& H(A_{1:N}^{[k]}|\mathcal{W}, \mathcal{Q}) && (21)\\ =& H(A_{1:N}^{[k]}|\mathcal{Q}) - I(\mathcal{W}; A_{1:N}^{[k]}|\mathcal{Q}) && (22)\\ \overset{(6)(7)}{=}& H(A_{1:N}^{[k]}|\mathcal{Q}) - L. && (23)\end{aligned}$$

For the capacity upper bounds, we state the following lemmas. The first lemma generalizes [9, Lemma 2].

Lemma 1 *For any $\mathcal{J} \subseteq [K]$,*

$$H(A_n^{[k]}|W_{\mathcal{J}}, \mathcal{R}, Q_n^{[k]}) = H(A_n^{[k]}|W_{\mathcal{J}}, \mathcal{R}, Q_n^{[k]}, \mathcal{Q}). \quad (24)$$

The next lemma is a consequence of user privacy (5) and graph-replicated databases.

Lemma 2 *For any $\mathcal{J} \subseteq [K]$, and any server $n \in [N]$, for all $k, k' \in [K]$, we have*

$$\begin{aligned} H(A_n^{[k]}|W_{\mathcal{J}}, \mathcal{R}, Q_n^{[k]}) &= H(A_n^{[k']}|W_{\mathcal{J}}, \mathcal{R}, Q_n^{[k']}), && (25)\\ H(A_n^{[k]}|W_{\mathcal{J}}, Q_n^{[k]}) &= H(A_n^{[k']}|W_{\mathcal{J}}, Q_n^{[k']}). && (26)\end{aligned}$$

The following lemma is an extension of [19, Lemma 3].

Lemma 3 *For $k \in [K]$, let W_k be replicated on servers servers i and j. Then,*

$$H(A_i^{[k]}|W_{\overline{k}}, \mathcal{R}, \mathcal{Q}) + H(A_j^{[k]}|W_{\overline{k}}, \mathcal{R}, \mathcal{Q}) \geq L. \tag{27}$$

A. Converse for $\mathbb{P}_3$

Suppose W_1 is the desired message. Then,

$$\begin{aligned}
L \overset{(6)}{=} & H(W_1|\mathcal{Q}) - H(W_1|A_{1:3}^{[1]}, \mathcal{Q}) = I(A_{1:3}^{[1]}; W_1|\mathcal{Q}) && (28)\\
\overset{(3)}{=} & H(A_{1:3}^{[1]}|\mathcal{Q}) - H(A_1^{[1]}|W_1, Q_1^{[1]}, \mathcal{Q}) \\
& - H(A_{2:3}^{[1]}|W_1, A_1^{[1]}, \mathcal{Q}) && (29)\\
\leq & H(A_{1:3}^{[1]}|\mathcal{Q}) - H(A_1^{[1]}|W_1, Q_1^{[1]}) \\
& - H(A_{2:3}^{[1]}|W_1, A_1^{[1]}, \mathcal{Q}, \mathcal{R}) && (30)\\
= & H(A_{1:3}^{[1]}|\mathcal{Q}) - H(A_1^{[2]}|W_1, Q_1^{[2]}) - H(A_2^{[1]}|W_1, \mathcal{Q}, \mathcal{R}) \\
& - H(A_3^{[1]}|W_1, A_2^{[1]}, \mathcal{Q}, \mathcal{R}) && (31)\\
\overset{(7)(5)}{=} & H(A_{1:3}^{[1]}|\mathcal{Q}) - H(A_1^{[1]}|Q_1^{[1]}) - H(A_2^{[1]}|W_1, \mathcal{Q}, \mathcal{R}) \\
& - H(A_3^{[1]}|W_1, A_2^{[1]}, \mathcal{Q}, \mathcal{R}) && (32)\\
\leq & H(A_{1:3}^{[1]}|\mathcal{Q}) - H(A_1^{[1]}|\mathcal{Q}) - H(A_2^{[1]}|W_1, \mathcal{Q}, \mathcal{R}) \\
& - H(A_3^{[1]}|W_1, W_2, A_2^{[1]}, \mathcal{Q}, \mathcal{R}) && (33)\\
\overset{(4)}{=} & H(A_{1:3}^{[1]}|\mathcal{Q}) - H(A_1^{[1]}|\mathcal{Q}) - H(A_2^{[2]}|W_1, \mathcal{Q}, \mathcal{R}), && (34)
\end{aligned}$$

where (30) follows from Lemma 1 and conditioning with $\mathcal{R}$, (31) holds since $A_1^{[1]}$ is completely determined by $(W_1, \mathcal{Q}, \mathcal{R})$, and Lemma 2, and (34) also follows from Lemma 2. By interchanging the positions of $A_2^{[1]}$ and $A_3^{[1]}$ in (31), we obtain

$$L \leq H(A_{1:3}^{[1]}|\mathcal{Q}) - H(A_1^{[1]}|\mathcal{Q}) - H(A_3^{[2]}|W_1, \mathcal{Q}, \mathcal{R}). \tag{35}$$

From (34) and (35), and applying Lemma 3, we get

$$\begin{aligned}
L \leq & H(A_{1:3}^{[1]}|\mathcal{Q}) - H(A_1^{[1]}|\mathcal{Q}) \\
& - \frac{1}{2}\left(H(A_2^{[2]}|W_1, \mathcal{Q}, \mathcal{R}) + H(A_3^{[2]}|W_1, \mathcal{Q}, \mathcal{R})\right) && (36)\\
\leq & H(A_{1:3}^{[1]}|\mathcal{Q}) - H(A_1^{[1]}|\mathcal{Q}) - \frac{L}{2}, && (37)
\end{aligned}$$

which, since $H(A_{1:3}^{[1]}|\mathcal{Q}) - H(A_1^{[1]}|\mathcal{Q}) \leq H(A_2^{[1]}|\mathcal{Q}) + H(A_3^{[1]}|\mathcal{Q})$, upon rearrangement gives

$$H(A_2^{[1]}|\mathcal{Q}) + H(A_3^{[1]}|\mathcal{Q}) \geq \frac{3L}{2}. \tag{38}$$

Repeating the steps in (28)-(37) for W_2, we get

$$H(A_1^{[2]}|\mathcal{Q}) + H(A_2^{[2]}|\mathcal{Q}) \geq \frac{3L}{2}. \tag{39}$$

Moreover, for $k = 1, 2$, by bounding $H(A_{1:3}^{[k]}|W_k, \mathcal{Q})$ as

$$\begin{aligned}
& H(A_{1:3}^{[k]}|W_k, \mathcal{Q}) \\
& \quad = H(A_2^{[k]}|W_k, \mathcal{Q}) + H(A_1^{[k]}, A_3^{[k]}|W_k, A_2^{[k]}, \mathcal{Q}) && (40)\\
& \quad \geq H(A_2^{[k]}|\mathcal{Q}) + H(A_1^{[k]}, A_3^{[k]}|W_1, W_2, \mathcal{R}, A_2^{[k]}, \mathcal{Q}) && (41)\\
& \quad = H(A_2^{[k]}|\mathcal{Q}) + H(A_1^{[k]}, A_3^{[k]}|W_1, W_2, \mathcal{R}, \mathcal{Q}) && (42)\\
& \quad = H(A_2^{[k]}|\mathcal{Q}), && (43)
\end{aligned}$$

we get

$$H(A_1^{[k]}|\mathcal{Q}) + H(A_3^{[k]}|\mathcal{Q}) \geq L. \tag{44}$$

Finally, summing (34), (35) and (44), for $k = 1, 2$ by (5), we obtain

$$2\big(H(A_1^{[k]}|\mathcal{Q}) + H(A_2^{[k]}|\mathcal{Q}) + H(A_3^{[k]}|\mathcal{Q})\big) \geq 4L, \tag{45}$$

which gives $\mathscr{C}(\mathbb{P}_3) \leq \frac{1}{2}$. To determine the bound on $H(\mathcal{R})$, from (23), we have $H(\mathcal{R}) \geq H(A_{1:3}^{[k]}|\mathcal{Q}) - L$, which satisfies

$$H(A_{1:3}^{[1]}|\mathcal{Q}) - L \geq H(A_1^{[k]}|\mathcal{Q}) + \frac{L}{2}, \tag{46}$$

$$H(A_{1:3}^{[k]}|\mathcal{Q}) - L \geq H(A_2^{[k]}|\mathcal{Q}), \tag{47}$$

$$H(A_{1:3}^{[2]}|\mathcal{Q}) - L \geq H(A_3^{[k]}|\mathcal{Q}) + \frac{L}{2}. \tag{48}$$

From this, we obtain $3H(\mathcal{R}) \geq H(A_1^{[k]}|\mathcal{Q}) + H(A_2^{[k]}|\mathcal{Q}) + H(A_3^{[k]}|\mathcal{Q}) + L \geq 3L$, i.e., $\rho_{total}(\mathbb{P}_3) \geq 1$.

B. Proof of Upper Bounds for $\mathbb{P}_N$ and $\mathbb{C}_N$

Recall that for $\mathbb{P}_N$, $K = N - 1$, and W_n is replicated on servers n and $n + 1$, for all $n \in [N - 1]$ while for $\mathbb{C}_N$, $\mathcal{W}_n = \{W_{n-1}, W_n\}, n \in [N]$ with $W_0 = W_N$. The next two lemmas lower bound the size of $A_{[1:N]\setminus\{n\}}^{[k]}$, conditioned on $A_n^{[k]}, W_k$ and $\mathcal{Q}$, for $\mathbb{P}_N$ and $\mathbb{C}_N$, respectively. We omit the proof of Lemma 5 since it follows similarly as that of Lemma 4.

Lemma 4 *For any feasible SPIR scheme on $\mathbb{P}_N$, we have*

$$H(A_{2:N}^{[1]}|A_1^{[1]}, W_1, \mathcal{Q}) \geq \frac{(N-2)L}{2}, \tag{49}$$

$$H(A_{1:N-1}^{[N-1]}|A_N^{[N-1]}, W_{N-1}, \mathcal{Q}) \geq \frac{(N-2)L}{2}, \tag{50}$$

$$H(A_{1:N\setminus\{n\}}^{[k]}|A_n^{[k]}, W_k, \mathcal{Q}) \geq \frac{(N-3)L}{2}, \quad k \in \{n-1, n\}, n \in \{2, \ldots, N-2\}. \tag{51}$$

Proof: To show (49), we have that $H(A_{2:N}^{[1]}|A_1^{[1]}, W_1, \mathcal{Q}) \geq H(A_{2:N}^{[1]}|A_1^{[1]}, W_1, \mathcal{R}, \mathcal{Q}) = H(A_{2:N}^{[1]}|W_1, \mathcal{R}, \mathcal{Q})$ where the equality is because $H(A_1^{[k]}|W_1, \mathcal{R}, \mathcal{Q}) = 0$. Now,

$$\begin{aligned}
& H(A_{2:N}^{[1]}|W_1, \mathcal{R}, \mathcal{Q}) \\
& \quad = \sum_{n\in[2:N]} H(A_n^{[1]}|A_{2:n-1}^{[1]}, W_1, \mathcal{R}, \mathcal{Q}) && (52)\\
& \quad \geq \sum_{n\in[2:N]} H(A_n^{[1]}|A_{2:n-1}^{[1]}, W_{1:n-1}, \mathcal{R}, \mathcal{Q}) && (53)\\
& \quad = \sum_{n\in[2:N-1]} H(A_n^{[n]}|W_{1:n-1}, \mathcal{R}, \mathcal{Q}) && (54)\\
& \quad \geq \sum_{n\in[2:N-1]} H(A_n^{[n]}|W_{\overline{n}}, \mathcal{R}, \mathcal{Q}). && (55)
\end{aligned}$$

Here (54) holds since $H(A_N^{[1]}|W_{1:N-1}, \mathcal{R}, \mathcal{Q}) = 0$ and by Lemma 2. Expanding (52) in reverse order of server indices,

$$H(A_{2:N}^{[1]}|A_1^{[1]}, W_1, \mathcal{R}, \mathcal{Q})$$

$$= \sum_{n\in[2:N]} H(A_n^{[1]}|A_{n+1:N}^{[1]}, W_1, \mathcal{R}, \mathcal{Q}) \quad (56)$$

$$\geq \sum_{n\in[2:N]} H(A_n^{[1]}|A_{n+1:N}^{[1]}, W_1, W_{n:N-1}, \mathcal{R}, \mathcal{Q}) \quad (57)$$

$$= \sum_{n\in[3:N]} H(A_n^{[n-1]}|W_1, W_{n:N-1}, \mathcal{R}, \mathcal{Q}) \quad (58)$$

$$\geq \sum_{n\in[3:N]} H(A_n^{[n-1]}|W_{\overline{n-1}}, \mathcal{R}, \mathcal{Q}). \quad (59)$$

Summing (55) and (59), and using Lemma 3, we obtain

$$2H(A_{2:N}^{[1]}|A_1^{[1]}, W_1, \mathcal{R}, \mathcal{Q})$$
$$\geq \sum_{n=2}^{N-1} H(A_n^{[n]}|W_{\overline{n}}, \mathcal{R}, \mathcal{Q}) + H(A_{n+1}^{[n]}|W_{\overline{n}}, \mathcal{R}, \mathcal{Q}) \quad (60)$$
$$\geq (N-2)L. \quad (61)$$

Proving (49) and (50) follows similarly. For (51), proceeding similarly, and using $H(A_{1:N\setminus\{n\}}^{[k]}|A_n^{[k]}, W_k, \mathcal{R}, \mathcal{Q}) \geq H(A_{1:N\setminus\{n\}}^{[k]}|W_{n-1}, W_n, \mathcal{R}, \mathcal{Q})$, $k = n-1, n$ yields the desired result. ■

To prove the upper bound for $\mathbb{P}_N$, for $k \in [N-1]$ we have

$$L = I(A_{1:N}^{[k]}; W_k|\mathcal{Q}) = H(A_{1:N}^{[k]}|\mathcal{Q}) - H(A_{1:N}^{[k]}|W_k, \mathcal{Q}) \quad (62)$$
$$= H(A_{1:N}^{[k]}|\mathcal{Q}) - H(A_n^{[k]}|\mathcal{Q}) - H(A_{1:N\setminus\{n\}}^{[k]}|A_n^{[k]}, W_k, \mathcal{Q}) \quad (63)$$
$$\leq \sum_{n'\in[N]\setminus\{n\}} H(A_{n'}^{[k]}|\mathcal{Q}) - H(A_{1:N\setminus\{n\}}^{[k]}|A_n^{[k]}, W_k, \mathcal{Q}). \quad (64)$$

Bounding the second term by Lemma 4 we obtain for all k,

$$\sum_{n'\in[N]\setminus\{n\}} H(A_{n'}^{[k]}|\mathcal{Q}) \geq \begin{cases} \frac{NL}{2}, & n = 1, N, \\ \frac{(N-1)L}{2}, & n \in [2:N-1], \end{cases} \quad (65)$$

by user privacy (5). Summing over all $n \in [N]$, we get

$$\sum_{n=1}^{N} \sum_{n'\in[N]\setminus\{n\}} H(A_{n'}^{[k]}|\mathcal{Q})$$
$$\geq 2\frac{NL}{2} + (N-2)\frac{(N-1)L}{2} = \frac{(N^2-N+2)L}{2}, \quad (66)$$

resulting in

$$(N-1)\sum_{n=1}^{N} H(A_n^{[k]}|\mathcal{Q}) \geq \frac{(N^2-N+2)L}{2}, \quad (67)$$

which gives $\frac{L}{\sum_{n=1}^N H(A_n^{[k]})} \leq \frac{L}{\sum_{n=1}^N H(A_n^{[k]}|\mathcal{Q})} \leq \frac{2(N-1)}{N^2-N+2} = \frac{2}{N+\frac{2}{N-1}}$, proving the upper bound for $\mathbb{P}_N$.

For $\mathbb{C}_N$, we need the following lemma, where $k = 0$ implies $k = N$.

Lemma 5 *For any feasible SPIR scheme on $\mathbb{C}_N$, and $n \in [N]$,*

$$H(A_{1:N\setminus\{n\}}^{[k]}|A_n^{[k]}, W_k, \mathcal{Q}) \geq \frac{(N-2)L}{2},\ k = n-1, n. \quad (68)$$

Then, (62)-(64) and Lemma 5 result in

$$(N-1)\sum_{n=1}^{N} H(A_n^{[k]}|\mathcal{Q}) \geq N\left(L + \frac{(N-2)L}{2}\right), \quad (69)$$

for any achievable scheme, which yields

$$\frac{L}{\sum_{n=1}^N H(A_n^{[k]})} \leq \frac{2(N-1)}{N^2} = \frac{2}{N+1+\frac{1}{N-1}}. \quad (70)$$

This completes the proof of Theorem 3.

REFERENCES

[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval. *Journal of the ACM*, 45(6):965–981, November 1998.

[2] H. Sun and S. A. Jafar. The capacity of private information retrieval. *IEEE Trans. Inf. Theory*, 63(7):4075–4088, March 2017.

[3] K. Banawan and S. Ulukus. The capacity of private information retrieval from coded databases. *IEEE Trans. Inf. Theory*, 64(3):1945–1956, January 2018.

[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 and Geometry*, 1(1):647–664, 2017.

[5] Q. Wang, H. Sun, and M. Skoglund. The capacity of private information retrieval with eavesdroppers. *IEEE Trans. Inf. Theory*, 65(5):3198–3214, December 2018.

[6] A. Heidarzadeh, S. Kadhe, S. El Rouayheb, and A. Sprintson. Single-server multi-message individually-private information retrieval with side information. In *IEEE ISIT*, July 2019.

[7] S. Ulukus, S. Avestimehr, M. Gastpar, S. A. Jafar, R. Tandon, and C. Tian. Private retrieval, computing, and learning: Recent progress and future challenges. *IEEE J. Sel. Areas Commun.*, 40(3):729–748, March 2022.

[8] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin. Protecting data privacy in private information retrieval schemes. *J. Computer System Sciences*, 60(3):592–629, 2000.

[9] H. Sun and S. A. Jafar. The capacity of symmetric private information retrieval. *IEEE Trans. Inf. Theory*, 65(1):322–329, June 2018.

[10] Q. Wang and M. Skoglund. Symmetric private information retrieval from MDS coded distributed storage with non-colluding and colluding servers. *IEEE Trans. Inf. Theory*, 65(8):5160–5175, March 2019.

[11] Q. Wang, H. Sun, and M. Skoglund. Symmetric private information retrieval with mismatched coded messages and randomness. In *IEEE ISIT*, July 2019.

[12] Q. Wang and M. Skoglund. On PIR and symmetric PIR from colluding databases with adversaries and eavesdroppers. *IEEE Trans. Inf. Theory*, 65(5):3183–3197, October 2019.

[13] Z. Wang, K. Banawan, and S. Ulukus. Private set intersection: A multi-message symmetric private information retrieval perspective. *IEEE Trans. Inf. Theory*, 68(3):2001–2019, November 2021.

[14] Z. Wang and S. Ulukus. Symmetric private information retrieval at the private information retrieval rate. *IEEE J. Sel. Areas Inf. Theory*, 3(2):350–361, 2022.

[15] A. M. Jafarpisheh, M. Mirmohseni, and M. A. Maddah-Ali. Distributed attribute-based private access control. In *IEEE ISIT*, June 2022.

[16] S. Meel and S. Ulukus. HetDAPAC: Distributed attribute-based private access control with heterogeneous attributes. In *IEEE ISIT*, July 2024.

[17] N. Raviv, I. Tamo, and E. Yaakobi. Private information retrieval in graph-based replication systems. *IEEE Trans. Inf. Theory*, 66(6):3590–3602, November 2019.

[18] K. Banawan and S. Ulukus. Private information retrieval from non-replicated databases. In *IEEE ISIT*, July 2019.

[19] B. Sadeh, Y. Gu, and I. Tamo. Bounds on the capacity of private information retrieval over graphs. *IEEE Trans. Inf. Forensics Security*, 18:261–273, November 2023.

[20] S. Meel, X. Kong, T. J. Maranzatto, I. Tamo, and S. Ulukus. Private information retrieval on multigraph-based replicated storage. In *IEEE ISIT*, June 2025.

[21] S. Meel and S. Ulukus. Symmetric private information retrieval (SPIR) on graph-based replicated systems. In *IEEE Globecom*, December 2025.

[22] X. Kong, S. Meel, T. J. Maranzatto, I. Tamo, and S. Ulukus. New capacity bounds for PIR on graph and multigraph-based replicated storage, 2025. Available online at arXiv:2504.20888.