Private Information Retrieval from Non-Replicated Databases

Karim Banawan Sennur Ulukus
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
kbanawan@umd.edu *ulukus@umd.edu*

***Abstract*—We consider the problem of private information retrieval (PIR) of a single message out of K messages from N non-colluding and non-replicated databases. Different from the majority of the existing literature, here, we consider the case of *non-replicated databases* under a special non-replication structure where each database stores M out of K messages and each message is stored across R different databases. This generates an R-*regular graph* structure for the storage system where the vertices of the graph are the messages and the edges are the databases. We derive a general upper bound for $M = 2$ that depends on the graph structure. We then specialize the problem to storage systems described by two special types of graph structures: *cyclic graphs* and *fully-connected graphs*. We prove that the PIR capacity for the case of cyclic graphs is $\frac{2}{K+1}$, and the PIR capacity for the case of fully-connected graphs is $\min\{\frac{2}{K}, \frac{1}{2}\}$. In both cases, the results show severe degradation in PIR capacity due to non-replication.**

I. Introduction

Private information retrieval (PIR) is a canonical problem to study the privacy of downloaded content from public databases [1]. In the classical setting, a user is interested in retrieving a message out of K messages from N replicated and non-colluding databases, such that no database can know the identity of the user's desired file. The PIR problem has become a vibrant research topic within information theory starting with [2]–[5]. In [6], Sun and Jafar characterize the PIR capacity, which is the supremum of the ratio of the number of bits of desired information L that can be retrieved privately to the total downloaded information. The fundamental limits of many variants of the problem have been investigated in [7]–[35].

A common assumption in most of these works is that the entire message set is *replicated* across all databases. However, the replication assumption may not be practical in next-generation storage systems and networks. From a storage point of view, message replication incurs high storage cost. From a network structure point of view, in next-generation networks where peer-to-peer (P2P) connections will be prevalent, nodes may not necessarily possess the same set of messages. These scenarios motivate investigating PIR in *non-replicated* storage systems. In this work, we aim at devising achievable schemes that do not rely on message replication, while being more efficient than the trivial scheme of downloading the contents of all databases. We aim at evaluating the loss in the PIR rate due to non-replication and investigating the interplay between the storage structure and the resulting PIR rate.

This work was supported by NSF Grants CNS 15-26608, CCF 17-13977 and ECCS 18-07348.

A few works have considered relaxing the replication assumption: Reference [9] investigates the case when the contents of the databases are encoded via an (N, K_c) MDS code (see also [31] for arbitrary codes). Reference [23] studies the PIR problem from homogeneous storage constrained databases. In this problem, each database is constrained to store μKL uncoded bits with $\mu \leq 1$. This problem is extended to the decentralized and heterogeneous storage settings in [32], [33], respectively. The work that is most closely related to our work here is [34]. In [34] databases store *full messages* and not portions of every message. In particular, [34] investigates the case when every message is replicated across two databases only. This storage system can be represented by a graph, in which every two databases are connected via an edge corresponding to the common message. [34] proposes an achievable PIR scheme that is immune against colluding databases, that do not form a cycle in the graph. The scheme in [34] achieves a retrieval rate of $\frac{1}{N}$. In the extended version of [34] in [35], an upper bound is proposed to show that their PIR rate is at most a factor of 2 from the optimal value for regular graphs, and the techniques are extended to larger replication factors.

In this paper, we consider PIR of a message out of K messages from N non-replicated and non-colluding databases. In our formulation, each message appears in R databases, and every database stores M different messages. We focus on the case $M = 2$. For this case, the storage system can be uniquely specified by an R-*regular graph*. In our graph, the messages correspond to the vertices and the databases correspond to the edges. This is in contrast to [34], where $R = 2$, and the roles of messages and databases are reversed on the graph. Our goal is to characterize the PIR capacity of this system.

First, we derive a general upper bound on the retrieval rate for storage systems described by R-regular graphs. Interestingly, the upper bound depends on the structure of the graph and not only on (K, R, M, N). We specialize the problem to two classes of graphs, namely, cyclic graphs and fully-connected graphs, where we obtain exact results. In *cyclic graphs*, each message (vertex) is common among two adjacent databases (edges) which are arranged in a cycle. For this type of graphs, we show that $C_{\text{PIR}} = \frac{2}{K+1}$. The achievable scheme starts from the greedy algorithm of Sun and Jafar [6] and then compresses the requests to $K - 2$ databases by

replacing the individual symbols of the scheme in [6] by sum of two messages. In *fully-connected graphs*, each vertex is connected to all of the remaining $K-1$ vertices. In this case, we show that $C_{\text{PIR}} = \min\{\frac{2}{K}, \frac{1}{2}\}$. For $K \geq 4$, we propose a novel achievable scheme, which is based on retrieving a single weighted sum of two symbols from every database. For the comparable cases with [34], our scheme outperforms their scheme in terms of the PIR rate. We note that, in both cyclic and fully-connected graph cases, the PIR capacity converges to zero as $N \to \infty$, which implies a severe degradation in the PIR efficiency due to non-replication. Due to space limitations here, proof details, extra examples, remarks, and figures can be found in the longer version in [36].

II. Problem Formulation

Consider the problem of PIR from N non-replicated and non-colluding databases. We denote the databases by $\mathcal{D} = \{D_1, D_2, \cdots, D_N\}$. The storage system stores K messages in total, each message is stored across R different databases, and each database stores locally M different messages. We denote the message set by $\mathcal{W} = \{W_1, W_2, \cdots, W_K\}$. Each message $W_k \in \mathbb{F}_q^L$ is a vector of length L picked in an i.i.d. fashion from a sufficiently large finite field $\mathbb{F}_q^L$, i.e., in q-ary units

$$H(W_k) = L,\ k \in \{1, \cdots, K\}, \quad H(\mathcal{W}) = KL \tag{1}$$

The storage system is parameterized by (K, R, M, N), where $KR = MN$. In this work, we focus on the case $M = 2$. We characterize the storage system by a (V, E) graph, where $V = \mathcal{W} = \{W_1, W_2, \cdots, W_K\}$ is the set of vertices, and $E = \mathcal{D} = \{D_1, D_2, \cdots, D_N\}$ is the set of edges. An edge D_j drawn between messages W_m and W_k means that the contents of database D_j is $Z_j = \{W_m, W_k\}$. This graph is an *R-regular graph*, since each message is repeated R times across the storage system. In the following, we define specific parameters of the graph, which are needed for the converse.

Definition 1 (Graph reduction) *The graph $(V, E) = (\mathcal{W}, \mathcal{D})$ is reduced iteratively starting with the vertex W_1 by enumerating all the edges connecting to W_1, and removing all neighboring vertices connected to enumerated edges except one, which we denote by $\tilde{W}_2$. The process of enumerating edges and removing corresponding neighbors iteratively continues until one vertex is left $\tilde{W}_{\kappa+1}$ after κ reductions.*

Definition 2 (Spread of the graph) *The spread of a graph δ is the largest sequence of edges (databases) that results from the graph reduction procedure given in Definition 1.*

Definition 3 (Cyclic graphs) *The graph $(V, E) = (\mathcal{W}, \mathcal{D})$ is called a cyclic graph if each two adjacent vertices are connected by an edge and no non-adjacent vertices are connected by an edge, i.e., the contents of the databases are:*

$$Z_1 = \{W_1, W_2\},\ Z_2 = \{W_2, W_3\}, \cdots, Z_N = \{W_N, W_1\} \tag{2}$$

The cyclic graph is parameterized by $(K, R, M, N) = (K, 2, 2, K)$, and the spread of the graph is $\delta = K - 1$.

Definition 4 (Fully-connected graphs) *The graph $(V, E) = (\mathcal{W}, \mathcal{D})$ is called fully-connected if every two vertices are connected by a unique edge. Hence, the contents of the databases can be written as the $\binom{K}{2}$ subsets of $\{1, \cdots, K\}$ with 2 elements. The fully-connected graph is parameterized by $(K, R, M, N) = (K, K-1, 2, \binom{K}{2})$, and the spread of the graph is $\delta = K - 1$.*

In PIR, the user wants to retrieve a message W_k without leaking any information about k to any individual database. The user sends N queries to the databases, one query to each database. These queries are independent of the messages as the user has no information about the messages, hence,

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

The databases respond to the user queries by answer strings $A_{1:N}^{[k]}$. The answer string $A_n^{[k]}$ is a deterministic function of the query $Q_n^{[k]}$ and the contents of the nth database Z_n, therefore,

$$H(A_n^{[k]} | Q_n^{[k]}, Z_n) = 0, \quad n \in [N] \tag{4}$$

To ensure privacy, the retrieval strategy intended to retrieve W_i must be indistinguishable from the retrieval strategy intended to retrieve W_j for any i and j, i.e.,

$$(Q_n^{[i]}, A_n^{[i]}, \mathcal{W}) \sim (Q_n^{[j]}, A_n^{[j]}, \mathcal{W}), \quad n \in [N],\ i, j \in [K] \tag{5}$$

where $\sim$ denotes statistical equivalence.

The user needs to be able to reconstruct W_k perfectly from the collected answers, i.e.,

$$H(W_k | Q_{1:N}^{[k]}, A_{1:N}^{[k]}) = 0 \tag{6}$$

An achievable retrieval scheme satisfies (5), (6) for some message length L. The retrieval rate is the ratio between the length of the desired message L and the total download,

$$R_{\text{PIR}} = \frac{L}{\sum_{n=1}^{N} H(A_n^{[k]})} \tag{7}$$

The PIR capacity is the largest PIR rate over all achievable schemes, i.e., $C_{\text{PIR}} = \sup R_{\text{PIR}}$.

III. Main Results

In this section, we present the main results of this paper. Our first result is a general upper bound for storage systems defined by R-regular graphs with $M = 2$ and arbitrary (K, R, N) which is given in the following theorem. The proof of Theorem 1 is given in Section IV.

Theorem 1 (Upper-bound for R-regular graphs) *In an R-regular graph storage system with $(K, R, M, N) = (K, R, 2, N)$, the retrieval rate is upper bounded by*

$$R_{PIR} \leq \min\left\{\frac{R}{N}, \frac{1}{1 + \frac{\delta}{R}}\right\} \tag{8}$$

In the following, we characterize the PIR capacity of cyclic graphs and fully-connected graphs. The converse proofs are

corollaries of Theorem 1 (see [36, Remark 2]). The achievability proofs of Theorems 2 and 3 are given in Section V.

Theorem 2 (Capacity of cyclic graphs) *For a cyclic graph storage system, the PIR capacity is* $C_{PIR} = \frac{2}{K+1}$.

Theorem 3 (Capacity of fully-connected graphs) *For a fully-connected graph storage system with* $M = 2$*, the PIR capacity is given by*

$$C_{PIR} = \begin{cases} \frac{1}{2}, & K = 2, 3 \\ \frac{2}{K}, & K \geq 4 \end{cases} \tag{9}$$

IV. Converse Proof

In this section, we prove Theorem 1. Let $\mathcal{Q}$ denote the collection of all queries to all databases for all desired messages,

$$\mathcal{Q} \triangleq \left\{ Q_n^{[k]} : k \in [K],\ n \in [N] \right\} \tag{10}$$

We assume that the retrieval scheme is symmetric across databases (see [36, eqn. (13)]). This is without loss of generality, since any asymmetric scheme can be transformed into a symmetric one by means of time-sharing without changing the retrieval rate. We need the following lemma (proof in [36]).

Lemma 1 *Let* $\mathcal{R}_m$ *denote the set of databases containing message* W_m*, then*

$$H(A_n^{[m]} | \mathcal{W} \setminus \{W_m\}, \mathcal{Q}) \geq \frac{L}{R}, \quad n \in \mathcal{R}_m \tag{11}$$

We are now ready to prove Theorem 1. We first prove that $R_{\text{PIR}} \leq \frac{R}{N}$. From Lemma 1, we have

$$L \leq RH(A_n^{[m]} | \mathcal{W} \setminus \{W_m\}, \mathcal{Q}), \quad n \in \mathcal{R}_m \tag{12}$$

$$= \frac{R}{N} NH(A_n^{[m]} | \mathcal{W} \setminus \{W_m\}, \mathcal{Q}) \leq \frac{R}{N} \sum_{n=1}^{N} H(A_n^{[m]} | \mathcal{Q}) \tag{13}$$

where (13) follows from the database symmetry. Therefore,

$$R_{\text{PIR}} = \frac{L}{\sum_{n=1}^{N} H(A_n^{[m]})} \leq \frac{L}{\sum_{n=1}^{N} H(A_n^{[m]} | \mathcal{Q})} \leq \frac{R}{N} \tag{14}$$

Next, we prove that $R_{\text{PIR}} \leq \frac{1}{1+\frac{\delta}{R}}$, where δ is the spread of the graph. To obtain the spread of the graph, we perform the graph reduction as in Definition 1. Let $W_1, \tilde{W}_2, \tilde{W}_3, \cdots$ be the remaining messages at every graph reduction (the leading message), respectively, and W_{n_i} be a message in the n_ith database that is different from the leading message, then,

$$L = H(W_1) \tag{15}$$

$$= H(W_1 | \mathcal{Q}) - H(W_1 | A_{1:N}^{[1]}, \mathcal{Q}) \tag{16}$$

$$= I(W_1; A_{1:N}^{[1]} | \mathcal{Q}) \tag{17}$$

$$= H(A_{1:N}^{[1]} | \mathcal{Q}) - H(A_{1:N}^{[1]} | W_1, \mathcal{Q}) \tag{18}$$

$$\leq NH(A_1^{[1]} | \mathcal{Q}) - H(A_{\Delta}^{[1]} | W_1, \mathcal{Q}) \tag{19}$$

$$= NH(A_1^{[1]} | \mathcal{Q}) - \sum_{i=1}^{\delta} H(A_i^{[1]} | W_1, \mathcal{Q}, A_{1:i-1}^{[1]}) \tag{20}$$

$$\leq NH(A_1^{[1]} | \mathcal{Q}) - \sum_{i=1}^{\delta} H(A_i^{[1]} | W_1, \mathcal{W} \setminus \{W_{n_i}\}, \mathcal{Q}, A_{1:i-1}^{[1]}) \tag{21}$$

$$= NH(A_1^{[1]} | \mathcal{Q}) - \sum_{i=1}^{\delta} H(A_i^{[1]} | W_1, \mathcal{W} \setminus \{W_{n_i}\}, \mathcal{Q}) \tag{22}$$

where (16) follows from the reliability constraint and the independence of queries and messages, (19) follows from the independence bound and non-negativity of the entropy function where $A_{\Delta}^{[1]}$ denotes the answer strings returned by the sequence of the databases that define the spread of the graph, and (21) follows from the fact that conditioning on $\mathcal{W} \setminus \{W_i\}$ cannot increase entropy. To show (22), we note that the leading message at the jth graph reduction $\tilde{W}_j$ belongs to the set of the connected messages in the $(j-1)$th graph reduction, and at the jth graph reduction, the nodes connecting to $\tilde{W}_j$ are removed from the graph, we have $\{\tilde{W}_2, \tilde{W}_3, \cdots, \tilde{W}_{j(i)}\} \subseteq \{W_{n_1}, W_{n_2}, \cdots, W_{n_{i-1}}\} \subseteq \mathcal{W} \setminus \{W_{n_i}\}$, where $j(i)$ is the index of the leading message in the ith database. Hence, we drop $A_{1:i-1}^{[1]}$ as they are deterministic functions of $(\mathcal{Q}, W_1, \mathcal{W} \setminus \{W_{n_i}\})$.

Now, we have

$$L \leq NH(A_1^{[1]} | \mathcal{Q}) - \sum_{i=1}^{\delta} H(A_i^{[1]} | W_1, \mathcal{W} \setminus \{W_{n_i}\}, \mathcal{Q}) \tag{23}$$

$$= NH(A_1^{[1]} | \mathcal{Q}) - \sum_{i=1}^{\delta} H(A_i^{[n_i]} | W_1, \mathcal{W} \setminus \{W_{n_i}\}, \mathcal{Q}) \tag{24}$$

$$\leq NH(A_1^{[1]} | \mathcal{Q}) - \frac{\delta L}{R} \tag{25}$$

where (24) follows from the privacy constraint, and (25) follows from Lemma 1. Reordering terms, we have

$$R_{\text{PIR}} = \frac{L}{\sum_{n=1}^{N} H(A_n^{[m]})} \leq \frac{L}{NH(A_1^{[1]} | \mathcal{Q})} \leq \frac{1}{1 + \frac{\delta}{R}} \tag{26}$$

which together with (14) concludes the proof of Theorem 1.

V. Achievability Proof

We show the basic ingredients of the achievable scheme by a motivating example of $(K, R, M, N) = (3, 2, 2, 3)$. In fact, the graph for this example is both cyclic and fully-connected (see Fig. 1). Then, we present general capacity-achieving schemes for cyclic graphs and fully-connected graphs.

A. Motivating Example: $K = 3$, $R = 2$, $M = 2$, $N = 3$

In this example, we consider a storage system that consists of $N = 3$ databases. The system stores $K = 3$ messages in total, namely W_1, W_2, W_3. Each message is replicated across $R = 2$ databases, such that each database stores $M = 2$ messages (see Table I). This is a cyclic and also a fully-connected graph as shown in Fig. 1.

Without loss of generality, assume that the desired message is W_1. To construct the capacity-achieving scheme, the user randomly permutes the indices of the symbols of W_1, W_2, W_3 independently, uniformly, and privately from the

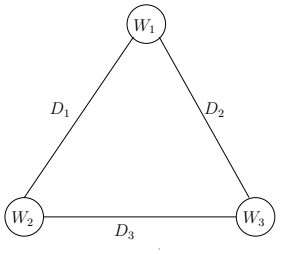

Fig. 1: Graph structure for the $(3, 2, 2, 3)$ system.

Database 1 (D_1)	Database 2 (D_2)	Database 3 (D_3)
W_1	W_1	W_2
W_2	W_3	W_3

TABLE I: Contents of databases for the $(3, 2, 2, 3)$ system.

databases. Denote the permuted version of W_1 by the vector $(a_1, \cdots, a_L)$, the permuted version of W_2 by $(b_1, \cdots, b_L)$, and the permuted version of W_3 by $(c_1, \cdots, c_L)$. Pick $L = 12$.

A straightforward solution for this problem is to apply Sun and Jafar scheme in [6]. Since every database contains $M = 2$ messages, the user downloads a single bit from each message from each database in round 1, i.e., the user downloads a_1, b_1 from database 1, a_2, c_1 from database 2, and b_2, c_2 from database 3. Now, the user exploits b_2, c_2 as side information by downloading $a_3 + b_2$ from database 1, and $a_4 + c_2$ from database 2. Finally, the user downloads the sum $b_3 + c_3$ from database 3. The query table for this scheme is shown in Table II. Note that although the sum $b_3 + c_3$ is irrelevant to the decodability of W_1, the user needs to download it to satisfy the privacy constraint. With this scheme, the user downloads 4 bits from W_1 out of the total 9 downloads, hence $R_{\text{PIR}} = \frac{4}{9}$.

Database 1	Database 2	Database 3
a_1 b_1	a_2 c_1	b_2 c_2
$a_3 + b_2$	$a_4 + c_2$	$b_3 + c_3$

TABLE II: Sun and Jafar scheme for the $K = 3$, $R = 2$, $M = 2$, $N = 3$ example.

Although this scheme outperforms the scheme in [34] in terms of the retrieval rate, there is room for improving it. The main source of inefficiency of the scheme is the downloads from database 3, as the user downloads 3 bits and exploits only 2 of them. Moreover, the user downloads new independent bit $b_3 + c_3$. If the user introduces *dependency* to the downloads of database 3, the user may *compress* the requests from database 3, and improve the retrieval rate. In order to do this, the user downloads the sums $b_1 + c_2$ and $b_2 + c_1$ from database 3. For the decodability, the user can decode c_2 by canceling b_1 from $b_1 + c_2$ and b_2 by canceling c_1 from $b_2 + c_1$. Therefore, a_3, a_4 are decodable by canceling b_2 and c_2.

Nevertheless, this scheme is *not private* because the user still downloads 2 bits from database 3 in the form of sum of 2 bits. To remedy this problem, the user should repeat the compression of the downloads over all databases. Hence, in repetition 2, the user compresses the downloads from database 2 and downloads $a_7 + c_3$, $a_8 + c_4$. Similarly, in repetition 3, the user downloads $a_{10} + b_5$ and $a_{11} + b_6$ from database 1. The complete query structure is given in Table III.

	Database 1	Database 2	Database 3
rep. 1	a_1 b_1	a_2 c_1	
	$a_3 + b_2$	$a_4 + c_2$	$b_1 + c_2$ $b_2 + c_1$
rep. 2	a_5 b_3		b_4 c_3
	$a_6 + b_4$	$a_7 + c_3$ $a_8 + c_4$	$b_3 + c_4$
rep. 3		a_9 c_5	b_5 c_6
	$a_{10} + b_5$ $a_{11} + b_6$	$a_{12} + c_6$	$b_6 + c_5$

TABLE III: Complete query structure for the capacity-achieving scheme for $K = 3$, $R = 2$, $M = 2$, $N = 3$.

Since, the query structure is now symmetric across the databases, and the indices of the bits from each message are chosen uniformly, independently and privately, all queries are equally likely, and the scheme is private. The scheme is decodable as each repetition is decodable separately. As we discussed above, $a_1, \cdots, a_4$ are decodable in repetition 1. For repetition 2, a_5 is decodable directly, a_6 is decodable by canceling b_4 from $a_6 + b_4$, and a_7 is decodable by canceling c_3 from $a_7 + c_3$. Finally, c_4 is decodable by canceling b_3 from $b_3 + c_4$ and therefore a_8 is decodable by further canceling c_4 from $a_8 + c_4$. The decodability of repetition 3 follows in a similar way to the decodability of repetition 2 by exchanging the roles of W_2, W_3. The user downloads 12 bits from W_1 out of a total of 24 downloads. Consequently, $R_{\text{PIR}} = \frac{12}{24} = \frac{1}{2}$ which matches the upper bound in Theorem 1.

B. General Achievability for Cyclic Graphs

The new ingredient in this scheme is the *compression* of the queries submitted for a subset of the databases. To satisfy the privacy constraint, the user performs the scheme along $\binom{N}{2} = \binom{K}{2}$ repetitions. In each repetition, the user chooses to submit the full query (according to [6]) to 2 databases. For the remaining databases, the user downloads two symbols in the form of 2-sums. The scheme works with $L = 4\binom{K}{2}$ symbols. The general scheme for cyclic graphs can be summarized as:

1) *Index preparation:* The symbol indices of each message are permuted independently, uniformly, and privately.
2) *Constructing full queries:* We apply the scheme of Sun and Jafar [6] to construct the full queries to all databases. We apply this scheme over blocks of $\tilde{L} = 4$. The user downloads 1 individual symbol from each message from each database in round 1. Next, the user downloads a 2-sum from the stored messages in each database. This sum exploits the side information generated from other databases. Note that since $R = 2$ in this graph, the user can generate 1 side information equation for each

database. Another change from [6] is that even for the $K-2$ databases that do not contain the desired message, the user exploits the side information generated at other databases by introducing dependency to the answers.

3) *Compressing queries:* The user chooses different $K-2$ databases at each repetition. The user downloads the sum of the individual symbols in round 1.
4) Repeat step 2, 3 over new symbols for $\binom{K}{2}$ repetitions.

C. General Achievability for Fully-Connected Graphs

For $K=2$, we have 1 database containing 2 messages; the capacity-achieving scheme is simply to download the contents of the entire database, hence $R_{\text{PIR}}=\frac{1}{2}$. For $K=3$, the capacity-achieving scheme is exactly the motivating example in Section V-A, hence $R_{\text{PIR}}=\frac{1}{2}$.

For $K\geq 4$, the upper bound $R_{\text{PIR}}\leq\frac{2}{K}$ is the active upper bound. The general achievability for this case is given below. The achievable scheme works with $L=R=K-1$ symbols from $\mathbb{F}_q$, where q is sufficiently large and is prime.

1) *Index preparation:* The symbol indices of each message are permuted independently, uniformly, and privately.
2) *Retrieval from database 1:* Denote the permuted contents of the nth database by $Z_n=\left\{X_1^{(n)},X_2^{(n)}\right\}$. Without loss of generality, assume that $X_1^{(1)}$ is the permuted version of the desired message. From database 1, the user downloads a weighted sum of two symbols from the two messages, i.e., the user downloads $\alpha_1^{(1)}X_1^{(1)}(1)+\alpha_2^{(1)}X_2^{(1)}(1)$ from database 1, where $\alpha_m^{(n)}\in\mathbb{F}_q$, $m\in\{1,2\}, n\in[N]$.
3) *Exploiting side information:* The user downloads different weighted sums from every database. If the nth database contains the desired message, the user downloads a new desired symbol in the sum. If the message stored in the nth database is undesired, the user exploits the same message symbol in all databases. I.e., the user downloads the weighted sum $\alpha_1^{(n)}X_1^{(n)}(i)+\alpha_2^{(1)}X_2^{(1)}(j)$, where indices i,j are chosen depending on the message.
4) *Database symmetry:* Repeat step 3 across all databases.

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] N. B. Shah, K. V. Rashmi, and K. Ramchandran. One extra bit of download ensures perfectly private information retrieval. In *IEEE ISIT*, June 2014.

[3] T. Chan, S. Ho, and H. Yamamoto. Private information retrieval for coded storage. In *IEEE ISIT*, June 2015.

[4] A. Fazeli, A. Vardy, and E. Yaakobi. Codes for distributed PIR with low storage overhead. In *IEEE ISIT*, June 2015.

[5] R. Tajeddine and S. El Rouayheb. Private information retrieval from MDS coded data in distributed storage systems. In *IEEE ISIT*, July 2016.

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

[7] H. Sun and S. A. Jafar. The capacity of robust private information retrieval with colluding databases. *IEEE Trans. on Info. Theory*, 64(4):2361–2370, April 2018.

[8] H. Sun and S. A. Jafar. The capacity of symmetric private information retrieval. *IEEE Trans. on Info. Theory*, 65(1):322–329, January 2019.

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

[10] Q. Wang and M. Skoglund. Symmetric private information retrieval for MDS coded distributed storage. In *IEEE ICC*, May 2017.

[11] R. Freij-Hollanti, O. Gnilke, C. Hollanti, and D. Karpuk. Private information retrieval from coded databases with colluding servers. *SIAM Journal on Applied Algebra and Geometry*, 1(1):647–664, 2017.

[12] K. Banawan and S. Ulukus. Multi-message private information retrieval: Capacity results and near-optimal schemes. *IEEE Trans. on Info. Theory*, 64(10):6842–6862, October 2018.

[13] K. Banawan and S. Ulukus. The capacity of private information retrieval from byzantine and colluding databases. *IEEE Trans. on Info. Theory*, 65(2):1206–1219, February 2019.

[14] Y.-P. Wei, K. Banawan, and S. Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. *IEEE Trans. on Info. Theory*, 65(5):3215–3232, May 2019.

[15] Y.-P. Wei, K. Banawan, and S. Ulukus. Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits. *IEEE JSAC*, 36(6):1126–1139, June 2018.

[16] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson. Private information retrieval with side information. Available at arXiv:1709.00112.

[17] Z. Chen, Z. Wang, and S. Jafar. The capacity of private information retrieval with private side information. Available at arXiv:1709.03022.

[18] Y.-P. Wei, K. Banawan, and S. Ulukus. The capacity of private information retrieval with partially known private side information. *IEEE Trans. on Info. Theory*. Submitted November 2017. Also available at arXiv:1710.00809.

[19] Y.-P. Wei and S. Ulukus. The capacity of private information retrieval with private side information under storage constraints. *IEEE Trans. on Info. Theory*. Submitted November 2018. Also available at arXiv:1806.01253.

[20] S. Li and M. Gastpar. Single-server multi-message private information retrieval with side information. Available at arXiv:1808.05797.

[21] M. Mirmohseni and M. A. Maddah-Ali. Private function retrieval. Available at arXiv:1711.04677.

[22] Z. Chen, Z. Wang, and S. Jafar. The asymptotic capacity of private search. In *IEEE ISIT*, June 2018.

[23] M. A. Attia, D. Kumar, and R. Tandon. The capacity of private information retrieval from uncoded storage constrained databases. Available at arXiv:1805.04104v2.

[24] K. Banawan and S. Ulukus. Asymmetry hurts: Private information retrieval under asymmetric-traffic constraints. *IEEE Trans. on Info. Theory*. Submitted January 2018. Also available at arXiv:1801.03079.

[25] K. Banawan and S. Ulukus. Private information retrieval through wiretap channel II: Privacy meets security. *IEEE Trans. on Info. Theory*. Submitted January 2018. Also available at arXiv:1801.06171.

[26] K. Banawan and S. Ulukus. Noisy private information retrieval: Separability of channel coding and information retrieval. *IEEE Trans. on Info. Theory*. Submitted July 2018. Also available at arXiv: 1807.05997.

[27] Z. Jia, H. Sun, and S. Jafar. Cross subspace alignment and the asymptotic capacity of X-secure T-private information retrieval. Available at arXiv:1808.07457.

[28] C. Tian, H. Sun, and J. Chen. Capacity-achieving private information retrieval codes with optimal message size and upload cost. Available at arXiv:1808.07536.

[29] R. Bitar and S. El Rouayheb. Staircase-PIR: Universally robust private information retrieval. Available at arXiv:1806.08825.

[30] S. Kumar, A. G. i Amat, E. Rosnes, and L. Senigagliesi. Private information retrieval from a cellular network with caching at the edge. Available at arXiv:1809.00872.

[31] S. Kumar, H.-Y. Lin, E. Rosnes, and A. G. i Amat. Achieving maximum distance separable private information retrieval capacity with linear codes. Available at arXiv:1712.03898.

[32] Y.-P. Wei, B. Arasli, K. Banawan, and S. Ulukus. The capacity of private information retrieval from decentralized uncoded caching databases. *IEEE Trans. on Info. Theory*. Submitted November 2018. Also available at arXiv:1811.11160.

[33] K. Banawan, B. Arasli, Y.-P. Wei, and S. Ulukus. The capacity of private information retrieval from heterogeneous uncoded caching databases. *IEEE Trans. on Info. Theory*. Submitted February 2019. Also available at arXiv:1902.09512.

[34] N. Raviv and I. Tamo. Private information retrieval in graph based replication systems. In *IEEE ISIT*, June 2018.

[35] N. Raviv, I. Tamo, and E. Yaakobi. Private information retrieval in graph based replication systems. Available at arXiv:1812.01566.

[36] K. Banawan and S. Ulukus. Private information retrieval from non-replicated databases. Available at arXiv:1901.00004.