Private Cache-aided Interference Alignment for Multiuser Private Information Retrieval

Xiang Zhang*, Kai Wan†, Hua Sun‡, Mingyue Ji*, and Giuseppe Caire†
Department of Electrical and Computer Engineering, University of Utah*
Department of Electrical Engineering and Computer Science, Technische Universität Berlin†
Department of Electrical Engineering, University of North Texas‡
Email: *{xiang.zhang, mingyue.ji}@utah.edu, †{kai.wan, caire}@tu-berlin.de, ‡hua.sun@unt.edu

***Abstract*—In the problem of cache-aided Multiuser Private Information Retrieval (MuPIR), a set of K_u cache-aided users wish to download their desired messages from a set of N distributed non-colluding databases each holding a library of K independent messages. The communication load of this problem is defined as the total number of bits downloaded (normalized by the message length) by the users. The goal is to find the optimal memory-load trade-off under the constraint of user demand privacy, which ensures that any individual database learns nothing about the demands of the users. In this paper, for the MuPIR problem with $K = 2$ messages, $K_u = 2$ users and $N \geq 2$ databases, we provide achievability for the memory-load pairs $(\frac{N-1}{2N}, \frac{N+1}{N})$ and $(\frac{2(N-1)}{2N-1}, \frac{N+1}{2N-1})$ by constructing specific achievable schemes based on the novel idea of *Private Cache-aided Interference Alignment (PCIA)*. We prove that the proposed scheme is optimal if the cache placement is uncoded (i.e., users directly cache a subset of the library bits). Computer-aided investigation also shows that the proposed schemes are optimal in general when $N = 2, 3$.**

I. Introduction

Introduced by Chor *et al.* in 1995 [1], the problem of private information retrieval (PIR) seeks the most efficient way for a user to retrieve a desired message from N distributed databases (each holding a library of K messages) while keeping the desired message identity private from the databases. Sun and Jafar recently characterized the capacity of the PIR problem as $C = (1 + \frac{1}{N} + \cdots \frac{1}{N^{K-1}})^{-1}$ [2], [3], which is achieved by leveraging the idea of Blind Interference Alignment (BIA) originally introduced to exploit the diversity of coherence patterns in wireless interference networks [4].

Coded caching was originally proposed by Maddah-Ali and Niesen (MAN) in [5] for a shared-link caching system containing a server with a library of K equal-length files, which is connected to K_u users through a noiseless shared-link, each of which can store M files in its cache. Each user demands one file in the delivery phase. The MAN scheme uses a combinatorial design in the placement phase so that during the delivery phase multicast messages can simultaneously satisfy the demands of multiple users. It was proved in [6] that the MAN caching scheme is order optimal within a factor of 2 when $K \geq K_u$. In addition, the optimal memory-load tradeoff was characterized in [5] for the case where $K = K_u = 2$.[1] Under the constraint of uncoded cache placement and for worst-case load, the MAN scheme was proved to be optimal when $K \geq K_u$ [7].

Characterization of the optimal memory-load trade-off for the *cache-aided PIR problem*, in which the effect of caching is taken into account, has gained significant attentions recently. Two different privacy models are commonly considered. In one line of research [8]–[10], the *user-against-database* privacy model is studied where individual databases are prevented from learning the single-user's demand. The author in [8] studied the case where a single cache-aided user is connected to a set of N replicated databases and showed that memory sharing between the memory-load pairs $(0, 1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}})$ and $(K, 0)$ (i.e., split the messages and cache memory proportionally and then implement two PIR schemes on independent parts of the messages) is actually optimal if the databases are aware of the user's cached content. However, if the databases are unaware of the user's cached content, then there is an "unawareness gain" in capacity as shown in [9], [10]. Another line of research [11]–[13] deals with the *user-against-user* privacy model where users are prevented from learning each other's demands. The authors in [11] first formulated the *coded caching with private demands* problem where a shared-link caching system with demand privacy, i.e., any user should not learn anything about the demands of other users, was considered. The goal is to design efficient delivery schemes such that the communication load is minimized while preserving user demand privacy. Order optimal schemes were proposed based on the novel concept of *virtual user*. In [14], the authors studied the subpacketizaiton issues for this problem. Later, coded caching with private demands was extended to the Device-to-Device (D2D) network [12]. In general, the exact capacity characterization remains open for these problems.

This paper formulates the problem of cache-aided multiuser PIR (MuPIR), where each of the K_u cache-equipped users aims to retrieve a message from N distributed databases while preserving the privacy of user demands given that the cached content of the users are known to the databases. Based on the novel idea of *Private Cache-aided Interference Alignment (PCIA)*, we construct cache placement and private delivery phases achieving the non-trivial memory-load pairs $(\frac{N-1}{2N}, \frac{N+1}{N})$ and $(\frac{2(N-1)}{2N-1}, \frac{N+1}{2N-1})$ for the MuPIR problem with $K = 2$ messages, $K_u = 2$ users and $N \geq 2$ databases.

[1] To the best of our knowledge, for the coded caching problem, this is the only non-trivial case (i.e., $\min\{K, K_u\} \geq 2$) where the exact optimality is characterized for arbitrary memory size.

ISBN 978-3-903176-29-4 © 2020 IFIP

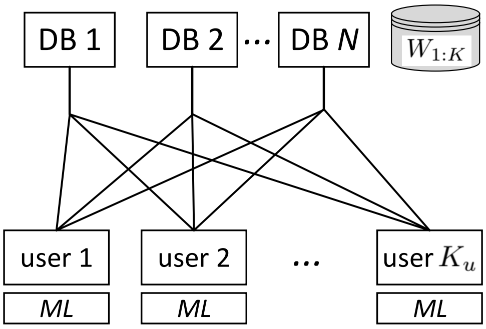

Fig. 1. Cache-aided MuPIR system with N replicated databases, K independent messages and K_u cache-equipped users. The users are connected to each DB via an error-free shared-link broadcast channel.

Different from the existing cache-aided interference algiment schemes in [15]–[18] which were designed for the cache-aided interference channels, our proposed private caching scheme is to let each server deliver symmetric messages (in order to keep privacy), each of which contains some uncached and undesired symbols (interferences) of each user. Our main strategy is to align these interferences for each user. We prove that the proposed scheme is optimal under the constraint of uncoded cache placement. Computer-aided investigation also shows that the proposed schemes are optimal in general when $N = 2, 3$.

Notation Convention: $|\cdot|$ represents the cardinality of a set. $[n] := \{1, 2, \cdots, n-1, n\}$, $[m : n] := \{m, m+1, m+2, \cdots, n\}$ and $(m : n) := (m, m+1, \cdots, n)$ for some integers $m \le n$. For two sets $\mathcal{A}$ and $\mathcal{B}$, let $\mathcal{A}\backslash\mathcal{B} := \{x \in \mathcal{A} : x \notin \mathcal{B}\}$. For an index set $\mathcal{I}$, the notation $A_{\mathcal{I}}$ represents the set $\{A_i : i \in \mathcal{I}\}$. When $\mathcal{I} = [m : n]$, we write $A_{[m:n]}$ as $A_{m:n}$ for simplicity. For an index vector $I = (i_1, i_2, \cdots, i_n)$, the notation A_I represents a new vector $A_I := (A_{i_1}, A_{i_2}, \cdots, A_{i_n})$. The operator $\oplus$ denotes the bit-wise XOR. $\mathbf{0}_n$ denotes the all-zero vector with length n, i.e., $\mathbf{0}_n := [\underbrace{0, 0, 0, \cdots, 0}_{n \text{ terms}}]$.

II. Problem Formulation

We consider a system with K_u users, each of which aims to privately retrieve a message from $N \ge 2$ replicated (non-colluding) databases (DBs). Each DB stores K independent messages, denoted by $W_1, W_2, \cdots, W_K$, each of which is uniformly distributed over $[2^L]$. Each user is equipped with a cache memory of size ML bits, where $0 \le M \le K$. Let the random variables $Z_1, Z_2, \cdots, Z_{K_u}$ denote the cached content of the users. The system operates in two phases, a *cache placement phase* followed by a *private delivery phase*. In the cache placement phase, the users fill up their cache memory without the knowledge of their future demands. It is assumed that the cached content of each user is a deterministic function of the messages $W_{1:K}$ and is known to all DBs. In the private delivery phase, each user $k \in [K_u]$ wishes to retrieve a message W_{θ_k} where $\theta_k \in [K]$. Let $\boldsymbol{\theta} := (\theta_1, \theta_2, \cdots, \theta_{K_u})$ denote the demands of the users. We assume that $\boldsymbol{\theta}$ is generated according to the uniform distribution over $[K]^{K_u}$. Depending on $\boldsymbol{\theta}$ and $Z_1, Z_2, \cdots, Z_{K_u}$, users cooperatively generate N queries $Q_1^{[\boldsymbol{\theta}]}, Q_2^{[\boldsymbol{\theta}]}, \cdots, Q_N^{[\boldsymbol{\theta}]}$, and then send the query $Q_n^{[\boldsymbol{\theta}]}$ to DB n. Upon receiving the query, DB n responds with an answer $A_n^{[\boldsymbol{\theta}]}$ broadcasted to all users. The answer $A_n^{[\boldsymbol{\theta}]}$ is a deterministic function of the query received by DB n, i.e., $Q_n^{[\boldsymbol{\theta}]}$, and the information available to DB n, i.e., $W_{1:K}$ and $Z_{1:K}$. Therefore,

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

After collecting all the answers from the N DBs, the users should be able to recover their desired messages correctly with the help of their caches. This decodability requirement can be written as $\forall k \in [K_u]$:

$$H(W_{\theta_k} | Q_{1:N}^{[\boldsymbol{\theta}]}, A_{1:N}^{[\boldsymbol{\theta}]}, Z_k) = \varepsilon_1, \tag{2}$$

where $\varepsilon_1 \to 0$ as $L \to \infty$. When $\varepsilon_1 = 0$, this corresponds to the zero decoding error case.

To preserve the privacy of the users' demands, from the viewpoint of any individual DB, the demand vector $\boldsymbol{\theta}$ should be independent of all the information available to that DB, i.e., the following privacy constraint should be satisfied $\forall n \in [N]$:

$$I(\boldsymbol{\theta}; Q_n^{[\boldsymbol{\theta}]}, A_n^{[\boldsymbol{\theta}]}, W_{1:K}, Z_{1:K}) = \varepsilon_2 \tag{3}$$

where $\varepsilon_2 \to 0$ as $L \to \infty$.

The *load* (or transmission rate) of the MuPIR problem, denoted by R, is defined as the average (over random demands) number of bits downloaded from the DBs per useful message bit. Let D denote the total number of bits broadcasted from the DBs, then

$$R := \frac{D}{L} = \frac{\sum_{n=1}^{N} H(A_n^{[\boldsymbol{\theta}]})}{L} \tag{4}$$

Note that R does not depend on $\boldsymbol{\theta}$, otherwise this leaks information of the user demands to the DBs. A memory-load pair (M, R) is said to be achievable if there exists a MuPIR scheme satisfying the decodability constraint (2) and the privacy constraint (3). The goal of the MuPIR problem is to design the cache placement and the corresponding private delivery phases such that the load is minimized. For any $0 \le M \le K$, let $R^\star(M)$ denote the minimal achievable load. In addition, if users directly cache a subset of the library bits, the placement phase is referred to as *uncoded cache placement*. We denote the minimum achievable load under the constraint of uncoded cache placement by $R^\star_{\mathrm{u}}(M)$.

III. Main Result

In this section we present the main results of this paper.

Theorem 1: For the cache-aided MuPIR problem with $K = 2$ messages, $K_u = 2$ users and $N \ge 2$ DBs, the following load is achievable

$$R(M) = \begin{cases} 2(1-M), & 0 \le M \le \frac{N-1}{2N} \\ \frac{(N+1)(3-2M)}{2N+1}, & \frac{N-1}{2N} \le M \le \frac{2(N-1)}{2N-1} \\ \frac{(N+1)(2-M)}{2N}, & \frac{2(N-1)}{2N-1} \le M \le 2 \end{cases} \tag{5}$$

Proof: The achievability proof of Theorem 1 is presented in Section IV where specific schemes are constructed based on the novel idea of PCIA achieving the memory-load pairs $(\frac{N-1}{2N}, \frac{N+1}{N})$ and $(\frac{2(N-1)}{2N-1}, \frac{N+1}{2N-1})$. In addition, the memory-load pairs $(0,2)$ and $(2,0)$ are trivially achievable. By memory sharing, the lower convex envelope of the above four corner points can be achieved, which gives the achievable rate in Theorem 1. ■

Remark 1: Computer-aided investigation using the information theory toolbox [19] showed that the achievable rate provided in Theorem 1 is optimal when $N=2$ and 3. □

Theorem 2: For the cache-aided MuPIR problem with $K=2$ messages, $K_u=2$ users and $N\geq 2$ DBs, the optimal memory-load trade-off under the constraint of uncoded cache placement is given by

$$R_u^\star(M)=\begin{cases}2-\frac{3}{2}M, & 0\leq M\leq \frac{2(N-1)}{2N-1}\\ \frac{(N+1)(2-M)}{2N}, & \frac{2(N-1)}{2N-1}\leq M\leq 2\end{cases} \quad (6)$$

Proof: For the achievability part, the corner points in Theorem 2 are the memory-load pairs $(0,2)$, $(\frac{2(N-1)}{2N-1}, \frac{N+1}{2N-1})$, and $(2,0)$, which can be achieved by the same achievable schemes as Theorem 1.

For the converse part, under the constraint of uncoded cache placement, the converse bound for the original MAN caching problem without privacy constraint in [7] is also the converse for our considered problem. When $M\in[0,1]$, it was proved in [7] that

$$R_{\rm u}^\star(M)\geq 2-\frac{3}{2}M. \quad (7)$$

In addition, the converse bound for the single-user cache-aided PIR problem in [8] is also the converse for our considered problem since increasing the number of users while keeping demand privacy can only possibly increase the load. When $M\in[0,2]$,

$$R_{\rm u}^\star(M)\geq\left(1-\frac{M}{2}\right)\left(1+\frac{1}{N}\right)=\frac{(N+1)(2-M)}{2N}. \quad (8)$$

By combining Eqs. (7) and (8), the converse bound of Eq. (6) can be obtained, which coincides with the achievability. ■

IV. Achievability

In this section we provide the proof of Theorems 1 and 2 by showing the achievability of the non-trivial memory-load pairs $(\frac{N-1}{2N}, \frac{N+1}{N})$ and $(\frac{2(N-1)}{2N-1}, \frac{N+1}{2N-1})$ for the MuPIR problem with $K=2$ messages, $K_u=2$ users and $N\geq 2$ DBs. The design of the achievable schemes utilizes the novel idea of PCIA where certain undesired message symbols (i.e., interferences) are aligned into certain dimensions such that the users can decode their desired messages by solving linear equations after removing the aligned interference.

A. Achievability of $(M,R)=(\frac{N-1}{2N}, \frac{N+1}{N})$

In this section we present the achievability of the memory-load pair $(\frac{N-1}{2N}, \frac{N+1}{N})$. We first provide an example to illustrate the basic design idea and then generalize the proposed scheme to arbitrary number of DBs.

Example 1: (Achievability of $(\frac{1}{4}, \frac{3}{2})$) Consider the MuPIR problem with $K=2$ messages, $K_u=2$ users and $N=2$ DBs. For simplicity, let $W_1=A$ and $W_2=B$ denote the two messages. Let $\boldsymbol{\theta}=(\theta_1,\theta_2)$ be the user demands. We next show that the memory-load pair $(\frac{1}{4}, \frac{3}{2})$ can be achieved.

1) Cache placement: Assume that each message consists of $F=4$ symbols over some finite field $\mathbb{F}_q$, i.e, $A=(A_1,A_2,A_3,A_4), B=(B_1,B_2,B_3,B_4)$ where $A_i,B_i\in\mathbb{F}_q,\forall i\in[4]$ for some prime power q. For this case, binary field is sufficient for the proposed scheme to work. Each user stores a linear combination of the message symbols, i.e.,

$$Z_1 = \{\boldsymbol{\alpha}_1 A_{(1:2)}^{\rm T}+\boldsymbol{\beta}_1 B_{(1:2)}^{\rm T}\}, \quad (9)$$

$$Z_2 = \{\boldsymbol{\alpha}_2 A_{(3:4)}^{\rm T}+\boldsymbol{\beta}_2 B_{(3:4)}^{\rm T}\}, \quad (10)$$

in which $\boldsymbol{\alpha}_i,\boldsymbol{\beta}_i,\forall i=1,2$ are the linear combination coefficients randomly picked from $[\mathbb{F}_2]^{1\times 2}\backslash\{[0,0]\}$. Without loss of generality, we use the coefficients

$$\boldsymbol{\alpha}_1=\boldsymbol{\alpha}_2=\boldsymbol{\beta}_1=\boldsymbol{\beta}_2=[1,1]. \quad (11)$$

2) Private delivery: In this phase, the users download an answer from each DB. These answers take the form of linear combinations of the message symbols. For different user demands $\boldsymbol{\theta}=(\theta_1,\theta_2)$, the linear combination coefficients are designed accordingly such that the users can correctly decode their desired messages while preserving demand privacy. More specifically, the answer from DB 1 consists of two linear combinations $A_{1,1}^{[\boldsymbol{\theta}]}$ and $A_{1,2}^{[\boldsymbol{\theta}]}$ which are

$$A_{1,1}^{[\boldsymbol{\theta}]} = \mathbf{u}_{1,1}A_{(1:2)}^{\rm T}+\mathbf{v}_{1,1}B_{(1:2)}^{\rm T}, \quad (12)$$

$$A_{1,2}^{[\boldsymbol{\theta}]} = \mathbf{u}_{1,2}A_{(3:4)}^{\rm T}+\mathbf{v}_{1,2}B_{(3:4)}^{\rm T}. \quad (13)$$

The answer from DB 2 consists of four linear combinations $A_{2,1}^{[\boldsymbol{\theta}]}, A_{2,2}^{[\boldsymbol{\theta}]}, A_{2,3}^{[\boldsymbol{\theta}]}$ and $A_{2,4}^{[\boldsymbol{\theta}]}$ which are

$$A_{2,1}^{[\boldsymbol{\theta}]} = \mathbf{g}_1 A_{(1:2)}^{\rm T}, \quad (14)$$

$$A_{2,2}^{[\boldsymbol{\theta}]} = \mathbf{g}_2 B_{(1:2)}^{\rm T}, \quad (15)$$

$$A_{2,3}^{[\boldsymbol{\theta}]} = \mathbf{g}_3 A_{(3:4)}^{\rm T}, \quad (16)$$

$$A_{2,4}^{[\boldsymbol{\theta}]} = \mathbf{g}_4 B_{(3:4)}^{\rm T}, \quad (17)$$

in which the coefficients $\mathbf{u}_{1,1},\mathbf{v}_{1,1},\mathbf{u}_{1,2},\mathbf{v}_{1,2},\mathbf{g}_1,\mathbf{g}_2,\mathbf{g}_3,\mathbf{g}_4\in[\mathbb{F}_q]^{1\times 2}\backslash\{[0,0],[1,1]\}$ are subject to design. The answers can be written in a more compact form as

$$\begin{bmatrix}A_{1,1}^{[\boldsymbol{\theta}]}\\ A_{1,2}^{[\boldsymbol{\theta}]}\\ A_{2,1}^{[\boldsymbol{\theta}]}\\ A_{2,2}^{[\boldsymbol{\theta}]}\\ A_{2,3}^{[\boldsymbol{\theta}]}\\ A_{2,4}^{[\boldsymbol{\theta}]}\end{bmatrix}=\begin{bmatrix}\mathbf{u}_{1,1} & \mathbf{0}_2 & \mathbf{v}_{1,1} & \mathbf{0}_2\\ \mathbf{0}_2 & \mathbf{u}_{1,2} & \mathbf{0}_2 & \mathbf{v}_{1,2}\\ \mathbf{g}_1 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2\\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{g}_2 & \mathbf{0}_2\\ \mathbf{0}_2 & \mathbf{g}_3 & \mathbf{0}_2 & \mathbf{0}_2\\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{g}_4\end{bmatrix}\begin{bmatrix}A_{(1:4)}^{\rm T}\\ B_{(1:4)}^{\rm T}\end{bmatrix}. \quad (18)$$

We next show how these coefficients can be designed according to different user demands (θ_1,θ_2) such that decodabilty and privacy can be guaranteed. Due to space limitation, we will only illustrate the case of $(\theta_1,\theta_2)=(1,2)$ and omit other cases, which can be achieved similarly. In this case, users 1 and 2 request A and B respectively. We require that the following six coefficient matrices to be full rank:[2]

$$\begin{bmatrix}\boldsymbol{\alpha}_1\\ \mathbf{g}_1\end{bmatrix},\begin{bmatrix}\boldsymbol{\beta}_2\\ \mathbf{g}_4\end{bmatrix},\begin{bmatrix}\mathbf{u}_{1,2}\\ \mathbf{g}_3\end{bmatrix},\begin{bmatrix}\mathbf{v}_{1,1}\\ \mathbf{g}_2\end{bmatrix},\begin{bmatrix}\boldsymbol{\beta}_1\\ \mathbf{g}_2\end{bmatrix},\begin{bmatrix}\boldsymbol{\alpha}_2\\ \mathbf{g}_3\end{bmatrix}. \tag{19}$$

and $\mathbf{u}_{1,1}$ and $\mathbf{v}_{1,2}$ are chosen as

$$\text{[alignment]} \quad \mathbf{u}_{1,1}=\mathbf{g}_1,\ \mathbf{v}_{1,2}=\mathbf{g}_4. \tag{20}$$

Now we show that with the above full rank and alignment conditions the users can correctly decode their desired messages.

From Eq. (18), we obtain

$$\begin{bmatrix}A_{1,2}^{[(1,2)]}\\ A_{2,3}^{[(1,2)]}\end{bmatrix}=\begin{bmatrix}\mathbf{u}_{1,2} & \mathbf{v}_{1,2}\\ \mathbf{g}_3 & \mathbf{0}_2\end{bmatrix}\begin{bmatrix}A_{(3:4)}^{\mathrm{T}}\\ B_{(3:4)}^{\mathrm{T}}\end{bmatrix}. \tag{21}$$

Since $\mathbf{g}_4=\mathbf{v}_{1,2}$, we have $A_{2,4}^{[(1,2)]}=\mathbf{g}_4B_{(3:4)}^{\mathrm{T}}=\mathbf{v}_{1,2}B_{(3:4)}^{\mathrm{T}}$, i.e, the interferences B_3 and B_4 are aligned. Removing the interferences and due to $\text{rank}([\mathbf{u}_{1,2};\mathbf{g}_3])=2$, the symbols $A_{(3:4)}$ can be solved from Eq. (21) as

$$\begin{bmatrix}A_3\\ A_4\end{bmatrix}=\begin{bmatrix}\mathbf{u}_{1,2}\\ \mathbf{g}_3\end{bmatrix}^{-1}\begin{bmatrix}A_{1,2}^{[(1,2)]}-A_{2,4}^{[(1,2)]}\\ A_{2,3}^{[(1,2)]}\end{bmatrix}. \tag{22}$$

From Eq. (18), we also obtain

$$\begin{bmatrix}A_{1,1}^{[(1,2)]}\\ A_{2,2}^{[(1,2)]}\end{bmatrix}=\begin{bmatrix}\mathbf{u}_{1,1} & \mathbf{v}_{1,1}\\ \mathbf{0}_2 & \mathbf{g}_2\end{bmatrix}\begin{bmatrix}A_{(1:2)}^{\mathrm{T}}\\ B_{(1:2)}^{\mathrm{T}}\end{bmatrix}. \tag{23}$$

Since $\mathbf{g}_1=\mathbf{u}_{1,1}$, we have $A_{2,1}^{[(1,2)]}=\mathbf{g}_1A_{(1:2)}^{\mathrm{T}}=\mathbf{u}_{1,1}A_{(1:2)}^{\mathrm{T}}$, i.e, the interferences A_1,A_2 are aligned. Removing the interferences, $B_{(1:2)}$ can be solved from Eq. (23) as

$$\begin{bmatrix}B_1\\ B_2\end{bmatrix}=\begin{bmatrix}\mathbf{v}_{1,1}\\ \mathbf{g}_2\end{bmatrix}^{-1}\begin{bmatrix}A_{1,1}^{[(1,2)]}-A_{2,1}^{[(1,2)]}\\ A_{2,2}^{[(1,2)]}\end{bmatrix}. \tag{24}$$

Hence, both users can correctly decode A_3,A_4 and B_1,B_2.

Now, user 1 still needs A_1,A_2 while user 2 needs B_3,B_4. For user 1, the interference term $\boldsymbol{\beta}_1B_{(1:2)}^{\mathrm{T}}$ in Z_1 can be eliminated since B_1,B_2 are already available. Therefore, user 1 can decode A_1,A_2 from

$$\begin{bmatrix}A_1\\ A_2\end{bmatrix}=\begin{bmatrix}\boldsymbol{\alpha}_1\\ \mathbf{g}_1\end{bmatrix}^{-1}\begin{bmatrix}Z_1-\boldsymbol{\beta}_1B_{(1:2)}^{\mathrm{T}}\\ A_{2,1}^{[(1,2)]}\end{bmatrix}. \tag{25}$$

Similarly, for user 2, the interference term $\boldsymbol{\alpha}_2A_{(3:4)}^{\mathrm{T}}$ can be eliminated since A_3,A_4 are already available. Therefore, user 2 can decode B_3,B_4 from

$$\begin{bmatrix}B_3\\ B_4\end{bmatrix}=\begin{bmatrix}\boldsymbol{\beta}_2\\ \mathbf{g}_4\end{bmatrix}^{-1}\begin{bmatrix}Z_2-\boldsymbol{\alpha}_2A_{(3:4)}^{\mathrm{T}}\\ A_{2,4}^{[(1,2)]}\end{bmatrix}. \tag{26}$$

As a result, both users can correctly recover their desired messages.

We next show that how the specific coefficients are chosen such that user demand privacy can be preserved.

Privacy: Due to the full rank requirements and alignment conditions, it must hold that $\mathbf{u}_{1,i},\mathbf{v}_{1,i}\neq[1,1],[0,0],\forall i\in[2]$ and $\mathbf{g}_j\neq[1,1],[0,0],\forall j\in[4]$. Therefore, there are $2^4=16$ choices for $(\mathbf{u}_{1,1},\mathbf{v}_{1,1},\mathbf{u}_{1,2},\mathbf{v}_{1,2})\in\{[0,1],[1,0]\}^4$. For each choice, we can uniquely determine $(\mathbf{g}_1,\mathbf{g}_2,\mathbf{g}_3,\mathbf{g}_4)\in\{[0,1],[1,0]\}^4$ for each demand $(\theta_1,\theta_2)\in[2]^2$ according to the corresponding full rank and alignment conditions. If we employ each choice of the linear coefficients of the DBs with equal probability, then for any choice of the DB 1 linear combination coefficients, (θ_1,θ_2) is equally likely to be $(1,2),(2,1),(1,1)$ or $(2,2)$ from the perspective of DB 1. On the other hand, there are also $2^4=16$ choices for $[\mathbf{g}_1,\mathbf{g}_2,\mathbf{g}_3,\mathbf{g}_4]\in\{[0,1],[1,0]\}^4$. For each choice, we can uniquely determine $[\mathbf{u}_{1,1},\mathbf{v}_{1,1},\mathbf{u}_{1,2},\mathbf{v}_{1,2}]\in\{[0,1],[1,0]\}^4$ for each demand $(\theta_1,\theta_2)\in[2]^2$ according to the full rank and alignment conditions therein. If we employ each choice of the linear coefficients of the DBs with equal probability, then for any choice of the DB 2 linear combination coefficients, (θ_1,θ_2) is equally likely to be $(1,2),(2,1),(1,1)$ or $(2,2)$ from the perspective of DB 2. Therefore, the proposed scheme is private.

Performance: Since $D=6$ linear combinations are downloaded in total, the achieved load is $R=\frac{D}{F}=\frac{3}{2}$. ◊

We next present the general achievable schemes for arbitrary number of databases. The cache placement and the corresponding private delivery phases are formally described as follows.

1) Cache placement: Let $W_1=A,W_2=B$ denote the two messages. Each message is assumed to consist of $F=2N$ symbols over some finite field $\mathbb{F}_q$, i.e.,

$$\begin{aligned} A &= (A_1,\cdots,A_{2N}), && (27)\\ B &= (B_1,\cdots,B_{2N}). && (28)\end{aligned}$$

Each user then stores $N-1$ linear combinations of the message symbols in its cache (therefore $M=\frac{N-1}{2N}$), i.e.,

$$Z_1=\left\{\boldsymbol{\alpha}_{1,j}A_{(1:N)}^{\mathrm{T}}+\boldsymbol{\beta}_{1,j}B_{(1:N)}^{\mathrm{T}}:j\in[N-1]\right\}, \tag{29}$$

$$Z_2=\left\{\boldsymbol{\alpha}_{2,j}A_{(N+1:2N)}^{\mathrm{T}}+\boldsymbol{\beta}_{2,j}B_{(N+1:2N)}^{\mathrm{T}}:j\in[N-1]\right\}. \tag{30}$$

in which the linear combination coefficients $\boldsymbol{\alpha}_{i,j},\boldsymbol{\beta}_{i,j},\forall i\in[2],\forall j\in[N-1]$ are randomly picked from $[\mathbb{F}_q]^{1\times N}\backslash\{\mathbf{0}_N\}$ such that $\text{rank}([\boldsymbol{\alpha}_{i,1};\boldsymbol{\alpha}_{i,2};\cdots;\boldsymbol{\alpha}_{i,N-1}])=N-1$ and $\text{rank}([\boldsymbol{\beta}_{i,1};\boldsymbol{\beta}_{i,2};\cdots;\boldsymbol{\beta}_{i,N-1}])=N-1,\forall i\in[2]$. Let $Z_{i,1},Z_{i,2},\cdots,Z_{i,N-1}$ denote the $N-1$ stored linear combinations in $Z_i,\forall i\in[2]$, i.e., $\forall j\in[N-1]$:

$$\begin{aligned} Z_{1,j} &= \boldsymbol{\alpha}_{1,j}A_{(1:N)}^{\mathrm{T}}+\boldsymbol{\beta}_{1,j}B_{(1:N)}^{\mathrm{T}}, && (31)\\ Z_{2,j} &= \boldsymbol{\alpha}_{2,j}A_{(N+1:2N)}^{\mathrm{T}}+\boldsymbol{\beta}_{2,j}B_{(N+1:2N)}^{\mathrm{T}}. && (32)\end{aligned}$$

2) Private delivery: The answer of each DB consists of several linear combinations of the message symbols. More specifically, the answer of DB $n\in[N-1]$ consists of two

[2] The way we choose the coefficients to satisfy these requirements will be described later. Here we only present the conditions which are necessary for correct decoding and preserving demand privacy.

linear combinations

$$A_{n,1}^{[\boldsymbol{\theta}]} = \mathbf{u}_{n,1}A_{(1:N)}^{\mathrm{T}} + \mathbf{v}_{n,1}B_{(1:N)}^{\mathrm{T}}, \tag{33}$$

$$A_{n,2}^{[\boldsymbol{\theta}]} = \mathbf{u}_{n,2}A_{(N+1:2N)}^{\mathrm{T}} + \mathbf{v}_{n,2}B_{(N+1:2N)}^{\mathrm{T}}, \tag{34}$$

and the answer of DB N consists of four linear combinations

$$A_{N,1}^{[\boldsymbol{\theta}]} = \mathbf{g}_1 A_{(1:N)}^{\mathrm{T}}, \tag{35}$$

$$A_{N,2}^{[\boldsymbol{\theta}]} = \mathbf{g}_2 B_{(1:N)}^{\mathrm{T}}, \tag{36}$$

$$A_{N,3}^{[\boldsymbol{\theta}]} = \mathbf{g}_3 A_{(N+1:2N)}^{\mathrm{T}}, \tag{37}$$

$$A_{N,4}^{[\boldsymbol{\theta}]} = \mathbf{g}_4 B_{(N+1:2N)}^{\mathrm{T}}. \tag{38}$$

The linear coefficients $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3, \mathbf{g}_4, \mathbf{u}_{n,j}, \mathbf{v}_{n,j}, \forall n \in [N-1], \forall j \in [2]$ belong to $[\mathbb{F}_q]^{1\times N}\backslash\{\mathbf{0}_N\}$ and are designed according to the user demands.

In the private delivery phase, the users download all the $2N+2$ linear combinations from the N DBs. The linear combination coefficients of the answers are chosen according to $\boldsymbol{\theta}$ such that both users can correctly decode their desired messages. In the following, we will focus on the case of $\boldsymbol{\theta} = (1,2)$ and omit the other demands, which can be designed similarly. In the case of $(\theta_1, \theta_2) = (1,2)$, user 1 and 2 demand message A and B respectively. We let

$$\mathbf{g}_1 = \mathbf{u}_{1,1} = \mathbf{u}_{2,1} = \cdots = \mathbf{u}_{N-1,1}, \tag{39}$$

$$\mathbf{g}_4 = \mathbf{v}_{1,2} = \mathbf{v}_{2,2} = \cdots = \mathbf{v}_{N-1,2}, \tag{40}$$

and the linear coefficient vectors

$$\mathbf{g}_1, \ldots, \mathbf{g}_4, \mathbf{u}_{1,2}, \ldots, \mathbf{u}_{N-1,2}, \mathbf{v}_{1,1}, \ldots, \mathbf{v}_{N-1,1}$$

are picked independently and uniformly at random from $[\mathbb{F}_q]^{1\times N}\backslash\{\mathbf{0}_N\}$ such that with high probability,[3] the following six matrices are full rank

$$\begin{bmatrix} \boldsymbol{\alpha}_{1,1} \\ \vdots \\ \boldsymbol{\alpha}_{1,N-1} \\ \mathbf{g}_1 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\beta}_{2,1} \\ \vdots \\ \boldsymbol{\beta}_{2,N-1} \\ \mathbf{g}_4 \end{bmatrix}, \begin{bmatrix} \mathbf{u}_{1,2} \\ \vdots \\ \mathbf{u}_{N-1,2} \\ \mathbf{g}_3 \end{bmatrix}, \begin{bmatrix} \mathbf{v}_{1,1} \\ \vdots \\ \mathbf{v}_{N-1,1} \\ \mathbf{g}_2 \end{bmatrix},$$

$$\begin{bmatrix} \boldsymbol{\alpha}_{2,1} \\ \vdots \\ \boldsymbol{\alpha}_{2,N-1} \\ \mathbf{g}_3 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\beta}_{1,1} \\ \vdots \\ \boldsymbol{\beta}_{1,N-1} \\ \mathbf{g}_2 \end{bmatrix}. \tag{41}$$

With such a design of the linear coefficients, we now demonstrate that the users can correctly recover their desired messages.

Due to the alignment condition of Eq. (40), we have $A_{2,4}^{[(1,2)]} = \mathbf{g}_4 B_{(N+1:2N)}^{\mathrm{T}} = \mathbf{v}_{1,2}B_{(N+1:2N)}^{\mathrm{T}} = \cdots = \mathbf{v}_{N-1,2}B_{(N+1:2N)}^{\mathrm{T}}$, i.e., the interfering symbols $B_{(N+1:2N)}$ are aligned among the linear combinations $A_{1,2}^{[(1,2)]}, A_{2,2}^{[(1,2)]}, \cdots, A_{N-1,2}^{[(1,2)]}$. Subtracting $\mathbf{g}_4 B_{(N+1:2N)}^{\mathrm{T}}$ from $A_{1,2}^{[(1,2)]}, A_{2,2}^{[(1,2)]}, \cdots, A_{N-1,2}^{[(1,2)]}$, we obtain

$$\begin{bmatrix} A_{1,2}^{[(1,2)]} - A_{2,4}^{[(1,2)]} \\ \vdots \\ A_{N-1,2}^{[(1,2)]} - A_{2,4}^{[(1,2)]} \\ A_{N,3}^{[(1,2)]} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_{1,2} \\ \vdots \\ \mathbf{u}_{N-1,2} \\ \mathbf{g}_3 \end{bmatrix} A_{(N+1:2N)}^{\mathrm{T}}. \tag{42}$$

Since the linear coefficient matrix on the RHS of Eq. (42) is full rank, $A_{(N+1:2N)}$ can be solved from

$$A_{(N+1:2N)}^{\mathrm{T}} = \begin{bmatrix} \mathbf{u}_{1,2} \\ \vdots \\ \mathbf{u}_{N-1,2} \\ \mathbf{g}_3 \end{bmatrix}^{-1} \begin{bmatrix} A_{1,2}^{[(1,2)]} - A_{2,4}^{[(1,2)]} \\ \vdots \\ A_{N-1,2}^{[(1,2)]} - A_{2,4}^{[(1,2)]} \\ A_{N,3}^{[(1,2)]} \end{bmatrix}. \tag{43}$$

Therefore, both users can correctly decode the symbols $A_{(N+1:2N)}$. Similarly, due to the alignment condition of Eq. (39), we have $A_{2,1}^{[(1,2)]} = \mathbf{g}_1 A_{(1:N)}^{\mathrm{T}} = \mathbf{u}_{1,1}A_{(1:N)}^{\mathrm{T}} = \mathbf{u}_{2,1}A_{(1:N)}^{\mathrm{T}} = \cdots = \mathbf{u}_{N-1,1}A_{(1:N)}^{\mathrm{T}}$, i.e., the interfering symbols $A_{(1:N)}$ are aligned among the linear combinations $A_{1,1}^{[(1,2)]}, A_{2,1}^{[(1,2)]}, \cdots, A_{N-1,1}^{[(1,2)]}$. Subtracting $\mathbf{g}_1 A_{(1:N)}^{\mathrm{T}}$ from $A_{1,1}^{[(1,2)]}, A_{2,1}^{[(1,2)]}, \cdots, A_{N-1,1}^{[(1,2)]}$, we obtain

$$\begin{bmatrix} A_{1,1}^{[(1,2)]} - A_{2,1}^{[(1,2)]} \\ \vdots \\ A_{N-1,1}^{[(1,2)]} - A_{2,1}^{[(1,2)]} \\ A_{N,2}^{[(1,2)]} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{1,1} \\ \vdots \\ \mathbf{v}_{N-1,1} \\ \mathbf{g}_2 \end{bmatrix} B_{(1:N)}^{\mathrm{T}}. \tag{44}$$

Since the linear coefficient matrix on the RHS of Eq. (44) is full rank, $B_{(1:N)}$ can be solved from

$$B_{(1:N)}^{\mathrm{T}} = \begin{bmatrix} \mathbf{v}_{1,1} \\ \vdots \\ \mathbf{v}_{N-1,1} \\ \mathbf{g}_2 \end{bmatrix}^{-1} \begin{bmatrix} A_{1,1}^{[(1,2)]} - A_{2,1}^{[(1,2)]} \\ \vdots \\ A_{N-1,1}^{[(1,2)]} - A_{2,1}^{[(1,2)]} \\ A_{N,2}^{[(1,2)]} \end{bmatrix}. \tag{45}$$

Therefore, both users can correctly decode the packets $B_{(1:N)}$.

Now the message symbols $A_{(N+1:2N)}$ and $B_{(1:N)}$ are available to both users. User 1 still needs $A_{(1:N)}$ and user 2 still needs $B_{(N+1:2N)}$. For user 1, since $B_{(1:N)}$ are already available, it obtains $N-1$ linear combinations of $A_{(1:N)}$ according to Eq. (29). Together with $A_{N,1}^{[(1,2)]} = \mathbf{g}_1 A_{(1:N)}^{\mathrm{T}}$, user 1 obtains N independent linear combinations of $A_{(1:N)}$, from which $A_{(1:N)}$ can be solved as

$$A_{(1:N)}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{\alpha}_{1,1} \\ \vdots \\ \boldsymbol{\alpha}_{1,N-1} \\ \mathbf{g}_1 \end{bmatrix}^{-1} \begin{bmatrix} Z_{1,1} - \boldsymbol{\beta}_{1,1}B_{(1:N)}^{\mathrm{T}} \\ \vdots \\ Z_{1,N-1} - \boldsymbol{\beta}_{1,N-1}B_{(1:N)}^{\mathrm{T}} \\ A_{N,1}^{[(1,2)]} \end{bmatrix}. \tag{46}$$

Therefore, user 1 can correctly decode all the symbols $A_{(1:2N)}$ of the desired message A. Similarly, for user 2, since $A_{(N+1:2N)}$ are already available, it obtains $N-1$ linear

[3] The base field q is assumed to be large enough such that the probability of the random matrices being full rank is arbitrarily close to 1.

combinations of $B_{(N+1:2N)}$ according to Eq. (30). Together with $A_{N,4}^{[(1,2)]}=\mathbf{g}_4B_{(N+1:2N)}^{\mathrm{T}}$, user 2 obtains N independent linear combinations of $B_{(N+1:2N)}$, from which $B_{(N+1:2N)}$ can be solved as

$$B_{(N+1:2N)}^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{\beta}_{2,1}\\ \vdots\\ \boldsymbol{\beta}_{2,N-1}\\ \mathbf{g}_4\end{bmatrix}^{-1}\begin{bmatrix}Z_{2,1}-\boldsymbol{\alpha}_{2,1}A_{(N+1:2N)}^{\mathrm{T}}\\ \vdots\\ Z_{2,N-1}-\boldsymbol{\alpha}_{2,N-1}A_{(N+1:2N)}^{\mathrm{T}}\\ A_{N,4}^{[(1,2)]}\end{bmatrix}. \tag{47}$$

Therefore, user 2 can correctly decode all the symbols $B_{(1:2N)}$ of the desired message B. As a result, both users can correctly decode their desired messages.

Privacy: For any $\boldsymbol{\theta}$, for each DB $n\in[N-1]$, its answer contains one linear combinations of $A_{(1:N)},B_{(1:N)}$ and one linear combination of $A_{(N+1:2N)},B_{(N+1:2N)}$, which are generated randomly and independently from its viewpoint. For DB N, its answer contains four linear combinations of $A_{(1:N)},A_{(N+1:2N)},B_{(1:N)}$, and $B_{(N+1:2N)}$ respectively, which are generated randomly and independently from its viewpoint. In addition, all of these linear combinations are generated independently to the cache content of each user. Since each individual DB does not know the exact full rank and alignment conditions, the user demands (θ_1,θ_2) can be arbitrary from that DB's point of view. As a result, the proposed scheme is private.

Performance: Since $D=2N+2$ linear combinations are downloaded in total, the achieved load is $R=\frac{D}{F}=\frac{N+1}{N}$.

B. Achievability of $(M,R)=(\frac{2(N-1)}{2N-1},\frac{N+1}{2N-1})$

In this section we present the achievability of the memory-load pair $(\frac{2(N-1)}{2N-1},\frac{N+1}{2N-1})$, which is also based on the PCIA strategy. We provide an example first and then generalize the scheme to arbitrary number of databases.

Example 2: (Achievability of $(\frac{2}{3},1)$) Consider the same setting as Example 1 and each user has a cache memory of $M=\frac{2}{3}$ messages. We assume that each message consists of $L=3$ bits, i.e., $A=(A_1,A_2,A_3)$ and $B=(B_1,B_2,B_3)$.

1) Cache placement: The cache placement is

$$Z_1=\{A_1,B_1\}, \tag{48}$$
$$Z_2=\{A_2,B_2\}. \tag{49}$$

2) Private delivery: The answer of each DB 1 consists of a linear combination $A_1^{[\boldsymbol{\theta}]}=\mathbf{u}_1A_{(1:3)}^{\mathrm{T}}+\mathbf{v}_1B_{(1:3)}^{\mathrm{T}}$ and the answer of DB 2 consists of two linear combinations $A_{2,1}^{[\boldsymbol{\theta}]}=\mathbf{g}_1A_{(1:3)}^{\mathrm{T}}$ and $A_{2,2}^{[\boldsymbol{\theta}]}=\mathbf{g}_2B_{(1:3)}^{\mathrm{T}}$ where the linear coefficients $\mathbf{u}_1=[u_{1,1},u_{1,2},u_{1,3}]$, $\mathbf{v}_1=[v_{1,1},v_{1,2},v_{1,3}]$, $\mathbf{g}_1=[g_{1,1},g_{1,2},g_{1,3}]$ and $\mathbf{g}_2=[g_{2,1},g_{2,2},g_{2,3}]$ belong to $[\mathbb{F}_2]^{1\times3}\backslash\{[0,0,0]\}$ and are subject to design according to different user demands (θ_1,θ_2). The answers from the DBs can be written in a more compact form as

$$\begin{bmatrix}A_1^{[\boldsymbol{\theta}]}\\ A_{2,1}^{[\boldsymbol{\theta}]}\\ A_{2,2}^{[\boldsymbol{\theta}]}\end{bmatrix}=\begin{bmatrix}\mathbf{u}_1&\mathbf{v}_1\\ \mathbf{g}_1&\mathbf{0}_3\\ \mathbf{0}_3&\mathbf{g}_2\end{bmatrix}\begin{bmatrix}A_{(1:3)}^{\mathrm{T}}\\ B_{(1:3)}^{\mathrm{T}}\end{bmatrix}. \tag{50}$$

In the private delivery phase, the users download all the three linear combinations from the two DBs. In the following, we will only illustrate the case of $(\theta_1,\theta_2)=(1,2)$. In this case, we require that the following two coefficient matrices to be full rank:[4]

$$\begin{bmatrix}u_{1,2}&u_{1,3}\\ g_{1,2}&g_{1,3}\end{bmatrix},\begin{bmatrix}v_{1,1}&v_{1,3}\\ g_{2,1}&g_{2,3}\end{bmatrix}, \tag{51}$$

and

$$\begin{aligned}[g_{1,1},g_{1,3}]&=[u_{1,1},u_{1,3}],\\ [g_{2,2},g_{2,3}]&=[v_{1,2},v_{1,3}].\end{aligned} \tag{52}$$

i.e., the symbols A_1,A_3 are aligned among the linear combinations $A_1^{[(1,2)]}$ and $A_{2,1}^{[(1,2)]}$, the symbols B_2,B_3 are aligned among $A_1^{[(1,2)]}$ and $A_{2,2}^{[(1,2)]}$. We next show that with the above full rank and alignment conditions, both users can correctly decode their desired messages.

For user 1, since $[g_{2,2},g_{2,3}]=[v_{1,2},v_{1,3}]$, we have $[v_{1,2},v_{1,3}][B_2,B_3]^{\mathrm{T}}=[g_{2,2},g_{2,3}][B_2,B_3]^{\mathrm{T}}=A_{2,2}^{[(1,2)]}-g_{2,1}B_1$. Then from Eq. (50), we obtain that

$$\begin{bmatrix}A_2\\ A_3\end{bmatrix}=\begin{bmatrix}u_{1,2}&u_{1,3}\\ g_{1,2}&g_{1,3}\end{bmatrix}^{-1}\mathbf{y} \tag{53}$$

where

$$\mathbf{y}=\begin{bmatrix}A_1^{[(1,2)]}\\ A_{2,1}^{[(1,2)]}\end{bmatrix}-\begin{bmatrix}A_{2,2}^{[(1,2)]}-g_{2,1}B_1\\ 0\end{bmatrix}-\begin{bmatrix}u_{1,1}\\ g_{1,1}\end{bmatrix}A_1-\begin{bmatrix}v_{1,1}\\ 0\end{bmatrix}B_1. \tag{54}$$

Since A_1,B_1 are already stored by user 1, it can correctly decode the desired symbols A_2 and A_3. For user 2, since $[g_{1,1},g_{1,3}]=[u_{1,1},u_{1,3}]$, we have $[u_{1,1},u_{1,3}][A_1,A_3]^{\mathrm{T}}=[g_{1,1},g_{1,3}][A_1,A_3]^{\mathrm{T}}=A_{2,1}^{[(1,2)]}-g_{1,2}A_2$. Then from Eq. (50), we obtain that

$$\begin{bmatrix}B_1\\ B_3\end{bmatrix}=\begin{bmatrix}v_{1,1}&v_{1,3}\\ g_{2,1}&g_{2,3}\end{bmatrix}^{-1}\mathbf{y}' \tag{55}$$

where

$$\mathbf{y}'=\begin{bmatrix}A_1^{[(1,2)]}\\ A_{2,2}^{[(1,2)]}\end{bmatrix}-\begin{bmatrix}A_{2,1}^{[(1,2)]}-g_{2,1}A_2\\ 0\end{bmatrix}-\begin{bmatrix}u_{1,2}\\ 0\end{bmatrix}A_2-\begin{bmatrix}u_{1,2}\\ g_{1,2}\end{bmatrix}B_2. \tag{56}$$

Since A_2,B_2 are already stored by user 2, it can correctly decode the desired symbols B_1 and B_3. Together with the cached symbols, both users can recover their desired messages.

Next we show how to choose the specific linear coefficients such that the user demand privacy can be preserved.

Privacy: From the full rank and alignment requirements, to preserve demand privacy, it must hold that $u_{1,3}=v_{1,3}=1$, $g_{1,3}=g_{2,3}=1$ for any demand $(\theta_1,\theta_2)\in[2]^2$. Therefore, there are $2^4=16$ choices for $[u_{1,1},u_{1,2},v_{1,1},v_{1,2}]\in\{0,1\}^4$. For each choice, we can uniquely determine the DB 2 linear combination coefficients $\mathbf{g}_1,\mathbf{g}_2$ for each $(\theta_1,\theta_2)\in[2]^2$. If we employ each choice of the linear coefficients of the DBs

[4] The way we choose the coefficients to satisfy these requirements will be described later. Here we only present the conditions which are necessary for correct decoding and preserving demand privacy.

with equal probability, then for any choice of the DB 1 linear combination coefficients, (θ_1, θ_2) is equally likely to be $(1,2), (2,1), (1,1)$ or $(2,2)$ from DB 1's perspective. On the other hand, there are also $2^4 = 16$ choices of the DB 2 linear coefficients $[g_{1,1}, g_{1,2}, g_{2,1}, g_{2,2}] \in \{0,1\}^4$ (note that $g_{1,3} = g_{2,3} = 1$). For each choice, we can uniquely determine the DB 1 linear coefficients $\mathbf{u}_1, \mathbf{v}_1$ for each $(\theta_1, \theta_2) \in [2]^2$. If we employ each choice of the linear coefficients of the DBs with equal probability, then for any choice of the DB 2 linear combination coefficients, (θ_1, θ_2) is equally likely to be $(1,2), (2,1), (1,1)$ or $(2,2)$ from DB 2's perspective. As a result, the proposed scheme is private.

Performance: Since $D = 3$ linear combinations are downloaded in total, the achieved load is $R = \frac{D}{F} = 1$. $\Diamond$

Next we present the general achievable schemes for arbitrary number of databases. The cache placement and private delivery phases are formally described as follows.

1) Cache placement: Each message is assumed to consist of $F = 2N - 1$ symbols over some finite field, i.e., $A = (A_1, A_2, \cdots, A_{2N-1})$, $B = (B_1, B_2, \cdots, B_{2N-1})$ for which $A_i, B_i \in \mathbb{F}_q, \forall i \in [2N-1]$ where q is assumed to be large enough. The cache placement is

$$\begin{aligned} Z_1 &= \{A_{1:N-1}, B_{1:N-1}\}, && (57) \\ Z_2 &= \{A_{N:2N-2}, B_{N:2N-2}\}. && (58) \end{aligned}$$

2) Private delivery: We first construct the answers from the DBs. For DB $n, n \in [N-1]$, the answer is

$$A_n^{[\boldsymbol{\theta}]} = \mathbf{u}_n A^{\mathrm{T}}_{(1:2N-1)} + \mathbf{v}_n B^{\mathrm{T}}_{(1:2N-1)}. \tag{59}$$

For DB N, the answer consists of two linear combinations

$$\begin{aligned} A_{N,1}^{[\boldsymbol{\theta}]} &= \mathbf{g}_1 A^{\mathrm{T}}_{(1:2N-1)}, && (60) \\ A_{N,2}^{[\boldsymbol{\theta}]} &= \mathbf{g}_2 B^{\mathrm{T}}_{(1:2N-1)}, && (61) \end{aligned}$$

in which the linear combination coefficients $\mathbf{u}_n = [u_{n,1}, u_{n,2}, \cdots, u_{n,2N-1}], \forall n \in [N-1]$, $\mathbf{v}_n = [v_{n,1}, v_{n,2}, \cdots, v_{n,2N-1}], \forall n \in [N-1]$, $\mathbf{g}_1 = [g_{1,1}, g_{1,2}, \cdots, g_{1,2N-1}]$ and $\mathbf{g}_2 = [g_{2,1}, g_{2,2}, \cdots, g_{2,2N-1}]$ belong to $[\mathbb{F}_q]^{1 \times (2N-1)} \backslash \{\mathbf{0}_{2N-1}\}$ and are subject to design according to different user demands (θ_1, θ_2).

The answers can be written in a more compact form as

$$\begin{bmatrix} A_1^{[\boldsymbol{\theta}]} \\ \vdots \\ A_{N-1}^{[\boldsymbol{\theta}]} \\ A_{N,1}^{[\boldsymbol{\theta}]} \\ A_{N,2}^{[\boldsymbol{\theta}]} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{v}_1 \\ \vdots & \vdots \\ \mathbf{u}_{N-1} & \mathbf{v}_{N-1} \\ \mathbf{g}_1 & \mathbf{0}_{2N-1} \\ \mathbf{0}_{2N-1} & \mathbf{g}_2 \end{bmatrix} \begin{bmatrix} A^{\mathrm{T}}_{(1:2N-1)} \\ B^{\mathrm{T}}_{(1:2N-1)} \end{bmatrix} \tag{62}$$

We will only consider the case of $\boldsymbol{\theta} = (1,2)$ in the following. We let $\forall n \in [N-1]$:

$$\begin{aligned} [\mathbf{g}_{1,(1:N-1)}, g_{1,2N-1}] &= [\mathbf{u}_{n,(1:N-1)}, u_{n,2N-1}], && (63) \\ \mathbf{g}_{2,(N:2N-1)} &= \mathbf{v}_{n,(N:2N-1)}, && (64) \end{aligned}$$

i.e., the message symbols $A_{(1:N-1)}, A_{2N-1}$ are aligned among $A_1^{[(1,2)]}, A_2^{[(1,2)]}, \cdots, A_{N,1}^{[(1,2)]}$ and the symbols $B_{(N:2N-1)}$ are aligned among $A_1^{[(1,2)]}, A_2^{[(1,2)]}, \cdots, A_{N,2}^{[(1,2)]}$. The linear combination coefficients $\mathbf{g}_1$, $\mathbf{g}_2$, $\mathbf{u}_{n,(N:2N-2)}$, $\mathbf{v}_{n,(1:N-1)}, \forall n \in [N-1]$ are picked randomly (with uniform probability) and independently from $[\mathbb{F}_q]^{1 \times (2N-1)} \backslash \{\mathbf{0}_{2N-1}\}$ such that with high probability the following two coefficient matrices are full rank:

$$\begin{bmatrix} \mathbf{u}_{1,(N:2N-1)} \\ \vdots \\ \mathbf{u}_{N-1,(N:2N-1)} \\ \mathbf{g}_{1,(N:2N-1)} \end{bmatrix}, \begin{bmatrix} [\mathbf{v}_{1,(1:N-1)}, v_{1,2N-1}] \\ \vdots \\ [\mathbf{v}_{N-1,(1:N-1)}, v_{N-1,2N-1}] \\ [\mathbf{g}_{2,(1:N-1)}, g_{2,2N-1}] \end{bmatrix} \tag{65}$$

We next show that with such a design of linear coefficients, both users can correctly decode their desired messages.

For user 1, due to the alignment of $B_{(N:2N-1)}$, we have $A_{N,2}^{[(1,2)]} - \mathbf{g}_{2,(1:N-1)} B^{\mathrm{T}}_{(1:N-1)} = \mathbf{g}_{2,(N:2N-1)} B^{\mathrm{T}}_{(N:2N-1)} = \mathbf{v}_{n,(N:2N-1)} B^{\mathrm{T}}_{(N:2N-1)}, \forall n \in [N-1]$. Subtracting $A_{N,2}^{[(1,2)]} - \mathbf{g}_{2,(1:N-1)} B^{\mathrm{T}}_{(1:N-1)}$ from $A_1^{[(1,2)]}, A_2^{[(1,2)]}, \cdots, A_{N-1}^{[(1,2)]}$ in Eq. (62), together with $A_{N,1}^{[(1,2)]} = \mathbf{g}_1 A^{\mathrm{T}}_{(1:2N-1)}$, we obtain N independent linear combinations of $A_{(N:2N-1)}$, from which $A_{(N:2N-1)}$ can be solved as

$$A^{\mathrm{T}}_{(N:2N-1)} = \begin{bmatrix} \mathbf{u}_{1,(N:2N-1)} \\ \vdots \\ \mathbf{u}_{N-1,(N:2N-1)} \\ \mathbf{g}_{1,(N:2N-1)} \end{bmatrix}^{-1} \mathbf{y} \tag{66}$$

where

$$\begin{aligned} \mathbf{y} = & \begin{bmatrix} A_1^{[(1,2)]} - A_{N,2}^{[(1,2)]} + \mathbf{g}_{2,(1:N-1)} B^{\mathrm{T}}_{(1:N-1)} \\ \vdots \\ A_{N-1}^{[(1,2)]} - A_{N,2}^{[(1,2)]} + \mathbf{g}_{2,(1:N-1)} B^{\mathrm{T}}_{(1:N-1)} \\ A_{N,1}^{[(1,2)]} \end{bmatrix} \\ & - \begin{bmatrix} \mathbf{u}_{1,(1:N-1)} & \mathbf{v}_{1,(1:N-1)} \\ \vdots & \\ \mathbf{u}_{N-1,(1:N-1)} & \mathbf{v}_{N-1,(1:N-1)} \\ \mathbf{g}_{1,(1:N-1)} & \mathbf{0}_{N-1} \end{bmatrix} \begin{bmatrix} A^{\mathrm{T}}_{(1:N-1)} \\ B^{\mathrm{T}}_{(1:N-1)} \end{bmatrix}. \end{aligned} \tag{67}$$

Since the symbols $A_{(1:N-1)}, B_{(1:N-1)}$ are already cached by user 1, it can correctly decode the symbols $A_{(N:2N-1)}$ and recover message A.

For user 2, due to the alignment of $A_{(1:N-1)}, A_{2N-1}$, we have $A_{N,1}^{[(1,2)]} - \mathbf{g}_{1,(N:2N-2)} A^{\mathrm{T}}_{(N:2N-2)} = [\mathbf{g}_{1,(1:N-1)}, g_{1,2N-1}][A_{(1:N-1)}, A_{2N-1}]^{\mathrm{T}} = \mathbf{u}_{n,(1:N-1)}, u_{n,2N-1}][A_{(1:N-1)}, A_{2N-1}]^{\mathrm{T}}, \forall n \in [N-1]$. Subtracting $A_{N,1}^{[(1,2)]} - \mathbf{g}_{1,(N:2N-2)} A^{\mathrm{T}}_{(N:2N-2)}$ from $A_1^{[(1,2)]}, A_2^{[(1,2)]}, \cdots, A_{N-1}^{[(1,2)]}$, together with $A_{N,2}^{[(1,2)]} = \mathbf{g}_2 B^{\mathrm{T}}_{(1:2N-1)}$, we obtain N independent linear combinations of $B_{(1:N-1)}, B_{2N-1}$, from which $B_{(1:N-1)}, B_{2N-1}$ can be solved as

$$\begin{bmatrix} B^{\mathrm{T}}_{(1:N-1)} \\ B_{2N-1}] \end{bmatrix} = \begin{bmatrix} [\mathbf{v}_{1,(1:N-1)}, v_{1,2N-1}] \\ \vdots \\ [\mathbf{v}_{N-1,(1:N-1)}, v_{N-1,2N-1}] \\ [\mathbf{g}_{2,(1:N-1)}, g_{2,2N-1}] \end{bmatrix}^{-1} \mathbf{y}' \tag{68}$$

where

$$\begin{aligned} \mathbf{y}' \; = & \begin{bmatrix} A_1^{[(1,2)]} - A_{N,1}^{[(1,2)]} + \mathbf{g}_{1,(N:2N-2)} A^{\mathrm{T}}_{(N:2N-2)} \\ \vdots \\ A_{N-1}^{[(1,2)]} - A_{N,1}^{[(1,2)]} + \mathbf{g}_{1,(N:2N-2)} A^{\mathrm{T}}_{(N:2N-2)} \\ A_{N,1}^{[(1,2)]} \end{bmatrix} \\ & - \begin{bmatrix} \mathbf{u}_{1,(N:2N-2)} & \mathbf{v}_{1,(N:2N-2)} \\ \vdots & \\ \mathbf{u}_{N-1,(N:2N-2)} & \mathbf{v}_{N-1,(N:2N-2)} \\ \mathbf{0}_{N-1} & \mathbf{g}_{2,(N:2N-2)} \end{bmatrix} \begin{bmatrix} A^{\mathrm{T}}_{(N:2N-2)} \\ B^{\mathrm{T}}_{(N:2N-2)} \end{bmatrix}. \end{aligned} \tag{69}$$

Since the symbols $A_{(N:2N-2)}, B_{(N:2N-2)}$ are already cached by user 2, it can correctly decode the symbols $B_{(1:N-1)}, B_{2N-1}$ and recover message B. Therefore, both users can correctly decode their desired messages.

Privacy: First note that each DB only knows its own linear combination coefficients and does not know what symbols are aligned and what coefficients form full rank matrices (since the linear coefficients are picked randomly and independently). Since for different configurations of the symbol alignment and full rank coefficient matrices, different messages can be decoded by the users, each individual DB can not determine the actual user demands (θ_1, θ_2). Therefore, the proposed scheme is private.

Performance: Since $D = N + 1$ linear combinations are downloaded in total, the achieved load is $R = \frac{D}{F} = \frac{N+1}{2N-1}$.

V. Conclusion

In this paper, we formulated the problem of cache-aided multiuser Private Information Retrieval (MuPIR) where the users wish to retrieve a set of messages from a set of distributed databases while keeping the user demands private from any individual database. For the MuPIR problem with two users, two messages and arbitrary number of databases, general achievable schemes utilizing the novel idea of PCIA were proposed for the memory-load pairs $\left(\frac{N-1}{2N}, \frac{N+1}{N}\right)$ and $\left(\frac{2(N-1)}{2N-1}, \frac{N+1}{2N-1}\right)$. This is the very first achievability result on the cache-aided MuPIR problem with arbitrary number of databases. On-going directions include: 1) Characterization of the optimal memory-load trade-off for the MuPIR problem with two messages, two users and arbitrary number of databases; 2) Extending the achievability to more general cases where the number of messages, users and databases are arbitrary. Future directions may include extension to colluding databases and characterization of the optimal memory-load trade-off under the assumption of uncoded cache placement.

Acknowledgement

This work is supported through the INL Laboratory Directed Research& Development (LDRD) Program under DOE Idaho Operations Office Contract DE-AC07-05ID14517. The work of M. Ji is supported in part by NSF Awards 1817154 and 1824558.

References

[1] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *Proceedings of IEEE 36th Annual Foundations of Computer Science*. IEEE, 1995, pp. 41–50.

[2] H. Sun and S. A. Jafar, "Blind interference alignment for private information retrieval," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 560–564.

[3] ——, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4075–4088, 2017.

[4] S. A. Jafar, "Blind interference alignment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 3, pp. 216–227, June 2012.

[5] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *Information Theory, IEEE Transactions on*, vol. 60, no. 5, pp. 2856–2867, 2014.

[6] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "Characterizing the rate-memory tradeoff in cache networks within a factor of 2," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 647–663, 2018.

[7] K. Wan, D. Tuninetti, and P. Piantanida, "On the optimality of uncoded cache placement," in *2016 IEEE Information Theory Workshop (ITW)*, Sept 2016, pp. 161–165.

[8] R. Tandon, "The capacity of cache aided private information retrieval," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2017, pp. 1078–1082.

[9] Y.-P. Wei, K. Banawan, and S. Ulukus, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 3215–3232, 2018.

[10] Y. Wei, K. Banawan, and S. Ulukus, "Private information retrieval with partially known private side information," in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, March 2018, pp. 1–6.

[11] K. Wan and G. Caire, "On coded caching with private demands," *arXiv preprint arXiv:1908.10821*, 2019.

[12] K. Wan, H. Sun, M. Ji, D. Tuninetti, and G. Caire, "Device-to-device private caching with trusted server," *arXiv preprint arXiv:1909.12748*, 2019.

[13] S. Kamath, "Demand private coded caching," *arXiv preprint arXiv:1909.03324*, 2019.

[14] V. R. Aravind, P. Sarvepalli, and A. Thangaraj, "Subpacketization in coded caching with demand privacy," *arXiv preprint arXiv:1909.10471*, 2019.

[15] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," *in IEEE Int. Symp. Inf. Theory*, June. 2015.

[16] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr, "Fundamental limits of cache-aided interference management," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3092–3107, May 2017.

[17] F. Xu, M. Tao, and K. Liu, "Fundamental tradeoff between storage and latency in cache-aided wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7464–7491, Nov 2017.

[18] J. Hachem, U. Niesen, and S. N. Diggavi, "Degrees of freedom of cache-aided wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5359–5380, Jul. 2018.

[19] C. Tian, J. S. Plank, and B. Hurst, "An open-source toolbox for computer-aided investigation on the fundamental limits of information systems, version 0.1," https://github.com/ct2641/CAI/releases/tag/0.1, October 2019.