Cache-aided Multiuser Private Information Retrieval

Xiang Zhang*, Kai Wan†, Hua Sun‡ and Mingyue Ji*
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@tu-berlin.de, ‡hua.sun@unt.edu

***Abstract*—This paper formulates the cache-aided multi-user Private Information Retrieval (MuPIR) problem, including K_u cache-equipped users, each of which wishes to retrieve a desired message efficiently from N distributed databases with access to K independent messages. Privacy of the users' demands requires that any individual database can not learn anything about the demands of the users. The *load* of this problem is defined as the average number of downloaded bits per desired message bit. The goal is to find the optimal memory-load trade-off while preserving the demand privacy. Besides the formulation of the MuPIR problem, the contribution of this paper is two-fold. First, we characterize the optimal memory-load trade-off for a system with $N = 2$ databases, $K = 2$ messages and $K_u = 2$ users demanding distinct messages; Second, a *product design* with order optimality guarantee is proposed. In addition, the product design can achieve the optimal load when the cache memory is large enough. The product design embeds the well-known Sun-Jafar PIR scheme into coded caching, in order to benefit from the coded caching gain while preserving the privacy of the users' demands.**

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], which strictly outperforms the previously best-known result $1-\frac{1}{N}$ [3]. Many variants of the PIR problem have been studied. In [4], the PIR capacity for arbitrary message length was characterized since the original scheme of [2] only deals with messages of certain length. Multi-message PIR was considered by [5] where the user demands multiple messages at a time and the new achievable scheme outperforms the simple concatenation of multiple rounds of the Sun-Jafar scheme in [2]. PIR with storage-constrained databases was considered in [6]–[11] where the capacity for MDS-coded and uncoded storage-constrained databases were characterized.

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 [12]–[14], the *user-against-database* privacy model is studied where individual databases are prevented from learning the single-user's demand. The author in [12] studied the case where a single cache-aided user is connected to a set of N replicated databases and showed that memory sharing 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 [13], [14]. More specifically, the authors in [13] studied the case where the users' cached content is uncoded and unknown to the databases, and the achievability therein strictly outperforms the scheme of [12], demonstrating the unawareness gain. The authors in [14] studied a similar setting where the user cache is partially known the databases. Except certain cache memory regimes, the capacity characterization of the cache-aided single-user PIR problem with unknown cache placement remains an open problem. Another line of research [15]–[17] deals with the *user-against-user* privacy model where users are prevented from learning each other's demands. The authors in [15] 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 [18], the authors studied the subpacketizaiton issues for this problem. Later, coded caching with private demands was extended to the Device-to-Device (D2D) scenario [16]. In general, the exact capacity characterization still remains open for these problems.

This paper formulates the cache-aided multi-user (MuPIR) problem, where each of the K_u cache-equipped users wishes to retrieve a message from N distributed databases while preserving the privacy of user demands given that the cached content at all users are known to the databases. The main contribution of this paper includes:

1) Characterization of the optimal memory-load trade-off for the two-user two-message two-database system: Under the assumption of distinct user demands, the optimal memory-load trade-off for the $K_u = K = N = 2$ system is characterized for arbitrary cache memory size $M(0 \leq M \leq 2)$.

2) Product design: We show that the multicast gain of coded caching can be efficiently exploited via the incorporation of the general PIR codes into the coded deliveries, leading to the idea of the product design. By comparing with existing caching converse bounds, the product design is shown to be order optimal within a multiplicative factor of 8.

978-1-7281-6432-8/20/$31.00 ©2020 IEEE

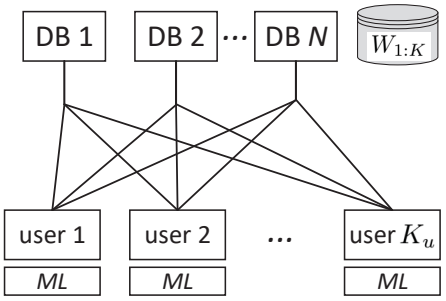

Fig. 1. Cache-aided MuPIR system with N replicated databases, K independent messages and K_u cache-equipped users. The users are connected to each DB via an error-free shared-link broadcast channel.

Notation Convention: $|\cdot|$ represents the cardinality of a set. $\mathbb{Z}^+$ denotes the set of non-negative integers. $[n] := \{1, 2, \cdots, n-1, n\}$ and $[m:n] := \{m, m+1, m+2, \cdots, n\}$ for some integers $m \leq n$. 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. The operator $\oplus$ denotes the bit-wise XOR.

II. Problem Formulation

We consider a system (See Fig. 1.) with K_u users, each of which wishes to privately retrieve a message from $N \geq 2$ 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 $0 \leq M \leq K$. Let the random variables $Z_1, Z_2, \cdots, Z_{K_u}$ denote the cached content of all users. The system operates in two phases, a *cache placement phase* followed by a *private delivery phase*. In the cache placement phase, all 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 $W_{\theta_k} (\theta_k \in [K])$. Let $\boldsymbol{\theta} := (\theta_1, \theta_2, \cdots, \theta_{K_u})$ be the demands of the users. 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 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 $\forall k \in [K_u]$:

$$H(W_{\theta_k} | Q_{1:N}^{[\boldsymbol{\theta}]}, A_{1:N}^{[\boldsymbol{\theta}]}, Z_k) = 0, \tag{2}$$

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 $\forall n \in [N], \forall \boldsymbol{\theta} \in [K]^{K_u}$:

$$I(\boldsymbol{\theta}; Q_n^{[\boldsymbol{\theta}]}, A_n^{[\boldsymbol{\theta}]}, W_{1:K}, Z_{1:K}) = 0 \tag{3}$$

The *load* (or transmission rate) of the MuPIR problem, denoted by R, is defined as the average 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 $0 \leq M \leq K$, let $R^\star(M)$ denote the minimal achievable load.[1]

III. Main Result

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

Theorem 1: For the MuPIR problem with parameters $N = 2$ DBs, $K = 2$ messages and $K_u = 2$ users demanding distinct messages, the optimal memory-load trade-off is characterized as $\forall\, 0 \leq M \leq 2$:

$$R^\star(M) = \max\left\{2(1-M), \frac{5}{3} - M, \frac{3(2-M)}{4}\right\} \tag{5}$$

Proof: See Section IV. ∎

Remark 1: When $1 \leq M \leq 2$, the load $R(M) = \frac{3(2-M)}{4}$ is actually optimal for arbitrary demands since it can be achieved by the product design described in Section V.

Theorem 2: The proposed product design achieves the load of $R(M) = \min\{K(1 - \frac{M}{K}), R'(M)\}$ in which

$$R'(M) = \frac{K_u - t}{t+1}\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right), \tag{6}$$

where $t = \frac{K_u M}{N} \in \mathbb{Z}^+$. For non-integer values of t, the lower convex envelope the integer points $(t, R'(M))$ can be achieved. Moreover, the achieved rate $R(M)$ is order optimal within a factor of 8, i.e, $\frac{R(M)}{R^\star(M)} \leq 8$.

Proof: See Section V for achievability. Note that when $K(1 - \frac{M}{K}) \leq R'(M)$, there is a naive design as follows. We let each user cache the same $\frac{M}{K}$ portion of each message. In the private delivery phase, the remaining $1 - \frac{M}{M}$ portion of each message is broadcast to the users. It can be easily seen that this naive design is correct and private (each user can correctly decode any of the K messages), and the achieved load is $K(1 - \frac{M}{K})$. The order optimality is explained as follows. Note that when $N \geq 2$, the PIR retrieval component is upper

[1] Note that the capacity as defined in [4] is given by $C = \frac{1}{R^\star}$.

bounded by 2, i.e., $1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}} \leq 2$. Let $R^{\star}_{\text{peak}}$ denote the optimal coded caching peak rate of a shared-link coded caching system with K_u users each having cache memory M. The caching load component $R_{\text{peak}} = \min\{\frac{K_u-t}{t+1}, K(1-\frac{M}{K})\}$ is shown to be optimal within a factor of 4 by [19]. Since the optimal load $R^{\star}(M)$ with demand privacy is lower bounded by the load without privacy, i.e., $R^{\star} \geq R^{\star}_{\text{peak}}$, we have

$$\frac{R(M)}{R^{\star}(M)} \leq \frac{R(M)}{R^{\star}_{\text{peak}}} = \frac{R_{\text{peak}}}{R^{\star}_{\text{peak}}}(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}) \leq 8, \quad (7)$$

which completes the proof of order optimality. ∎

Corollary 1: If $K \geq K_u$, the product design load (Eq. (6)) is optimal when $\frac{(K_u-1)}{K_u}K \leq M \leq K$.

Proof: For the case of $K \geq K_u$, when $M = \frac{K(K_u-1)}{K_u}$(i.e., $t = K_u-1$), the proposed product design achieves the load $R'(M) = \frac{1}{K_u}\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}\right)$. On the other hand, the author in [12] showed that when $K_u = 1$, the optimal cache-aided single-user PIR load is equal to $R^{\text{single-user}}(M) = \left(1-\frac{M}{K}\right)\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}\right)$. It can be seen that when $M = \frac{K(K_u-1)}{K_u}$, the product design achieves the same load as the single-user PIR load, implying its optimality. By memory sharing between $(\frac{K(K_u-1)}{K_u}, R(\frac{K(K_u-1)}{K_u}))$ and $(K, 0)$, we conclude that the product design is optimal when $\frac{(K_u-1)}{K_u}K \leq M \leq K$. ∎

IV. Proof of Theorem 1

In this section we present the characterization of the optimal memory-load trade-off for the MuPIR problem with $K = 2$ messages W_1 and W_2, $N = 2$ DBs and $K_u = 2$ users demanding distinct messages, which provides the proof of Theorem 1. The achievability and converse are described as follows.

A. Achievability

We propose schemes achieving the memory-load pairs $(0, 2), (\frac{1}{3}, \frac{4}{3}), (\frac{2}{3}, 1)$, and $(2, 0)$. Any other point on the lower convex envelope of the above corner points can be achieved by memory sharing.

1) Points $(0, 2)$ and $(2, 0)$: When $M = 0$, let either of the two DBs broadcast the two messages to both users. It is easy to check the correctness and privacy. When $M = 2$, let each user cache the two files, then there is no need to download anything from the DBs, which is trivially private.

2) Point $(\frac{1}{3}, \frac{4}{3})$: Assume that the users have distinct demands, i.e., the demand vector (θ_1, θ_2) can only be $(1, 2)$ or $(2, 1)$. Let $A_{1,1}$ and $A_{1,2}$ be two different answers from DB 1, and let $A_{2,1}$ and $A_{2,2}$ be two different answers from DB 2. Assume that each message contains $L = 3$ bits, i.e., $W_1 = (a_1, a_2, a_3), W_2 = (b_1, b_2, b_3)$. The cache placement is $Z_1 = \{a_1 \oplus b_1\}, Z_2 = \{a_2 \oplus b_2\}$ and the answers are constructed as

$$A_{1,1} = (a_3, b_1 \oplus b_2 \oplus b_3),\ A_{2,1} = (a_2 \oplus a_3, b_2 \oplus b_3) \quad (8)$$

$$A_{1,2} = (a_1 \oplus a_2 \oplus a_3, b_3),\ A_{2,2} = (a_1 \oplus a_3, b_1 \oplus b_3) \quad (9)$$

The users randomly choose $A_{1,1}$ or $A_{1,2}$ to request from DB 1 with equal probabilities. We then consider the two cases:

- $(\theta_1, \theta_2) = (1, 2)$. If $A_{1,1}$ is chosen, then go to DB 2 to download $A_{2,1}$; Otherwise, if $A_{1,2}$ is chosen, go to DB 2 to download $A_{2,2}$.
- $(\theta_1, \theta_2) = (2, 1)$. If $A_{1,1}$ is chosen, then go to DB 2 to download $A_{2,2}$; Otherwise if $A_{1,2}$ is chosen, go to DB 2 to download $A_{2,1}$.

This scheme is both correct and private due to the following reasons. For correctness, one can check that

$$(A_{1,1}, A_{2,1}, Z_1) \to W_1,\ (A_{1,1}, A_{2,1}, Z_2) \to W_2 \quad (10)$$

$$(A_{1,2}, A_{2,2}, Z_1) \to W_1,\ (A_{1,2}, A_{2,2}, Z_2) \to W_2 \quad (11)$$

$$(A_{1,1}, A_{2,2}, Z_1) \to W_2,\ (A_{1,1}, A_{2,2}, Z_2) \to W_1 \quad (12)$$

$$(A_{1,2}, A_{2,1}, Z_1) \to W_2,\ (A_{1,2}, A_{2,1}, Z_2) \to W_1 \quad (13)$$

Therefore, all users can decode their desired messages. For privacy, note that the answer from DB 1 is equally likely to be $A_{1,1}$ or $A_{1,2}$, and the answer from DB 2 is also equally likely to be $A_{2,1}$ or $A_{2,2}$. Therefore, we have

$$P\{\boldsymbol{\theta} = (1, 2)\} = P\{(A_{1,1}, A_{2,1})\} + P\{(A_{1,2}, A_{2,2})\} = \frac{1}{2} \quad (14)$$

$$P\{\boldsymbol{\theta} = (2, 1)\} = P\{(A_{1,1}, A_{2,2})\} + P\{(A_{1,2}, A_{2,1})\} = \frac{1}{2} \quad (15)$$

i.e., the demand vector $\boldsymbol{\theta}$ is equally likely to be $(1, 2)$ or $(2, 1)$ from each DB's perspective. Therefore, the privacy constraint (3) is satisfied (for distinct demands). Since $D = 4$ bits are downloaded in total, the achieved load is $R = \frac{D}{L} = \frac{4}{3}$.

3) Point $(\frac{2}{3}, 1)$: Let $L = 3$ and $W_1 = (a_1, a_2, a_3), W_2 = (b_1, b_2, b_3)$. The cache placement is $Z_1 = \{a_1, b_1\}, Z_2 = \{a_2, b_2\}$ and the answers are constructed as

$$A_{1,1} = (a_3 \oplus b_3 \oplus b_1 \oplus b_2),\ A_{2,1} = (a_2 \oplus a_3, b_2 \oplus b_3) \quad (16)$$

$$A_{1,2} = (a_3 \oplus b_3 \oplus a_1 \oplus a_2),\ A_{2,2} = (a_1 \oplus a_3, b_1 \oplus b_3) \quad (17)$$

The private delivery phase works similarly to the above corner point $(\frac{1}{3}, \frac{4}{3})$. The correctness of this scheme can be checked via Eqs. (10)-(13). The privacy argument is similar, i.e., from each DB's perspective, the demand vector $\boldsymbol{\theta}$ is equally likely to be $(1, 2)$ or $(2, 1)$. Since $D = 3$ bits are downloaded in total, the achieved load is $R = \frac{D}{L} = 1$. This completes the achievability proof of Theorem 1.

B. Converse

The converse curve consists of three piece-wise linear segments: $R(M) = 2(1-M)$ for $0 \leq M \leq \frac{1}{3}$, $R(M) = \frac{5}{3} - M$ for $\frac{1}{3} \leq M \leq \frac{2}{3}$, and $R(M) = \frac{3(2-M)}{4}$ for $\frac{2}{3} \leq M \leq 2$. We now prove the three segments respectively.

1) $0 \leq M \leq \frac{1}{3}$: In this regime, the cut-set bound without the privacy constriant is tight. Let $A_1^{[(1,2)]}$ and $A_2^{[(1,2)]}$ be two answers from DB 1 and DB 2 respectively when the user

demands are $\boldsymbol{\theta} = (1, 2)$. Since W_1, W_2 can be recovered from $\{A_1^{[(1,2)]}, A_2^{[(1,2)]}, Z_1, Z_2\}$, we have

$$H(A_1^{[(1,2)]}) + H(A_2^{[(1,2)]}) + 2ML \tag{18a}$$
$$\geq H(A_1^{[(1,2)]}) + H(A_1^{[(1,2)]}) + H(Z_1) + H(Z_2) \tag{18b}$$
$$\geq H(A_1^{[(1,2)]}, A_2^{[(1,2)]}, Z_1, Z_2) \tag{18c}$$
$$\geq 2L \tag{18d}$$

Similarly, we can obtain $H(A_1^{[(2,1)]}) + H(A_2^{[(2,1)]}) + 2ML \geq 2L$ for another two answers $A_1^{[(2,1)]}$ and $A_2^{[(2,1)]}$ corresponding to $\boldsymbol{\theta} = (2, 1)$. Therefore we have $RL = \frac{1}{2}(H(A_1^{[(1,2)]}) + H(A_2^{[(1,2)]}) + \frac{1}{2}(H(A_1^{[(2,1)]}) + H(A_2^{[(2,1)]})) \geq 2(1 - M)L$, yielding $R(M) \geq 2(1 - M)$.

2) $\frac{1}{3} \leq M \leq \frac{2}{3}$: Let $A_{1,1} = A_{1,1}^{[(1,2)]} = A_{1,1}^{[(2,1)]}$ be an answer of DB 1. Let $A_{2,1} = A_{2,1}^{[(1,2)]} = A_{2,1}^{[(2,1)]}$ be an answer of DB 2. It is clear that $(A_{1,1}^{[(1,2)]}, A_{2,1}^{[(1,2)]}, Z_1) \to W_1$, $(A_{1,1}^{[(1,2)]}, A_{2,1}^{[(1,2)]}, Z_2) \to W_2$. For privacy of DB 1, there must exist another answer $A_{2,2}^{[(2,1)]}$ of DB 2 such that $(A_{1,1}^{[(2,1)]}, A_{2,2}^{[(2,1)]}, Z_1) \to W_2$ and $(A_{1,1}^{[(2,1)]}, A_{2,2}^{[(2,1)]}, Z_2) \to W_1$. Also, for privacy of DB 2, there must exist another answer $A_{1,2}^{[(2,1)]}$ of DB 1 such that $(A_{1,2}^{[(2,1)]}, A_{2,1}^{[(2,1)]}, Z_1) \to W_2$ and $(A_{1,2}^{[(2,1)]}, A_{2,1}^{[(2,1)]}, Z_2) \to W_1$. Denote $R^{[\boldsymbol{\theta}]}(M) := (H(A_{1,i}^{[\boldsymbol{\theta}]}) + H(A_{2,j}^{[\boldsymbol{\theta}]}))/L, \forall i, j \in [2]$ and $X_{i,j,k}^{[\boldsymbol{\theta}]} := (A_{1,i}^{[\boldsymbol{\theta}]}, A_{2,j}^{[\boldsymbol{\theta}]}, Z_k), \forall i, j, k \in [2]$. We have

$$R^{[(1,2)]}(M)L + 2R^{[(2,1)]}(M)L + 3ML \tag{19a}$$
$$= H(A_{1,1}^{[(1,2)]}) + H(A_{2,1}^{[(1,2)]}) + H(A_{1,2}^{[(2,1)]}) + H(A_{2,1}^{[(2,1)]})$$
$$+ H(A_{1,1}^{[(2,1)]}) + H(A_{2,2}^{[(2,1)]}) + H(Z_1) + 2H(Z_2) \tag{19b}$$
$$\geq H(X_{1,1,1}^{[(1,2)]}) + H(X_{2,1,2}^{[(2,1)]}) + H(X_{1,2,2}^{[(2,1)]}) \tag{19c}$$
$$= 3L + H(X_{1,1,1}^{[(1,2)]}|W_1) + H(X_{2,1,2}^{[(2,1)]}|W_1) + H(X_{1,2,2}^{[(2,1)]}|W_1) \tag{19d}$$
$$\geq 3L + H(X_{1,1,1}^{[(1,2)]}|W_1) + H(Z_2|W_1) + H(A_{2,1}^{[(2,1)]}|W_1, Z_2)$$
$$+ H(X_{1,2,2}^{[(2,1)]}|W_1) \tag{19e}$$
$$\geq 3L + H(X_{1,1,1}^{[(1,2)]}, Z_2|W_1) + H(A_{2,1}^{[(2,1)]}|W_1, Z_2)$$
$$+ H(X_{1,2,2}^{[(2,1)]}|W_1) \tag{19f}$$
$$= 4L + H(A_{2,1}^{[(2,1)]}|W_1, Z_2) + H(X_{1,2,2}^{[(2,1)]}|W_1) \tag{19g}$$
$$\geq 4L + H(A_{2,1}^{[(2,1)]}|W_1, Z_2) + H(Z_2|W_1) + H(A_{1,1}^{[(2,1)]}|W_1, Z_2) \tag{19h}$$
$$\geq 4L + H(A_{1,1}^{[(2,1)]}, A_{2,1}^{[(2,1)]}|W_1, Z_2) + H(Z_2|W_1) \tag{19i}$$
$$= 4L + H(A_{1,1}^{[(2,1)]}, A_{2,1}^{[(2,1)]}, Z_2|W_1) \tag{19j}$$
$$= 4L + H(A_{1,1}^{[(1,2)]}, A_{2,1}^{[(1,2)]}, Z_2|W_1) \tag{19k}$$
$$= 5L \tag{19l}$$

Similarly, when assuming a different decoding structure, we can obtain $R^{[(1,2)]}(M)L + 2R^{[(2,1)]}(M)L + 3ML \geq 5L$. Therefore, we have $3R^{[(1,2)]}(M)L + 3R^{[(2,1)]}(M)L + 6ML \geq 10L$, which gives $R(M) = \frac{1}{2}(R^{[(1,2)]}(M) + R^{[(2,1)]}(M)) \geq \frac{5}{3} - M$.

3) $\frac{2}{3} \leq M \leq 2$: In this regime, the achievable load $\frac{3(2-M)}{4}$ coincides with the single-user cache-aided PIR bound $R^{\text{single-user}}(M) = \frac{3(2-M)}{4}$ given in [12]. Since increasing the number of users while keeping the user demands private from the DBs can only possibly increase the load, we conclude that $R(M) \geq \frac{3(2-M)}{4}$. This completes the converse proof of Theorem 1.

V. Product Design

In this section, we present the product design which is inspired by both coded caching and the Sun-Jafar PIR schemes and enjoys combined coding benefits from both coded caching and PIR. An example is provided to illustrate the basic idea. By comparing with the already established converse bounds for caching, we show that the product design is optimal within a factor 8 in general.

Example 1: Consider the MuPIR problem with $K = 3$ messages, $N = 2$ DBs and $K_u = 3$ users with cache memory $M = 1$ (therefore $t = \frac{K_u M}{N} = 1$). Let $W_1 = A, W_2 = B$ and $W_3 = C$ denote the three messages. Each message is assumed to have $L = 24$ bits. The cache placement and private delivery phases are described as follows.

1) Cache placement: The Maddah-Ali-Niesen (MAN) cache placment [20] is used. More specifically, each message is split into three packets each containing 8 bits, i.e., $A = (a_1, a_2, a_3), B = (b_1, b_2, b_3)$ and $C = (c_1, c_2, c_3)$. The cache placement is $Z_1 = \{a_1, b_1, c_1\}, Z_2 = \{a_2, b_2, c_2\}$ and $Z_3 = \{a_3, b_3, c_3\}$.

2) Private delivery: Suppose the user demands are $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3] = [1, 2, 3]$. We first construct three different coded messages $\{X_{\mathcal{S}} = (A_{1,\mathcal{S}}^{[\boldsymbol{\theta}]}, A_{2,\mathcal{S}}^{[\boldsymbol{\theta}]}) : \mathcal{S} \subseteq [3], |\mathcal{S}| = 2\}$ each being useful to a subset of two users in $\mathcal{S}$. $A_{1,\mathcal{S}}^{[\boldsymbol{\theta}]}$ and $A_{2,\mathcal{S}}^{[\boldsymbol{\theta}]}$ represents the answers from DB 1 and 2 respectively.

The first coded message is $X_{\{1,2\}} = (A_{1,\{1,2\}}^{[\boldsymbol{\theta}]}, A_{2,\{1,2\}}^{[\boldsymbol{\theta}]})$ in which

$$A_{1,\{1,2\}}^{[\boldsymbol{\theta}]} = A_1^{[\theta_1]}(a_2, b_2, c_2) \oplus A_1^{[\theta_2]}(a_1, b_1, c_1), \tag{20}$$

$$A_{2,\{1,2\}}^{[\boldsymbol{\theta}]} = A_2^{[\theta_1]}(a_2, b_2, c_2) \oplus A_2^{[\theta_2]}(a_1, b_1, c_1), \tag{21}$$

where the code components $A_1^{[\theta_1]}(a_2, b_2, c_2)$ and $A_2^{[\theta_1]}(a_2, b_2, c_2)$ represents the answer from DB 1 and DB 2 respectively in the Sun-Jafar PIR scheme when the messages are (First message, second message, third message) $= (a_2, b_2, c_2)$ and the user demands a_2. The meaning of $A_1^{[\theta_2]}(a_1, b_1, c_1)$ and $A_2^{[\theta_2]}(a_1, b_1, c_1)$ follow similarly. More specifically, let $a_i = (a_i^1, a_i^2, \cdots, a_i^8), b_i = (b_i^1, b_i^2, \cdots, b_i^8)$ and $c_i = (c_i^1, c_i^2, \cdots, c_i^8), \forall i \in [2]$ be six independent random permutations of the bits of the packets $a_i, b_i, c_i, i \in [2]$. Then the answers from the two DBs are constructed as

$A^{[\boldsymbol{\theta}]}_{1,\{1,2\}}$	$A^{[\boldsymbol{\theta}]}_{2,\{1,2\}}$
$a_2^1 \oplus a_1^1$	$a_2^2 \oplus a_1^2$
$b_2^1 \oplus b_1^1$	$b_2^2 \oplus b_1^2$
$c_2^1 \oplus c_1^1$	$c_2^2 \oplus c_1^2$
$a_2^3 \oplus b_2^3 \oplus a_1^2 \oplus b_1^2$	$a_2^5 \oplus b_2^1 \oplus a_1^1 \oplus b_1^5$
$a_2^4 \oplus c_2^2 \oplus a_1^3 \oplus c_1^3$	$a_2^6 \oplus c_2^1 \oplus a_1^4 \oplus c_1^4$
$b_2^3 \oplus c_2^3 \oplus b_1^4 \oplus c_1^2$	$b_2^4 \oplus c_2^4 \oplus b_1^6 \oplus c_1^1$
$a_2^7 \oplus b_2^4 \oplus c_2^4 \oplus a_1^4 \oplus b_1^7 \oplus c_1^4$	$a_2^8 \oplus b_2^3 \oplus c_2^3 \oplus a_1^3 \oplus b_1^8 \oplus c_1^3$

The second coded message is $X_{\{1,3\}} = (A^{[\boldsymbol{\theta}]}_{1,\{1,3\}}, A^{[\boldsymbol{\theta}]}_{2,\{1,3\}})$ in which

$$A^{[\boldsymbol{\theta}]}_{1,\{1,3\}} = A_1^{[\theta_1]}(a_3, b_3, c_3) \oplus A_1^{[\theta_3]}(a_1, b_1, c_1), \quad (22)$$

$$A^{[\boldsymbol{\theta}]}_{2,\{1,3\}} = A_2^{[\theta_1]}(a_3, b_3, c_3) \oplus A_2^{[\theta_3]}(a_1, b_1, c_1). \quad (23)$$

Using a another set of independent random permutations of the bits of the packets $a_i, b_i, c_i, \forall i \in \{1, 3\}$, the answers form the two DBs are constructed as

$A^{[\boldsymbol{\theta}]}_{1,\{1,3\}}$	$A^{[\boldsymbol{\theta}]}_{2,\{1,3\}}$
$a_3^1 \oplus a_1^1$	$a_3^2 \oplus a_1^2$
$b_3^1 \oplus b_1^1$	$b_3^2 \oplus b_1^2$
$c_3^1 \oplus c_1^1$	$c_3^2 \oplus c_1^2$
$a_3^3 \oplus b_3^2 \oplus a_1^3 \oplus b_1^3$	$a_3^5 \oplus b_3^1 \oplus a_1^4 \oplus b_1^4$
$a_3^4 \oplus c_3^2 \oplus a_1^2 \oplus c_1^3$	$a_3^6 \oplus c_3^1 \oplus a_1^1 \oplus c_1^5$
$b_3^3 \oplus c_3^3 \oplus b_1^2 \oplus c_1^4$	$b_3^4 \oplus c_3^4 \oplus b_1^1 \oplus c_1^6$
$a_3^7 \oplus b_3^4 \oplus c_3^4 \oplus a_1^4 \oplus b_1^4 \oplus c_1^7$	$a_3^8 \oplus b_3^3 \oplus c_3^3 \oplus a_1^3 \oplus b_1^3 \oplus c_1^8$

The third coded message is $X_{\{2,3\}} = (A^{[\boldsymbol{\theta}]}_{1,\{2,3\}}, A^{[\boldsymbol{\theta}]}_{2,\{2,3\}})$ in which

$$A^{[\boldsymbol{\theta}]}_{1,\{2,3\}} = A_1^{[\theta_2]}(a_3, b_3) \oplus A_1^{[\theta_3]}(a_2, b_2), \quad (24)$$

$$A^{[\boldsymbol{\theta}]}_{2,\{2,3\}} = A_2^{[\theta_2]}(a_3, b_3) \oplus A_2^{[\theta_3]}(a_2, b_2). \quad (25)$$

Applying a set of independent random permutations of the bits of the message packets $a_i, b_i, c_i, \forall i \in \{2, 3\}$, the answers can be constructed as

$A^{[\boldsymbol{\theta}]}_{1,\{2,3\}}$	$A^{[\boldsymbol{\theta}]}_{2,\{2,3\}}$
$a_3^1 \oplus a_2^1$	$a_3^2 \oplus a_2^2$
$b_3^1 \oplus b_2^1$	$b_3^2 \oplus b_2^2$
$c_3^1 \oplus c_2^1$	$c_3^2 \oplus c_2^2$
$a_3^2 \oplus b_3^3 \oplus a_2^3 \oplus b_2^3$	$a_3^1 \oplus b_3^5 \oplus a_2^4 \oplus b_2^4$
$a_3^3 \oplus c_3^3 \oplus a_2^2 \oplus c_2^3$	$a_3^4 \oplus c_3^4 \oplus a_2^1 \oplus c_2^5$
$b_3^4 \oplus c_3^2 \oplus b_2^2 \oplus c_2^4$	$b_3^6 \oplus c_3^1 \oplus b_2^1 \oplus c_2^6$
$a_3^4 \oplus b_3^7 \oplus c_3^4 \oplus a_2^4 \oplus b_2^4 \oplus c_2^7$	$a_3^3 \oplus b_3^8 \oplus c_3^3 \oplus a_2^3 \oplus b_2^3 \oplus c_2^8$

Note that for each coded message $X_{\mathcal{S}}$, a set of independent random permutations (not known to the DBs) are employed to the bits of the involved packets, which is key to privacy.

In the private delivery phase, the users download all the three coded messages from the DBs. We next verify the correctness (i.e., decodability) and privacy of the scheme.

Correctness: Let us look at $X_{\{1,2\}}$ first. Since the packets have been cached by user 1, user 1 can remove the interferences $A_1^{[\theta_2]}(a_1, b_1), A_2^{[\theta_2]}(a_1, b_1)$ from Eqs. (20) (21) to obtain the desired coded components $A_1^{[\theta_1]}(a_2, b_2)$ and $A_2^{[\theta_1]}(a_2, b_2)$. By the decodability of the Sun-Jafar PIR scheme, user 1 can correctly decode the desired packet a_2 from $A_1^{[\theta_1]}(a_2, b_2)$ and $A_2^{[\theta_1]}(a_2, b_2)$; Also because the packets a_2, b_2 are already cached by user 2, user 2 can remove the interferences $A_1^{[\theta_1]}(a_2, b_2)$ and $A_2^{[\theta_1]}(a_2, b_2)$ and obtain the desired code components $A_1^{[\theta_2]}(a_1, b_1)$ and $A_2^{[\theta_2]}(a_1, b_1)$, from which the packet a_1 can be decoded. Following a similar decoding process, it can be easily seen that from $X_{\{1,3\}}$, user 1 and 3 can decode a_3 and b_1 respectively, and from $X_{\{2,3\}}$, user 2 and 3 can decode a_3 and b_2 respectively. As a result,the three users can correctly decode their desired messages.

Privacy: First note that regardless of the user demands $[\theta_1, \theta_2, \theta_3] \in [2]^3$, three coded messages are downloaded from the DBs. So the DBs can not distinguish different user demands by simply observing the traffic load. Second, each component $A_{n,\mathcal{S}}, n = 1, 2$ of the coded message $X_{\mathcal{S}}, \forall \mathcal{S} \subseteq [3], |\mathcal{S}| = 2$ is independent of the demands of the users in $\mathcal{S}$. The reason is explained as follows. Without loss of generality, we show that both $A_{1,\{1,2\}}$ and $A_{2,\{1,2\}}$ are independent of θ_1 and θ_2. By the privacy of the Sun-Jafar scheme, both $A_1^{[\theta_1]}(a_2, b_2)$ and $A_2^{[\theta_1]}(a_2, b_2)$ are independent of θ_1. Also, both $A_1^{[\theta_2]}(a_1, b_1)$ and $A_2^{[\theta_2]}(a_1, b_1)$ are independent of θ_2. Therefore, both $A^{[\boldsymbol{\theta}]}_{1,\{1,2\}}$ and $A^{[\boldsymbol{\theta}]}_{2,\{1,2\}}$ are independent of θ_1 and θ_2. Similarly, both $A^{[\boldsymbol{\theta}]}_{1,\{1,3\}}$ and $A^{[\boldsymbol{\theta}]}_{2,\{1,3\}}$ are independent of θ_1 and θ_3, and both $A^{[\boldsymbol{\theta}]}_{1,\{2,3\}}$ and $A^{[\boldsymbol{\theta}]}_{2,\{2,3\}}$ are independent of θ_2 and θ_3. Moreover, since for each coded message $X_{\mathcal{S}}, \forall \mathcal{S} \subseteq [3], |\mathcal{S}| = 2$, a set of independent random permutations are applied to the corresponding packets, these coded messages are independent of each other. As a result, the answer from each DB $n \in [2]$, i.e., $\{A^{[\boldsymbol{\theta}]}_{n,\mathcal{S}} : \mathcal{S} \subseteq [3], |\mathcal{S}| = 2\}$ is independent of the user demands $[\theta_1, \theta_2, \theta_3]$. As a result, the scheme is private.

Performance: Since $D = 42$ bits are downloaded in total, the achieved load is $R = \frac{D}{L} = \frac{7}{4}$, which is better than the naive design with load $K - M = 2$. $\Diamond$

VI. Conclusion

In this paper we studied the cache-aided MuPIR problem where a set of cache-aided users wish to retrieve their desired messages while keeping their demands private from the DBs. We fully characterized the optimal memory-load trade-off for a system with $N = 2$ databases, $K = 2$ messages and $K_u = 2$ users when the users demand distinct messages. We also proposed a novel product design which captures the multicasting gain of coded caching in the delivery phase and was shown to be order optimal within a factor of 8. One on-going direction is to extend the scheme in Section IV to incorporate identical user demands and more general system parameters, in order to find the optimal memory-load trade-off for the cache-aided MuPIR problem.

Acknowledgement

This work is supported through the INL Laboratory Directed Research & Development (LDRD) Program under DOE Idaho Operations Office Contract DE-AC07-05ID14517, and 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, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4075–4088, 2017.

[3] N. B. Shah, K. Rashmi, and K. Ramchandran, "One extra bit of download ensures perfectly private information retrieval," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 856–860.

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

[5] K. Banawan and S. Ulukus, "Multi-message private information retrieval: Capacity results and near-optimal schemes," *IEEE Transactions on Information Theory*, vol. 64, no. 10, pp. 6842–6862, Oct 2018.

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

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

[8] Y. Wei, B. Arasli, K. Banawan, and S. Ulukus, "Private information retrieval from decentralized uncoded caching databases," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 2114–2118.

[9] N. Woolsey, R. Chen, and M. Ji, "A new design of private information retrieval for storage constrained databases," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1052–1056.

[10] K. Banawan, B. Arasli, Y. Wei, and S. Ulukus, "Private information retrieval from heterogeneous uncoded caching databases," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1267–1271.

[11] N. Woolsey, R.-R. Chen, and M. Ji, "An optimal iterative placement algorithm for pir from heterogeneous storage-constrained databases," *arXiv preprint arXiv:1904.02131*, 2019.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

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

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

[19] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4388–4413, 2017.

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