Cache-Aided Private Information Retrieval with Partially Known Uncoded Prefetching

Yi-Peng Wei Karim Banawan Sennur Ulukus
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
ypwei@umd.edu kbanawan@umd.edu ulukus@umd.edu

***Abstract*—We consider the problem of private information retrieval (PIR) from N non-colluding and replicated databases, when the user is equipped with a cache that holds an uncoded fraction r from each of the K stored messages in the databases. This model operates in a two-phase scheme, namely, the prefetching phase where the user acquires side information and the retrieval phase where the user privately downloads the desired message. In the prefetching phase, the user receives $\frac{r}{N}$ uncoded fraction of each message from the nth database. This side information is known only to the nth database and unknown to the remaining databases, i.e., the user possesses *partially known* side information. We investigate the optimal normalized download cost $D^*(r)$ as a function of K, N, r. For a fixed K, N, we develop an inner bound (converse) and an outer bound (achievability) for the $D^*(r)$ curve. The bounds match in general for the cases of very low caching ratio ($r \leq \frac{1}{N^{K-1}}$) and very high caching ratio ($r \geq \frac{K-2}{N^2-3N+KN}$). As a corollary, we fully characterize the optimal download cost caching ratio tradeoff for $K = 3$. For general K, N, and r, we show that the largest gap between the achievability and the converse bounds is $\frac{5}{32}$.**

I. Introduction

The private information retrieval (PIR) problem considers the privacy of the requested message by a user from distributed databases. In the classical setting of PIR [1], there are N non-communicating databases, each storing the same set of K messages. The user wishes to download a message without letting the databases know the identity of the desired message. One feasible scheme is to download all the K messages from a database. However, this results in excessive download cost since it results in a download K times the size of the desired message. The goal of the PIR problem is to construct an efficient retrieval scheme such that no database knows which message is retrieved. The PIR problem has originated in computer science society [1]–[3] and has drawn significant attention in information theory society in recent years [4]–[9].

Recently, Sun and Jafar [10] have characterized the optimal normalized download cost for the classical PIR problem to be $\frac{D}{L} = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right)$, where L is the message size and D is the total number of downloaded bits from the N databases. Since the work of Sun-Jafar [10], many interesting variants of the classical PIR problem have been investigated, such as, PIR from colluding databases, robust PIR, symmetric PIR, PIR from MDS-coded databases, PIR for arbitrary message lengths, multi-round PIR, multi-message PIR, PIR from Byzantine databases, secure symmetric PIR with adversaries, and their several combinations [11]–[25].

The achievability scheme proposed in [10] is based on three principles: database symmetry, message symmetry, and side information utilization. The side information in [10] comes from the undesired bits downloaded from the other $(N-1)$ databases. Side information plays a significant role in the PIR problem. For the case of $N = 1$, i.e., when no side information is available, the normalized download cost becomes K, which is the trivial download cost. The interplay between side information and the PIR problem has been studied in [26]–[30].

Reference [26] is the first to study the cache-aided PIR problem. In [26], the user has a memory of size KLr bits and can store an arbitrary function of the K messages, where $0 \leq r \leq 1$ is the caching ratio. Reference [26] considers the case that the cached content is known to all N databases, and shows the optimal normalized download cost to be $(1-r)\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right)$. Although the result is pessimistic since it implies that the user cannot utilize the cached content to further reduce the download cost, reference [26] reveals two new dimensions for the cache-aided PIR problem. The first one is the databases' awareness of the side information at its initial acquisition. Different from [26], in [27]–[29], all the databases are assumed to be unaware of the side information. In [30], the cached content is partially known by the databases. The second one is the structure of the side information. Instead of storing an arbitrary function of the K messages, references [27], [29], [30] consider caching M full messages out of total K messages. Reference [28] considers storing r fraction of each message in uncoded form.

This paper is closely related to [28]. In [28], all databases are assumed to be unaware of the side information. However, this may be practically challenging. Here, we consider a more natural model which uses the same set of databases for both prefetching and retrieval phases. Different from [28], each database gains partial knowledge about the side information it provides during prefetching. We aim at answering the following question: How much is the rate loss due to the partial knowledge of the cache from the fully unknown case in [28]?

In this work, we consider PIR with *partially known uncoded prefetching*. We consider the PIR problem with a two-phase scheme, namely, prefetching phase and retrieval phase. In the prefetching phase, the user caches uncoded $\frac{r}{N}$ fraction of each

This work was supported by NSF Grants CNS 13-14733, CCF 14-22111, CNS 15-26608 and CCF 17-13977.

message from the nth database. The nth database is aware of these $\frac{KLr}{N}$ bits side information, while it has no knowledge about the cached bits from the other $(N-1)$ databases. We aim at characterizing the optimal tradeoff between the normalized download cost $\frac{D(r)}{L}$ and the caching ratio r. For the outer bound, we explicitly determine the achievable download rates for specific $K+1$ caching ratios. Download rates for any other caching ratio can be achieved by memory-sharing between the nearest explicit points. Hence, the outer bound is a piece-wise linear curve which consists of K line segments. For the inner bound, we extend the techniques of [10], [28] to obtain a piece-wise linear curve which also consists of K line segments. We show that the inner and the outer bounds match exactly at three line segments for any K. Consequently, we characterize the optimal tradeoff for the very low ($r \leq \frac{1}{N^{K-1}}$) and the very high ($r \geq \frac{K-2}{N^2-3N+KN}$) caching ratios. As a direct corollary, we fully characterize the optimal download cost caching ratio tradeoff for $K = 3$ messages. For general K, N and r, we show that the worst-case gap between the inner and the outer bounds is $\frac{5}{32}$.

II. System Model

We consider a PIR problem with N non-communicating databases. Each database stores an identical copy of K statistically independent messages, $W_1, \ldots, W_K$. Each message is L bits long,

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

$$H(W_1, \ldots, W_K) = H(W_1) + \cdots + H(W_K). \tag{2}$$

The user (retriever) has a local cache memory which can store up to KLr bits, where $0 \leq r \leq 1$, and r is called the *caching ratio*. There are two phases in this system: the *prefetching phase* and the *retrieval phase*.

In the prefetching phase, for each message W_k, the user randomly and independently chooses Lr bits out of the L bits to cache. The user caches the Lr bits of each message by prefetching the same amount of bits from each database, i.e., the user prefetches $\frac{KLr}{N}$ bits from each database. $\forall n \in [N]$, where $[N] = \{1, 2, \ldots, N\}$, we denote the indices of the cached bits from the nth database by $\mathbb{H}_n$ and the cached bits from the nth database by the random variable Z_n. Therefore, the overall cached content Z is equal to $(Z_1, \ldots, Z_N)$, and

$$H(Z) = \sum_{n=1}^{N} H(Z_n) = KLr. \tag{3}$$

We further denote the indices of the cached bits by $\mathbb{H}$. Therefore, we have $\mathbb{H} = \bigcup_{n=1}^{N} \mathbb{H}_n$, where $\mathbb{H}_{n_1} \cap \mathbb{H}_{n_2} = \emptyset$, if $n_1 \neq n_2$. Since the user caches a subset of the bits from each message, this is called *uncoded prefetching*. Here, we consider the case where database n knows $\mathbb{H}_n$, but it does not know $\mathbb{H} \setminus \mathbb{H}_n$. We refer to Z as *partially known prefetching*.

In the retrieval phase, the user privately generates an index $\theta \in [K]$, and wishes to retrieve message W_θ such that it is impossible for any individual database to identify θ. Note that during the prefetching phase, the desired message is unknown a priori. Therefore, the cached bit indices $\mathbb{H}$ are independent of the desired message index θ. Note further that the cached bit indices $\mathbb{H}$ are independent of the message contents. Therefore, for random variables θ, $\mathbb{H}$, and $W_1, \ldots, W_K$, we have

$$\begin{aligned} &H(\theta, \mathbb{H}, W_1, \ldots, W_K) \\ &\quad = H(\theta) + H(\mathbb{H}) + H(W_1) + \cdots + H(W_K). \end{aligned} \tag{4}$$

The user sends N queries $Q_1^{[\theta]}, \ldots, Q_N^{[\theta]}$ to the N databases, where $Q_n^{[\theta]}$ is the query sent to the nth database for message W_θ. The queries are generated according to $\mathbb{H}$, which are independent of the realizations of the K messages. Therefore,

$$I(W_1, \ldots, W_K; Q_1^{[\theta]}, \ldots, Q_N^{[\theta]}) = 0. \tag{5}$$

After receiving the query $Q_n^{[\theta]}$, the nth database replies with an answering string $A_n^{[\theta]}$, which is a function of $Q_n^{[\theta]}$ and all the K messages. Therefore, $\forall \theta \in [K], \forall n \in [N]$,

$$H(A_n^{[\theta]} | Q_n^{[\theta]}, W_1, \ldots, W_K) = 0. \tag{6}$$

After receiving the answering strings $A_1^{[\theta]}, \ldots, A_N^{[\theta]}$ from all the N databases, the user needs to decode the desired message W_θ reliably. By using Fano's inequality, we have the following reliability constraint

$$H\left(W_\theta | Z, \mathbb{H}, Q_1^{[\theta]}, \ldots, Q_N^{[\theta]}, A_1^{[\theta]}, \ldots, A_N^{[\theta]}\right) = o(L), \tag{7}$$

where $o(L)$ denotes a function such that $\frac{o(L)}{L} \to 0$ as $L \to \infty$.

To ensure that individual databases do not know which message is retrieved, we need to satisfy the following privacy constraint, $\forall n \in [N]$, $\forall \theta \in [K]$,

$$\begin{aligned} &(Q_n^{[1]}, A_n^{[1]}, W_1, \ldots, W_K, \mathbb{H}_n) \\ &\quad \sim (Q_n^{[\theta]}, A_n^{[\theta]}, W_1, \ldots, W_K, \mathbb{H}_n), \end{aligned} \tag{8}$$

where $A \sim B$ means that A and B are identically distributed.

For a fixed N, K, and caching ratio r, a pair $(D(r), L)$ is achievable if there exists a PIR scheme for message of size L bits long with partially known uncoded prefetching satisfying the privacy constraint (8) and the reliability constraint (7), where $D(r)$ represents the expected number of downloaded bits (over all the queries) from the N databases via the answering strings $A_{1:N}^{[\theta]}$, where $A_{1:N}^{[\theta]} = (A_1^{[\theta]}, \ldots, A_N^{[\theta]})$, i.e.,

$$D(r) = \sum_{n=1}^{N} H\left(A_n^{[\theta]}\right). \tag{9}$$

In this work, we aim at characterizing the optimal normalized download cost $D^*(r)$ corresponding to every caching ratio $0 \leq r \leq 1$, where

$$D^*(r) = \inf\left\{\frac{D(r)}{L} : (D(r), L) \text{ is achievable}\right\}, \tag{10}$$

which is a function of the caching ratio r.

III. Main Results

We provide a PIR scheme for general K, N and r, which achieves the following normalized download cost, $\bar{D}(r)$.

Theorem 1 (Outer bound) *In the cache-aided PIR with partially known uncoded prefetching, let $s \in \{1, 2, \cdots, K-1\}$, for the caching ratio r_s, where*

$$r_s = \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}}, \tag{11}$$

the optimal normalized download cost $D^*(r_s)$ is upper bounded by,

$$D^*(r_s) \leq \bar{D}(r_s) = \frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^{i+1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}}. \tag{12}$$

Moreover, if $r_s < r < r_{s+1}$, and $\alpha \in (0,1)$ such that $r = \alpha r_s + (1-\alpha) r_{s+1}$, then

$$D^*(r) \leq \bar{D}(r) = \alpha \bar{D}(r_s) + (1-\alpha)\bar{D}(r_{s+1}). \tag{13}$$

The proof of Theorem 1 is provided in Section IV. Theorem 1 shows that the outer bound is a piece-wise linear curve, which consists of K line segments. These K line segments intersect at the points r_s. We characterize an inner bound (converse bound), which is denoted by $\tilde{D}(r)$, for the optimal normalized download cost $D^*(r)$ for general K, N, r.

Theorem 2 (Inner bound) *In the cache-aided PIR with partially known uncoded prefetching, the normalized download cost is lower bounded as,*

$$\begin{aligned} D^*(r) \geq \tilde{D}(r) &= \max_{i \in \{2,\cdots,K+1\}} \\ &(1-r)\sum_{j=0}^{K+1-i} \frac{1}{N^j} - r\left(1-\frac{1}{N}\right)\sum_{j=0}^{K-i} \frac{K+1-i-j}{N^j} \\ &= \max_{i \in \{2,\cdots,K+1\}} \sum_{j=0}^{K+1-i} \frac{1}{N^j} - (K+2-i)r. \end{aligned} \tag{14}$$

The proof of Theorem 2 is provided in Section V. Theorem 2 shows that the inner bound is a piece-wise linear curve, which consists of K line segments. Interestingly, these K line segments intersect at the points as follows,

$$\tilde{r}_i = \frac{1}{N^{K-i}}, \quad i = 1, \cdots, K-1. \tag{15}$$

The outer bounds provided in Theorem 1 and the inner bounds provided in Theorem 2 match for some caching ratios r summarized as follows. The proof can be found in [31].

Corollary 1 (Exact results for very low and very high r) *In the cache-aided PIR with partially known uncoded prefetching, for very low caching ratios, i.e., for $r \leq \frac{1}{N^{K-1}}$, the optimal normalized download cost is given by,*

$$D^*(r) = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right) - Kr. \tag{16}$$

On the other hand, for very high caching ratios, i.e., for $r \geq \frac{K-2}{N^2-3N+KN}$, the optimal normalized download cost is given by,

$$D^*(r) = \begin{cases} 1 + \frac{1}{N} - 2r, & \frac{K-2}{N^2-3N+KN} \leq r \leq \frac{1}{N} \\ 1 - r, & \frac{1}{N} \leq r \leq 1 \end{cases} \tag{17}$$

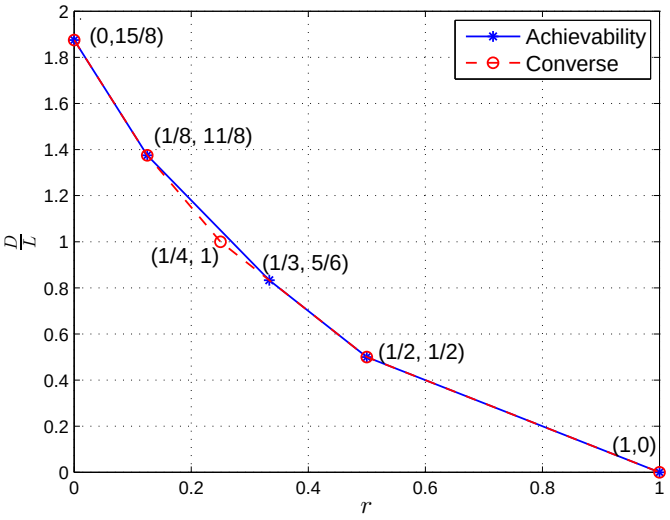

Fig. 1. Inner and outer bounds for $K = 4$, $N = 2$.

We use the example of $K = 4$, $N = 2$ to illustrate Corollary 1 (see Fig. 1). In this case, $r_1 = \tilde{r}_1 = \frac{1}{8}$, $r_{K-2} = \frac{1}{3}$, and $r_{K-1} = \tilde{r}_{K-1} = \frac{1}{2}$. Therefore, we have exact results for $0 \leq r \leq \frac{1}{8}$ (very low caching ratios) and $\frac{1}{3} \leq r \leq 1$ (very high caching ratios). We have a gap between the achievability and the converse for medium caching ratios in $\frac{1}{8} \leq r \leq \frac{1}{3}$. More specifically, line segments connecting $(0, \frac{15}{8})$ and $(\frac{1}{8}, \frac{11}{8})$; connecting $(\frac{1}{3}, \frac{5}{6})$ and $(\frac{1}{2}, \frac{1}{2})$; and connecting $(\frac{1}{2}, \frac{1}{2})$ and $(1, 0)$ are tight. For the case $K = 3$, we have exact tradeoff curve for any N, r as shown in the following corollary. The proof is provided in [31].

Corollary 2 (Exact result for $K = 3$) *In the cache-aided PIR with partially known uncoded prefetching with $K = 3$ messages, the optimal download cost caching ratio tradeoff is given explicitly as,*

$$D^*(r) = \begin{cases} 1 + \frac{1}{N} + \frac{1}{N^2} - 3r, & 0 \leq r \leq \frac{1}{N^2} \\ 1 + \frac{1}{N} - 2r, & \frac{1}{N^2} \leq r \leq \frac{1}{N} \\ 1 - r, & \frac{1}{N} \leq r \leq 1 \end{cases} \tag{18}$$

We find the asymptotic upper bound as $K \to \infty$ for the achievable bounds and conclude the worst-case gap to be $\frac{5}{32}$ in the following corollary. The proof is given in [31].

Corollary 3 (Asymptotics and the worst-case gap) *In the cache-aided PIR with partially known uncoded prefetching, as $K \to \infty$, the outer bound is upper bounded by,*

$$\bar{D}(r) \leq \frac{N}{N-1}(1-r)^2 \tag{19}$$

Hence, the worst-case gap is $\frac{5}{32}$.

Comparing to the asymptotic results of [26], [28], we conclude that $\bar{D}(r)$ is decreased by a factor of $1 - r \leq 1$ over the fully

known model in [26], while it is increased by a factor of $1+\frac{r}{N-1}\geq 1$ over the fully unknown model in [28].

IV. Achievable Scheme

In this section, we show the achievable scheme for the outer bounds provided in Theorem 1. Our achievable scheme is based on [10], [26], [28]. We first provide achievable schemes for the caching ratios r_s in (11) by applying the principles in [10]: 1) database symmetry, 2) message symmetry within each database, and 3) exploiting undesired messages as side information. For an arbitrary caching ratio $r\neq r_s$, we apply the memory-sharing scheme in [26]. We first use the case of $K=3$, $N=2$ to illustrate the main ideas of our achievability scheme.

A. Motivating Example: $K=3$ and $N=2$

We permutate the bits of messages W_1,W_2,W_3 randomly and independently, and use a_i, b_i, and c_i to denote the bits of each permutated message, respectively. We assume that the user wants to retrieve message W_1 privately without loss of generality.

1) Caching Ratio $r_1=\frac{1}{4}$: We choose the message size as 8 bits. In the prefetching phase, for caching ratio $r_1=\frac{1}{4}$, the user caches 2 bits from each message. Therefore, the user caches 1 bit from each database for each message. Therefore, $Z_1=(a_1,b_1,c_1)$ and $Z_2=(a_2,b_2,c_2)$.

In the retrieval phase, for $s=1$, we first mix 1 bit of side information with the desired bit. Therefore, the user queries a_3+b_2 and a_4+c_2 from database 1. Note that database 1 knows that the user has prefetched Z_1. Therefore, the user does not use side information Z_1 to retrieve information from database 1. To keep message symmetry, the user further queries b_3+c_3 from database 1. Similarly, the user queries a_5+b_1, a_6+c_1 and b_4+c_4 from database 2. Then the user exploits the side information b_4+c_4 to query $a_7+b_4+c_4$ from database 1 and the side information b_3+c_3 to query $a_8+b_3+c_3$ from database 2. After this step, no more side information can be used and the message symmetry is attained for each database. Therefore, the PIR scheme ends. The decodability of message W_1 can be shown easily, since the desired bits are either mixed with cached side information or the side information obtained from other databases. Overall, the user downloads 8 bits. Therefore, the normalized download cost is 1. We summarize the queries in Table. I.

TABLE I
Query table for $K=3$, $N=2$, $r_1=\frac{1}{4}$

s	DB1	DB2
$s=1$	a_3+b_2	a_5+b_1
	a_4+c_2	a_6+c_1
	b_3+c_3	b_4+c_4
	$a_7+b_4+c_4$	$a_8+b_3+c_3$
	$Z_1=(a_1,b_1,c_1)$	$Z_2=(a_2,b_2,c_2)$

2) Caching Ratio $r_2=\frac{1}{2}$: We choose the message size as 4 bits. In the prefetching phase, for caching ratio $r_2=\frac{1}{2}$, the user caches 2 bits from each message. Therefore, the user caches 1 bit from each database for each message. Therefore, $Z_1=(a_1,b_1,c_1)$ and $Z_2=(a_2,b_2,c_2)$. In the retrieval phase, for $s=2$, we first mix 2 bits of side information with the desired bit. Therefore, the user queries $a_3+b_2+c_2$ from database 1. Similarly, the user queries $a_4+b_1+c_1$ from database 2. After this, no more side information can be used and the message symmetry is attained for each database. Therefore, the PIR scheme ends. Overall, the user downloads 2 bits. Therefore, the normalized download cost is $\frac{1}{2}$. We summarize the queries in Table. II.

TABLE II
Query table for $K=3$, $N=2$, $r_2=\frac{1}{2}$

s	DB1	DB2
$s=2$	$a_3+b_2+c_2$	$a_4+b_1+c_1$
	$Z_1=(a_1,b_1,c_1)$	$Z_2=(a_2,b_2,c_2)$

3) Caching Ratio $r=\frac{1}{3}$: We choose the message size as 12 bits. In the prefetching phase, for caching ratio $r=\frac{1}{3}$, the user caches 4 bits from each message. Therefore, the user caches 2 bits from each database for each message. Therefore, $Z_1=(a_1,a_2,b_1,b_2,c_1,c_2)$ and $Z_2=(a_3,a_4,b_3,b_4,c_3,c_4)$. In the retrieval phase, we combine the achievable schemes in Section IV-A1 and IV-A2 as shown in Table III. The normalized download cost is $\frac{5}{6}$. By applying [26, Lemma 1] and taking $\alpha=\frac{2}{3}$, we can show that $\bar{D}(\frac{1}{3})=\bar{D}(\frac{2}{3}\cdot\frac{1}{4}+\frac{1}{3}\cdot\frac{1}{2})=\frac{2}{3}\bar{D}(\frac{1}{4})+\frac{1}{3}\bar{D}(\frac{1}{2})=\frac{2}{3}\cdot 1+\frac{1}{3}\cdot\frac{1}{2}=\frac{5}{6}$.

TABLE III
Query table for $K=3$, $N=2$, $r=\frac{1}{3}$

s	DB1	DB2
$s=1$	a_5+b_3	a_7+b_1
	a_6+c_3	a_8+c_1
	b_5+c_5	b_6+c_6
	$a_9+b_6+c_6$	$a_{10}+b_5+c_5$
$s=2$	$a_{11}+b_4+c_4$	$a_{12}+b_2+c_2$
	$Z_1=(a_1,a_2,b_1,b_2,c_1,c_2)$	$Z_2=(a_3,a_4,b_3,b_4,c_3,c_4)$

B. Achievable Scheme

We first present the achievable scheme for the caching ratios r_s given in (11). Then, we apply the memory sharing scheme provided in [26] for the other caching ratios.

Achievable Scheme for the Caching Ratio r_s: For fixed K and N, there are $K-1$ non-degenerate corner points (in addition to degenerate caching ratios $r=0$ and $r=1$). The caching ratios, r_s, corresponding to these non-degenerate corner points are indexed by s, which represents the number of cached bits used in the side information mixture at the first round of the querying. For each $s\in\{1,2,\ldots,K-1\}$, we choose the length of the message to be $L(s)$ for the corner point indexed by s, where

$$L(s)=N\binom{K-2}{s-1}+\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}N. \quad (20)$$

In the prefetching phase, for each message the user randomly and independently chooses $N\binom{K-2}{s-1}$ bits to cache, and

caches $\binom{K-2}{s-1}$ bits from each database for each message. Therefore, the caching ratio r_s is equal to

$$r_s = \frac{N\binom{K-2}{s-1}}{N\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}N}. \quad (21)$$

In the retrieval phase, the PIR scheme is as follows:

1) *Initialization:* Set the round index to $t = s+1$, where the tth round involves downloading sums of every t combinations of the K messages.
2) *Exploiting side information:* If $t = s+1$, for the first database, the user forms queries by mixing s undesired bits cached from the other $N-1$ databases in the prefetching phase to form one side information equation. Each side information equation is added to one bit from the uncached portion of the desired message. Therefore, for the first database, the user downloads $\binom{K-1}{s}(N-1)$ equations in the form of a desired bit added to a mixture of s cached bits from other messages. On the other hand, if $t > s+1$, for the first database, the user exploits the $\binom{K-1}{t-1}(N-1)^{t-s}$ side information equations generated from the remaining $(N-1)$ databases in the $(t-1)$th round.
3) *Symmetry across databases:* The user downloads the same number of equations with the same structure as in step 2 from every database. Consequently, the user decodes $\binom{K-1}{t-1}(N-1)^{t-s}$ desired bits from every database, which are done either using the cached bits as side information if $t = s+1$, or the side information generated in the $(t-1)$th round if $t > s+1$.
4) *Message symmetry:* To satisfy the privacy constraint, the user should download the same amount of bits from other messages. Therefore, the user downloads $\binom{K-1}{t}(N-1)^{t-s}$ undesired equations from each database in the form of sum of t bits from the uncached portion of the undesired messages.
5) *Repeat* steps 2, 3, 4 after setting $t = t+1$ until $t = K$.
6) *Shuffling the order of queries:* By shuffling the order of queries uniformly, all possible queries can be made equally likely regardless of the message index.

Since the desired bits are added to the side information which is either obtained from the cached bits (if $t = s+1$) or from the remaining $(N-1)$ databases in the $(t-1)$th round when $t > s+1$, the user can decode the uncached portion of the desired message by canceling out the side information bits. In addition, for each database, each message is queried equally likely with the same set of equations, which guarantees privacy as in [10]. Therefore, the privacy constraint in (8) and the reliability constraint in (7) are satisfied.

We now calculate the total number of downloaded bits for the caching ratio r_s in (21). For the round $t = s+1$, we exploit s cached bits to form the side information equation. Therefore, each download is a sum of $s+1$ bits. For each database, we utilize the side information cached from other $N-1$ databases. In addition to the message symmetry step enforcing symmetry across K messages, we download $\binom{K}{s+1}(N-1)$ bits from a database. Due to the database symmetry step, in total, we download $\binom{K}{s+1}(N-1)N$ bits. For the round $t = s+i > s+1$, we exploit $s+i-1$ undesired bits downloaded from the $(t-1)$th round to form the side information equation. Due to message symmetry and database symmetry, we download $\binom{K}{s+1+i}(N-1)^{i+1}N$ bits. In sum, the total number of downloaded bits is,

$$D(r_s) = \sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^{i+1}N. \quad (22)$$

By canceling out the undesired side information bits using the cached bits for the round $t = s+1$, we obtain $\binom{K-1}{s}(N-1)N$ desired bits. For the round $t = s+i > s+1$, we decode $\binom{K-1}{s+i}(N-1)^{i+1}N$ desired bits by using the side information obtained in $(t-1)$th round. In sum, we obtain $L(s) - N\binom{K-2}{s-1}$ desired bits. Therefore, the normalized download cost is,

$$\bar{D}(r_s) = \frac{D(r_s)}{L(s)} = \frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^{i+1}N}{N\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}N}. \quad (23)$$

Achievable Scheme for Caching Ratios not Equal to r_s: For caching ratios r which are not exactly equal to (21) for some s, we first find an s such that $r_s < r < r_{s+1}$. We choose $0 < \alpha < 1$ such that $r = \alpha r_s + (1-\alpha)r_{s+1}$. By using the memory sharing scheme in [26, Lemma 1], we achieve the following normalized download cost,

$$\bar{D}(r) = \alpha\bar{D}(r_s) + (1-\alpha)\bar{D}(r_{s+1}). \quad (24)$$

V. Converse Proof

In this section, we derive an inner bound for the cache-aided PIR with partially known uncoded prefetching. We extend the techniques in [10], [28] to our problem. The main difference between this proof and that in [28] is the usage of privacy constraint given in (8).

Lemma 1 (Interference lower bound) *For the cache-aided PIR with partially known uncoded prefetching, the interference from the undesired messages within the answering strings $D(r) - L(1-r)$ is lower bounded by,*

$$D(r) - L(1-r) + o(L) \geq I\left(W_{k:K}; \mathbb{H}, Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z\right) \quad (25)$$

for all $k \in \{2, \ldots, K\}$.

The proof of Lemma 1 is similar to [28, Lemma 1]. In the following lemma, we prove an inductive relation for the mutual information term on the right hand side of (25). The proof of Lemma 2 is provided in [31].

Lemma 2 (Induction lemma) *For all $k \in \{2, \ldots, K\}$, the mutual information term in Lemma 1 can be inductively lower bounded as,*

$$I\left(W_{k:K}; \mathbb{H}, Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z\right) \geq \frac{1}{N} I\left(W_{k+1:K}; \mathbb{H}, Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z\right)$$

$$+\frac{L(1-r)-o(L)}{N}+\frac{1-N}{N}(K-k+1)Lr. \quad (26)$$

Lemma 2 is a generalization of [10, Lemma 6] and [28, Lemma 2], and it reduces to [10, Lemma 6] when $r=0$. Compared to [28, Lemma 2], the lower bound in (26) is increased by $\frac{(K-k+1)Lr}{N}$, since the cached content is partially known by the databases.

Now we are ready to derive the general inner bound for arbitrary K, N, r. For fixed N, K and r, by applying Lemmas 1, 2 successively, we have

$$\begin{aligned}
D(r) &\overset{(25)}{\geq} L(1-r)-o(L) \\
&\quad + I\left(W_{k:K};\mathbb{H},Q_{1:N}^{[k-1]},A_{1:N}^{[k-1]}|W_{1:k-1},Z\right) \quad (27)\\
&\overset{(26)}{\geq} L(1-r)\left[1+\frac{1}{N}\right]+\left(\frac{1}{N}-1\right)(K-k+1)Lr \\
&\quad + \frac{1}{N}I\left(W_{k+1:K};\mathbb{H},Q_{1:N}^{[k]},A_{1:N}^{[k]}|W_{1:k},Z\right)-o(L) \quad (28)\\
&\overset{(26)}{\geq} \cdots \quad (29)\\
&\overset{(26)}{\geq} L(1-r)\sum_{j=0}^{K+1-k}\frac{1}{N^j} \\
&\quad - Lr\left(1-\frac{1}{N}\right)\sum_{j=0}^{K-k}\frac{K+1-k-j}{N^j}+o(L), \quad (30)
\end{aligned}$$

where $k=2,\ldots,K+1$. We conclude the proof by dividing by L and taking the limit as $L\to\infty$, which gives (14).

VI. Conclusion

In this paper, we studied the cache-aided PIR problem from N non-communicating and replicated databases, when the cache stores uncoded bits that are partially known to the databases. We determined inner and outer bounds for the optimal normalized download cost $D^*(r)$ as a function of the total number of messages K, the number of databases N, and the caching ratio r. Both inner and outer bounds are piecewise linear functions in r (for fixed N, K) that consist of K line segments. The bounds match in two specific regimes: the very low caching ratio regime, i.e., $r\leq\frac{1}{N^{K-1}}$, and the very high caching ratio regime, where $r\geq\frac{K-2}{N^2-3N+KN}$. As a direct corollary for this result, we characterized the exact tradeoff between the download cost and the caching ratio for $K=3$. For general K, N, and r, we showed that the largest gap between the achievability and the converse bounds is $\frac{5}{32}$.

References

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

[2] W. Gasarch. A survey on private information retrieval. In *Bulletin of the EATCS*, 2004.

[3] S. Yekhanin. Private information retrieval. *Communications of the ACM*, 53(4):68–73, 2010.

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

[5] G. Fanti and K. Ramchandran. Efficient private information retrieval over unsynchronized databases. *IEEE Journal of Selected Topics in Signal Processing*, 9(7):1229–1239, October 2015.

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

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

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

[9] H. Sun and S. A. Jafar. The capacity of symmetric private information retrieval. In *IEEE Globecom*, December 2016.

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

[11] H. Sun and S. A. Jafar. The capacity of robust private information retrieval with colluding databases. *IEEE Trans. on Info. Theory*, 2017.

[12] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, C. Hollanti, and S. El Rouayheb. Private information retrieval schemes for coded data with arbitrary collusion patterns. 2017. Available at arXiv:1701.07636.

[13] H. Sun and S. Jafar. The capacity of symmetric private information retrieval. 2016. Available at arXiv:1606.08828.

[14] K. Banawan and S. Ulukus. The capacity of private information retrieval from coded databases. *IEEE Trans. on Info. Theory*. Submitted September 2016. Also available at arXiv:1609.08138.

[15] H. Sun and S. A. Jafar. Optimal download cost of private information retrieval for arbitrary message length. *IEEE Trans. on Info. Forensics and Security*, 12(12):2920–2932, December 2017.

[16] H. Sun and S. A. Jafar. Multiround private information retrieval: Capacity and storage overhead. 2016. Available at arXiv:1611.02257.

[17] K. Banawan and S. Ulukus. Multi-message private information retrieval: Capacity results and near-optimal schemes. *IEEE Trans. on Info. Theory*. Submitted February 2017. Also available at arXiv:1702.01739.

[18] K. Banawan and S. Ulukus. The capacity of private information retrieval from Byzantine and colluding databases. *IEEE Trans. on Info. Theory*. Submitted June 2017. Also available at arXiv:1706.01442.

[19] Q. Wang and M. Skoglund. Symmetric private information retrieval for MDS coded distributed storage. 2016. Available at arXiv:1610.04530.

[20] 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, November 2017.

[21] H. Sun and S. A. Jafar. Private information retrieval from MDS coded data with colluding servers: Settling a conjecture by Freij-Hollanti et al. 2017. Available at arXiv: 1701.07807.

[22] Y. Zhang and G. Ge. A general private information retrieval scheme for MDS coded databases with colluding servers. 2017. Available at arXiv: 1704.06785.

[23] Y. Zhang and G. Ge. Multi-file private information retrieval from MDS coded databases with colluding servers. 2017. Available at arXiv: 1705.03186.

[24] Q. Wang and M. Skoglund. Linear symmetric private information retrieval for MDS coded distributed storage with colluding servers. 2017. Available at arXiv:1708.05673.

[25] Q. Wang and M. Skoglund. Secure symmetric private information retrieval from colluding databases with adversaries. 2017. Available at arXiv:1707.02152.

[26] R. Tandon. The capacity of cache aided private information retrieval. 2017. Available at arXiv: 1706.07035.

[27] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson. Private information retrieval with side information. 2017. Available at arXiv:1709.00112.

[28] Y.-P. Wei, K. Banawan, and S. Ulukus. Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching. 2017. Available at arXiv:1709.01056.

[29] Z. Chen, Z. Wang, and S. A. Jafar. The capacity of private information retrieval with private side information. 2017. Available at arXiv:1709.03022.

[30] Y.-P. Wei, K. Banawan, and S. Ulukus. The capacity of private information retrieval with partially known private side information. 2017. Available at arXiv:1710.00809.

[31] Y.-P. Wei, K. Banawan, and S. Ulukus. Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits. 2017. Available at arXiv:1712.07021.