Cache-Aided Private Information Retrieval with Unknown and 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. We assume that the databases are unaware of the cache content. We investigate $D^*(r)$ the optimal download cost normalized with the message size as a function of K, N, r. We develop inner and outer bounds for the optimal download cost. Both inner and outer bounds are piece-wise linear functions in r (for fixed N, K) that consist of K line segments. The inner and the outer bounds match in general for the cases of very low caching ratios and very high caching ratios. 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 additive gap between the achievability and the converse bounds is $\frac{1}{6}$.**

I. Introduction

The problem of private information retrieval (PIR) was introduced by Chor et al. [1] to investigate the privacy of the contents downloaded from public databases, and has become a major research area in the computer science society [2], [3]. In the classical form of the problem [1], a user requests to download a message from K messages from N non-communicating databases such that no database can distinguish individually which message has been retrieved. A naive PIR scheme is to download all of the K messages from a database. However, this trivial PIR scheme is quite inefficient. Consequently, the aim of the PIR problem is to retrieve the desired message correctly by downloading as few bits as possible from the N databases under the privacy constraint.

Recently, the PIR problem has been revisited by information theorists [4]–[9]. In the information-theoretic re-formulation of the problem, the length of the message L is assumed to be arbitrarily large to conform with the traditional Shannon-theoretic arguments, and the upload cost is neglected as it does not scale with the message length. In the influential paper by Sun and Jafar [10], the optimal download cost (total downloaded bits normalized with the message size) is shown to be $\frac{D}{L} = 1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}$. Following the work of [10], the fundamental limits of many interesting variants of the classical PIR problem have been considered [11]–[24].

Recently, reference [25] has considered cache-aided PIR, where the user has local cache memory of size rKL bits and it can store any function of the K messages. With the assumption that the cache content is known by all the N databases, reference [25] characterizes the optimal download cost to be $\frac{D(r)}{L} = (1-r)(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}})$. The result is quite pessimistic as it implies that the cached bits cannot be leveraged as side information and the user must download the uncached portion of the file (the remaining $L(1-r)$ bits) using the original PIR scheme in [10].

The above discussion motivates us to investigate the other extreme where the databases are fully unaware of the cache content. In this case, the user can leverage the cached bits as side information as the databases are unaware of the cached bits. This poses an interesting question: What is the optimal way to exploit the cached bits as side information in order to minimize the normalized download cost? The assumption of unknown prefetching can be interpreted in practice as either the prefetching phase is performed via an external database which does not participate in the retrieval (delivery) phase, or the cache content is dynamically refreshed to keep it essentially random from the perspective of the databases [25]. Concurrent to our paper (see arxiv version [26]), references [27], [28] have also considered the case where the databases are unaware of the cache content; different than our paper, [27], [28] consider the case where full messages are cached.

In this work, we consider PIR with unknown and uncoded prefetching, i.e., the cache content is unknown to the databases, and the cache stores only uncoded portions of all messages. We aim to characterize the optimal normalized download cost $\frac{D(r)}{L}$. For the outer bound, we 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 two explicit points. Therefore, 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], [25] to obtain a piece-wise linear curve consisting of K line segments. We show that the inner and the outer bounds match exactly at three of the K line segments for any number of messages K, which correspond to the very low ($r \leq \frac{1}{1+N+N^2+\cdots+N^{K-1}}$) and the very high ($r \geq \frac{K-2}{(N+1)K+N^2-2N-2}$) caching ratios. For $K = 3$ messages, we fully characterize the optimal download cost. We show that the largest additive gap between the achievability and the converse bounds is $\frac{1}{6}$. Our results show the benefits of the cached content when the databases are unaware of it over the

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

scenario in [25] where the databases are fully aware of it.

II. System Model

We consider a classic PIR problem with K independent messages $W_1, \dots, W_K$. Each message is of size L bits,

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

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

There are N non-communicating databases, and each database stores all the K messages, i.e., the messages are coded via $(N, 1)$ repetition code [14]. The user (retriever) has a local cache memory whose content is denoted by a random variable Z. For each message W_k of size L bits, the user randomly and independently caches Lr bits out of the L bits to Z, where $0 \leq r \leq 1$, and r is called the *caching ratio*. Therefore, $H(Z) = KLr$. Since the user caches a subset of the bits from each message, this is called *uncoded prefetching*. We denote the indices of the cached bits by random variable $\mathbb{H}$. For each message W_k, we have $H(W_k|Z, \mathbb{H}) = L(1-r)$. Here, different from [25], we consider the case where none of the databases knows the prefetched cache content.

After the uncoded prefetching phase, the user privately generates an index $\theta \in [K]$, where $[K] = \{1, \dots, K\}$, and wishes to retrieve message W_θ such that no database knows which message is retrieved. Note that during the prefetching phase, the desired message is unknown a priori. Note further that the cached bit indices $\mathbb{H}$ are independent of the message contents and the desired message index θ. Therefore, we have

$$\begin{aligned} &H\left(\theta, \mathbb{H}, W_1, \dots, W_K\right) \\ &\quad = H\left(\theta\right) + H\left(\mathbb{H}\right) + H(W_1) + \cdots + H(W_K). \end{aligned} \tag{3}$$

Suppose $\theta = k$. The user sends N queries $Q_1^{[k]}, \dots, Q_N^{[k]}$ to the N databases, where $Q_n^{[k]}$ is the query sent to the nth database for message W_k. The queries are generated according to $\mathbb{H}$ and Z, but are independent of the realizations of the uncached messages. Therefore,

$$I(W_1, \dots, W_K; Q_1^{[k]}, \dots, Q_N^{[k]} | Z, \mathbb{H}) = 0. \tag{4}$$

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

$$H(A_n^{[k]} | Q_n^{[k]}, W_1, \dots, W_K) = 0. \tag{5}$$

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

$$(Q_n^{[1]}, A_n^{[1]}, W_1, \dots, W_K) \sim (Q_n^{[k]}, A_n^{[k]}, W_1, \dots, W_K). \tag{6}$$

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

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

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

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 unknown and uncoded prefetching satisfying the privacy constraint (6) 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 $A_{1:N}^{[k]}$, i.e., $D(r) = \sum_{n=1}^{N} H\left(A_n^{[k]}\right)$. In this work, we aim to characterize the optimal normalized download cost $D^*(r)$ corresponding to every caching ratio $0 \leq r \leq 1$, where

$$D^*(r) = \inf \left\{ \frac{D(r)}{L} : (D(r), L) \text{ is achievable} \right\}, \tag{8}$$

which is a function of the caching ratio r.

III. Main Results and Discussions

Our first result characterizes an outer bound (achievable rate) for the normalized download cost $D^*(r)$.

Theorem 1 (Outer bound) *In the cache-aided PIR with uncoded and unknown prefetching, for the caching ratios*

$$r_s = \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N}, \tag{9}$$

where $s \in \{1, 2, \cdots, K-1\}$, the optimal normalized download cost $D^*(r_s)$ is upper bounded by,

$$D^*(r_s) \leq \bar{D}(r_s) = \frac{\sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^i N}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N}. \tag{10}$$

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

$$D^*(r) \leq \bar{D}(r) = \alpha \bar{D}(r_s) + (1-\alpha)\bar{D}(r_{s+1}) \tag{11}$$

The achievability in Theorem 1 is presented in Section IV. Theorem 1 implies that there exist $K+1$ *interesting* caching ratios denoted by r_s, where $s \in \{1, 2, \cdots, K-1\}$ in addition to $r = 0$ and $r = 1$ points. The index s represents the number of cached bits that can be used within one bit of the download (if this downloaded bit uses cached bits as side information). The achievability scheme for any other caching ratio r can be obtained by memory-sharing between the most adjacent interesting caching ratios that include r. Consequently, the outer bound is a piece-wise linear convex curve that connects the $K+1$ interesting caching ratio points. In the following corollary, we compare the outer bound in (10) with the case when the databases have the full knowledge about the cached bits in [25]. The proof can be found in [26].

Corollary 1 (Unawareness gain) *The achievable normalized download cost $\hat{D}(r) = (1-r)\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right)$ in the cache-aided PIR with known prefetching [25] is strictly larger than the achievable normalized download cost $\bar{D}(r)$ in* (10)*, i.e., the databases' unawareness contributes to reducing the download cost beyond the memory-sharing scheme in [25].*

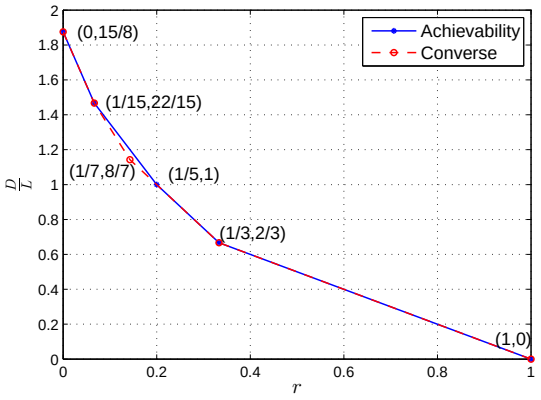

Fig. 1. Inner and outer bounds for $K = 4$, $N = 2$.

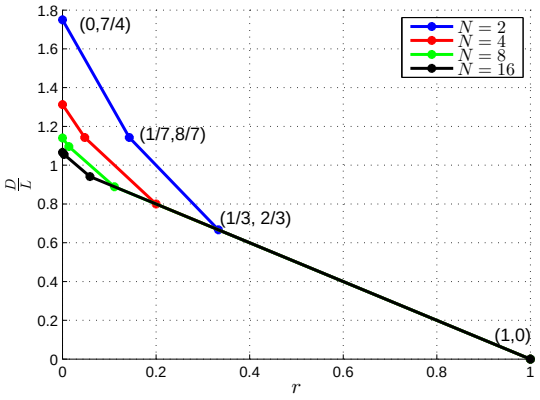

Fig. 2. Optimal download cost caching ratio tradeoff for the case of $K = 3$ messages.

Our second result characterizes an inner bound (converse bound) for the normalized download cost $D^*(r)$.

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

$$D^*(r) \geq \tilde{D}(r) = \max_{i\in\{2,\cdots,K+1\}} (1-r)\sum_{j=0}^{K+1-i}\frac{1}{N^j} - r\sum_{j=0}^{K-i}\frac{K+1-i-j}{N^j}. \quad (12)$$

The proof of Theorem 2 can be found in Section V. Theorem 2 implies that the inner bound is also a piecewise linear curve, which consists of K line segments with decreasing slope as r increases. The points at which the curve changes its slope are given by,

$$\tilde{r}_i = \frac{1}{1+N+N^2+\cdots+N^{K-i}}, \quad i = 1,\cdots,K-1. \quad (13)$$

We note that r_i in (9) and $\tilde{r}_i$ in (13) are the same for $i = 1$ and $i = K-1$. As a consequence of Theorem 1 and Theorem 2, we characterize the optimal download cost caching ratio tradeoff for very low and very high caching ratios in the following corollary. The proof can be found in [26]. As an example, the case of $K = 4$, $N = 2$ is shown in Fig. 1. In this case, $r_1 = \tilde{r}_1 = \frac{1}{15}$, $r_{K-2} = \frac{1}{5}$, and $r_{K-1} = \tilde{r}_{K-1} = \frac{1}{3}$. Therefore, we have exact results for $0 \leq r \leq \frac{1}{15}$ (very low caching ratios) and $\frac{1}{5} \leq r \leq 1$ (very high caching ratios). We have a gap between the achievability and the converse for medium caching ratios in $\frac{1}{15} \leq r \leq \frac{1}{5}$.

Corollary 2 (Exact results for very low and very high r) *In the cache-aided PIR with uncoded and unknown prefetching, for very low caching ratios, i.e., for $r \leq \frac{1}{1+N+N^2+\cdots+N^{K-1}}$, the optimal normalized download cost is given by,*

$$D^*(r) = (1-r)\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}\right) - r\left((K-1)+\frac{K-2}{N}+\cdots+\frac{1}{N^{K-2}}\right). \quad (14)$$

On the other hand, for very high caching ratios, i.e., for $r \geq \frac{K-2}{(N+1)K+N^2-2N-2}$, the optimal normalized download cost is given by,

$$D^*(r) = \begin{cases} (1-r)\left(1+\frac{1}{N}\right) - r, \\ \qquad \frac{K-2}{(N+1)K+N^2-2N-2} \leq r \leq \frac{1}{1+N} \\ 1-r, \quad \frac{1}{1+N} \leq r \leq 1 \end{cases}$$

Finally, we characterize the exact tradeoff curve for any N, r for the special case of $K = 3$ in the following corollary. The proof can be found in [26].

Corollary 3 (Optimal tradeoff for $K = 3$) *In the cache-aided PIR with uncoded and unknown prefetching with $K = 3$ messages, the optimal download cost caching ratio tradeoff is given explicitly as (see Fig. 2),*

$$D^*(r) = \begin{cases} (1-r)\left(1+\frac{1}{N}+\frac{1}{N^2}\right) - r\left(2+\frac{1}{N}\right), \\ \qquad 0 \leq r \leq \frac{1}{1+N+N^2} \\ (1-r)\left(1+\frac{1}{N}\right) - r, \quad \frac{1}{1+N+N^2} \leq r \leq \frac{1}{1+N} \\ 1-r, \quad \frac{1}{1+N} \leq r \leq 1 \end{cases}$$

For general N and K, we show the worst-case additive gap between the inner and outer bounds in the following corollary. The proof can be found in [26].

Corollary 4 (Asymptotics and the worst-case additive gap) *In cache-aided PIR with uncoded and unknown prefetching, as $K \to \infty$, the outer bound is tightly upper bounded by,*

$$\bar{D}(r) \leq \frac{N(1-r)^2}{(N-1)+r} \quad (15)$$

Hence, the worst-case additive gap is $\frac{1}{6}$.

IV. Achievability Scheme

Our achievability scheme is based on the PIR schemes in [10], [25]. Similar to [10], we apply the following three principles recursively: 1) database symmetry, 2) message symmetry

within each database, and 3) exploiting undesired messages as side information. Different from [10], we start the PIR scheme from the third principle due to the availability of pre-existing side information as a result of uncoded prefetching. Here, we present the achievable scheme for the case of $K = 3$, $N = 2$ as a motivating example to illustrate the main ideas of our achievability scheme. For general N, K and r, the achievability scheme can be found in [26].

The optimal download cost caching ratio tradeoff is shown in Fig. 2. We note that there are 4 corner points. Two of them are degenerate, corresponding to $r = 0$, $r = 1$ caching ratios. For $r = 0$, the achievable scheme in [10] achieves $\bar{D}(0) = \frac{7}{4} = \frac{1}{C}$. For $r = 1$, the user has already cached the entire desired file and does not download any extra bits from the databases, i.e., $\bar{D}(1) = 0$. We have two other corner points, corresponding to $r_1 = \frac{1}{7}$, and $r_2 = \frac{1}{3}$. In the sequel, we show the achievability of these two corner points. We also use an example of $r = \frac{1}{5}$ to present the achievability scheme for non-corner points.

1) Caching Ratio $r_1 = \frac{1}{7}$: Let s be the number of cached bits that are mixed together to form side information equation. The first corner point corresponds to $s = 1$. This means that the user exploits every bit in the cache individually as a side information. We next show how $s = 1$ suffices to achieve $r_1 = \frac{1}{7}$, $\bar{D}(\frac{1}{7}) = \frac{8}{7}$ for $K = 3$, $N = 2$; see Fig. 2. We assume that the user wants to retrieve message W_1 privately without loss of generality.

We initialize the process by permuting the indices 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 message, respectively. The steps of the retrieval can be followed in Table I. The user has already cached one bit from each message as denoted by Z in Table I. We start from the third principle by exploiting each bit in the cache as an individual side information. The user downloads $a_2 + b_1$ and $a_3 + c_1$ from the first database (DB1). Then, the user downloads $a_4 + b_1$ and $a_5 + c_1$ from the second database (DB2) to satisfy the database symmetry. Next, the user downloads $b_2 + c_2$ from DB1, and $b_3 + c_3$ from DB2 to ensure the message symmetry within the queries. At this point, all side information corresponding to the cached bits have been exploited. Next, we apply the third principle. The user downloads $a_6 + b_3 + c_3$ from DB1. Finally, we apply the first principle of database symmetry, and the user downloads $a_7 + b_2 + c_2$ from DB2. Since all the undesired side information is used and the symmetry across databases and symmetry within the queries is attained, the iterations stop.

Since the databases do not know the local cache memory Z, and for each database, the user's queries are symmetric across messages, the privacy constraint (6) is satisfied. In addition, the user can decode the desired message. Here, $L = 7$ and the user has cached 1 bit from each message. There are total of 8 downloads. Hence $r = \frac{1}{7}$, and $\bar{D}(\frac{1}{7}) = \frac{8}{7}$.

2) Caching Ratio $r_2 = \frac{1}{3}$: For the second non-degenerate corner point, we have $s = 2$. This means that each 2 bits from the cache are mixed together to form a side information equation. We use the process summarized in the query table in Table II to achieve $r_2 = \frac{1}{3}$, $\bar{D}(\frac{1}{3}) = \frac{2}{3}$ for $K = 3$, $N = 2$.

TABLE I
QUERY TABLE FOR $K = 3$, $N = 2$, $r_1 = \frac{1}{7}$

s	DB1	DB2
$s = 1$	$a_2 + b_1$	$a_4 + b_1$
	$a_3 + c_1$	$a_5 + c_1$
	$b_2 + c_2$	$b_3 + c_3$
	$a_6 + b_3 + c_3$	$a_7 + b_2 + c_2$

$Z = (a_1, b_1, c_1)$

TABLE II
QUERY TABLE FOR $K = 3$, $N = 2$, $r_2 = \frac{1}{3}$

s	DB1	DB2
$s = 2$	$a_2 + b_1 + c_1$	$a_3 + b_1 + c_1$

$Z = (a_1, b_1, c_1)$

3) Caching Ratio $r = \frac{1}{5}$: The achievability scheme for this case is a combination of the achievability schemes in Sections IV-1 and IV-2. Observe that by choosing $L = 10$, the achievable schemes in Sections IV-1 and IV-2 can be concatenated to achieve the caching ratio $r = \frac{1}{5}$. We summarize the process in the query table in Table III.

TABLE III
QUERY TABLE FOR $K = 3$, $N = 2$, $r = \frac{1}{5}$

s	DB1	DB2
$s = 1$	$a_3 + b_1$	$a_5 + b_1$
	$a_4 + c_1$	$a_6 + c_1$
	$b_3 + c_3$	$b_4 + c_4$
	$a_7 + b_4 + c_4$	$a_8 + b_3 + c_3$
$s = 2$	$a_9 + b_2 + c_2$	$a_{10} + b_2 + c_2$

$Z = (a_1, a_2, b_1, b_2, c_1, c_2)$

Here, we have $L = 10$, therefore $r = \frac{1}{5}$, and $\bar{D}(\frac{1}{5}) = \frac{10}{10} = 1$. In fact, by applying [25, Lemma 1] and taking $\alpha = \frac{7}{10}$, we can show that the normalized download cost of this example can be obtained from the download costs obtained in Sections IV-1 and IV-2, as $\bar{D}(\frac{1}{5}) = \bar{D}(\frac{1}{7} \cdot \frac{7}{10} + \frac{1}{3} \cdot \frac{3}{10}) = \frac{7}{10}\bar{D}(\frac{1}{7}) + \frac{3}{10}\bar{D}(\frac{1}{3}) = \frac{7}{10} \cdot \frac{8}{7} + \frac{3}{10} \cdot \frac{2}{3} = 1$.

V. CONVERSE PROOF

In this section, we derive an inner bound for the cache-aided PIR with uncoded and unknown prefetching. We extend the techniques presented in [10], [25] to our problem. We first characterize a lower bound on the length of the undesired portion of the answer strings. The proof can be found in [26].

Lemma 1 (Interference lower bound) *For the cache-aided PIR with unknown and uncoded prefetching, the interference*

from undesired messages within the answer strings $D(r) - L(1-r)$ is lower bounded by,

$$D(r) - L(1-r) + o(L) \geq I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \quad (16)$$

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

If the privacy constraint is absent, the user downloads only $L(1-r)$ bits in order to decode the desired message, however, when the privacy constraint is present, it should download $D(r)$. Lemma 1 provides $K-1$ lower bounds on $D(r) - L(1-r)$ by changing the index k from 2 to K. Each of these $K-1$ bounds contributes a different line segment for the final inner bound. Note that Lemma 1 reduces to [10, Lemma 5] when $k = 2$, $r = 0$. In the following lemma, we prove an inductive relation for the mutual information term on the right hand side of (16). The proof can be found in [26].

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

$$I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \geq \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) + \frac{L(1-r) - o(L)}{N} - (K-k+1)Lr. \quad (17)$$

Note that Lemma 2 reduces to [10, Lemma 6] when caching ratio $r = 0$. For the general inner bound, we use Lemma 1 to find K lower bounds on the length of the undesired portion of the answer strings $D(r) - L(1-r)$. Each lower bound is obtained by varying the index k in the lemma from $k = 2$ to $k = K$. Next, we inductively lower bound each result of Lemma 1 by using Lemma 2, precisely $(K-k+1)$ times, to get K explicit lower bounds. The detailed steps can be found in [26]. For N and K, we have

$$D(r) \geq L(1-r) \sum_{j=0}^{K+1-k} \frac{1}{N^j} - Lr \sum_{j=0}^{K-k} \frac{K+1-k-j}{N^j} - o(L), \quad (18)$$

where $k = 2, \ldots, K+1$. We conclude the converse proof by dividing (18) by L and taking the limit as $L \to \infty$, which gives (12).

VI. Conclusion

We studied the cache-aided PIR problem with unknown and uncoded prefetching. We determined inner and outer bounds for the optimal normalized download cost $D^*(r)$. The bounds match in two specific regimes: $r \leq \frac{1}{1+N+N^2+\cdots+N^{K-1}}$ and $r \geq \frac{K-2}{(N+1)K+N^2-2N-2}$. 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 additive gap between the achievability and the converse bounds is $\frac{1}{6}$.

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 private information retrieval. In *IEEE Globecom*, December 2016.

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

[11] 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.

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

[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 Transactions on Information Theory*, 64(3):1945–1956, March 2018.

[15] K. Banawan and S. Ulukus. Multi-message private information retrieval: Capacity results and near-optimal schemes. *IEEE Transactions on Information Theory*. To appear. Also available at arXiv:1702.01739.

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

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

[18] H. Sun and S. Jafar. Multiround private information retrieval: Capacity and storage overhead. *IEEE Transactions on Information Theory*, 2018.

[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, 2017.

[21] H. Sun and S. Jafar. Private information retrieval from MDS coded data with colluding servers: Settling a conjecture by Freij-Hollanti et al. *IEEE Transactions on Information Theory*, 64(2):1000–1022, February 2018.

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

[26] 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.

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