Fundamental Limits of Cache-Aided Private Information Retrieval With Unknown and Uncoded Prefetching

Yi-Peng Wei, *Student Member, IEEE*, Karim Banawan, *Member, IEEE*, and Sennur Ulukus, *Fellow, IEEE*

***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, and r. For a fixed K and N, we develop an inner bound (converse bound) for the $D^*(r)$ curve. The inner bound is a piece-wise linear function in r that consists of K line segments. For the achievability, we develop explicit schemes that exploit the cached bits as side information to achieve $K-1$ non-degenerate corner points. These corner points differ in the number of cached bits that are used to generate the one-side information equation. We obtain an outer bound (achievability) for any caching ratio by memory sharing between these corner points. Thus, the outer bound is also a piece-wise linear function in r that consists of K line segments. The inner and the outer bounds match in general for the cases of very low-caching ratio and very high-caching ratio. 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 1/6. Our results show that the download cost can be reduced beyond memory sharing if the databases are unaware of the cached content.**

***Index Terms*— Private information retrieval, caching, uncoded prefetching, PIR capacity, side information.**

I. Introduction

THE problem of private information retrieval (PIR) was introduced by Chor et al. [1] as a canonical problem to investigate the privacy of the contents downloaded from public databases. The PIR problem has become a major research area within the computer science literature subsequently, see e.g., [2]–[5]. In the classical form of the problem [1], a user requests to download a message (or a file) from K messages from N non-communicating databases such that no database can distinguish individually which message has been retrieved. The user performs this task by preparing N queries, one for each database, such that the queries do not reveal the user's interest in the desired message. Each database responds truthfully to the received query by an answer string. The user reconstructs the desired message from the collected answer strings. A naive PIR scheme is to download all of the K messages from a database. However, this trivial PIR scheme is quite inefficient from the retrieval rate perspective, which is defined as the number of desired bits per bit of downloaded data. 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 is revisited by information theorists [6]–[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. This formulation provides an absolute privacy guarantee by ensuring statistical independence between the queries and the identity of the desired message. In the influential paper by Sun and Jafar [9], the notion of PIR capacity is introduced, which is the supremum of PIR rates over all achievable retrieval schemes. In [9], the authors characterize the capacity of classical PIR. In [9], a greedy iterative algorithm is proposed for the achievability scheme and an induction based converse is provided to obtain an exact result. The achievable scheme is based on an interesting correspondence between PIR and blind interference alignment [10] as observed earlier in [11]. Sun and Jafar show that in order to privately retrieve a message, the optimal total downloaded bits normalized with the message size is $\frac{D}{L} = 1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}$. Consequently, the PIR capacity is the reciprocal of this optimal normalized download cost, i.e., $C = (1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}})^{-1}$.

Following the work of [9], the fundamental limits of many interesting variants of the classical PIR problem have been considered, such as: PIR with T colluding databases (TPIR) [12], [13], where any T of N databases might collude; robust PIR (RPIR) [12], [14]–[17], where some databases may fail to respond; symmetric PIR (SPIR) [18], which adds the constraint that the user should only learn the desired message; MDS-coded PIR (CPIR) [19], where the contents of the databases are not replicated, but coded via an MDS code;

Manuscript received September 5, 2017; revised March 30, 2018; accepted October 26, 2018. Date of publication November 26, 2018; date of current version April 19, 2019. This work was supported by NSF under Grant CNS 13-14733, Grant CCF 14-22111, Grant CNS 15-26608, and Grant CCF 17-13977. This paper was presented in part at the 2018 IEEE ISIT.

The authors are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: ypwei@umd.edu; kbanawan@umd.edu; ulukus@umd.edu).

Communicated by A. Ramamoorthy, Associate Editor for Coding Techniques.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2018.2883302

multi-message PIR (MPIR) [20], where the user wishes to jointly retrieve P messages; PIR from Byzantine databases (BPIR), where B databases are outdated or worse adversarial [21]; PIR under message size constraint L (LPIR) [22]; multi-round PIR, where the queries are permitted to be a function of the answer strings collected in previous rounds [23]; MDS-coded symmetric PIR [24]; MDS-coded PIR with colluding databases [25]–[27], and its multi-message [17], and symmetric [28] versions.

Recently, [29] 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 subject to this memory size constraint.[1,2] With the assumption that the cache content is known by all the N databases, [29] characterizes the optimal download cost. The achievability scheme is based on memory-sharing[3] and the converse bound is obtained with the aid of Han's inequality. To privately retrieve a message, the optimal total downloaded bits normalized with the message size is $\frac{D(r)}{L} = (1-r)(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}})$. The result is quite pessimistic as it implies that the cached bits cannot be used as side information within the retrieval scheme and the user must download the uncached portion of the file (the remaining $L(1-r)$ bits) using the original PIR scheme in [9]. The reason behind this result is that the databases are fully knowledgeable about the cached bits and can infer which message is desired if the user exploits these cached bits as side information in any form.

The above discussion motivates us to investigate the other extreme where the databases are fully unaware of the cache content, i.e., when the prefetched bits are unknown to all of the N databases (in contrast to having the cache content as public knowledge at all the N databases as in [29]). In this case, the user can leverage the cached bits as side information without sacrificing the privacy constraint 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, and what is the corresponding gain beyond memory-sharing if any? 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 in the context of dynamic cache-aided PIR, in which once the unknown cache is used, the user updates/refreshes its cached contents by some trusted mechanism which keeps the cached content essentially random from the perspective of each database as pointed out by [29]. In this paper, we further assume that the cache content is uncoded. This is a common assumption in the caching literature; see [30], [31], [33].

In this work, we consider PIR with unknown and uncoded prefetching, i.e., we assume that the cache content is unknown to all databases, and the cache supports only direct (uncoded) portions of all messages (smaller subfiles). We aim to characterize 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 proper memory-sharing between the nearest two explicit points. This implies that the outer bound is a piece-wise linear curve which consists of K line segments. For the inner bound, we extend the techniques of [9], [29] 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 of the K line segments for any number of messages K. This means that we characterize the optimal tradeoff for 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. 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 for fixed N, the outer bound monotonically increases as K increases. To characterize the worst-case gap between the inner and the outer bounds, we determine the asymptotic achievability bound as $K \to \infty$ for fixed N, r. We then show that the asymptotic gap monotonically decreases in N. Therefore, the worst-case gap happens at $N=2$ and $K \to \infty$. By maximizing this over r, we show that the largest 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 scenario in [29] where the databases are fully aware of the cached content.

II. System Model

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

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

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

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 [19]. 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,

$$H(Z) = KLr. \tag{3}$$

[1] Caching is an important technique to reduce the peak-time traffic in networks by pre-storing (prefetching) content to end-user's local memory [30], [31], and has been an active recent research field on its own right.

[2] In another related line of work, [32] investigates the privacy risks when the clients of an index coding based broadcast system possess a subset of the messages as side information and use them to retrieve the desired message privately against an external eavesdropper.

[3] Memory-sharing, introduced in [29], is an achievability concept similar to the classical achievability concept of time-sharing. Reference [29, Lemma 1] first shows that the download cost $D(S)$ is a convex function of the cache memory size S. That is, for two different cache sizes S_1 and S_2, we have $D(\alpha S_1 + (1-\alpha)S_2) \leq \alpha D(S_1) + (1-\alpha)D(S_2)$. Reference [29, Lemma 1] shows this by dividing the messages into two independent parts of sizes αL and $(1-\alpha)L$ and correspondingly scaling the cache memory sizes with α and $(1-\alpha)$, and applying two different PIR schemes to the two independent parts of the message. This implies that memory-sharing between zero caching (and requiring the download cost in [9]) and caching all the messages (and requiring zero download cost), a normalized download cost of $(1-r)\left(1+\frac{1}{N}+\cdots, \frac{1}{N^{K-1}}\right)$ is achievable with a caching ratio of r, which is linear in r.

The caching ratio r is known to the databases. 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

$$H(W_k|Z,\mathbb{H}) = L(1-r). \tag{4}$$

Here, different from [29], 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, for random variables θ, $\mathbb{H}$, and $W_1,\dots,W_K$, we have

$$\begin{aligned} &H(\theta,\mathbb{H},W_1,\dots,W_K) \\ &\quad = H(\theta) + H(\mathbb{H}) + H(W_1) + \cdots + H(W_K). \end{aligned} \tag{5}$$

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,

$$I(W_1,\dots,W_K; Q_1^{[k]},\dots,Q_N^{[k]}|Z,\mathbb{H}) = 0. \tag{6}$$

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

$$(Q_n^{[1]}, A_n^{[1]}, W_1,\dots,W_K) \sim (Q_n^{[k]}, A_n^{[k]}, W_1,\dots,W_K). \tag{7}$$

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, $\forall k \in [K], \forall n \in [N]$,

$$H(A_n^{[k]}|Q_n^{[k]}, W_1,\dots,W_K) = 0. \tag{8}$$

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

$$H\left(W_k|Z,\mathbb{H},Q_1^{[k]},\dots,Q_N^{[k]},A_1^{[k]},\dots,A_N^{[k]}\right) = o(L), \tag{9}$$

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 with unknown and uncoded prefetching satisfying the privacy constraint (7) and the reliability constraint (9), 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.,

$$D(r) = \sum_{n=1}^{N} H\left(A_n^{[k]}\right). \tag{10}$$

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

$$D^*(r) = \inf\left\{\frac{D(r)}{L} : (D(r), L) \text{ is achievable}\right\}, \tag{11}$$

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 $D^*(r)$ for general K, N and r.

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

$$r_s = \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i N}, \tag{12}$$

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

$$D^*(r_s) \le \bar{D}(r_s) = \frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i N}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i N}. \tag{13}$$

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

$$D^*(r) \le \bar{D}(r) = \alpha \bar{D}(r_s) + (1-\alpha)\bar{D}(r_{s+1}). \tag{14}$$

The proof of Theorem 1 can be found 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 $r = 0$ point (no caching) and $r = 1$ point (everything cached). The index s, which characterizes r_s for these points, 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). For example, if $s = 2$, this means that the user should use two of the cached bits as side information in the form of mixture of two bits if the caching ratio is r_2. 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 including the $(0, \frac{1}{C})$ point, where C is the PIR capacity without caching found in [9], and $(1,0)$ where everything is cached; here, in (x,y), x denotes the caching ratio and y denotes the normalized download cost.

As a direct corollary for Theorem 1, we note that since the databases do not know the cached bits, the download cost is strictly smaller than the case when the databases have the full knowledge about the cached bits in [29]. We state and prove this in the following corollary. As a concrete example, Fig. 1 shows the gain that can be achieved due to the unawareness of the databases about the cached bits.

Corollary 1 (Unawareness Gain): The achievable normalized download cost $\hat{D}(r)$ in the cache-aided PIR with known prefetching [29]

$$\hat{D}(r) = (1-r)\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right) \tag{15}$$

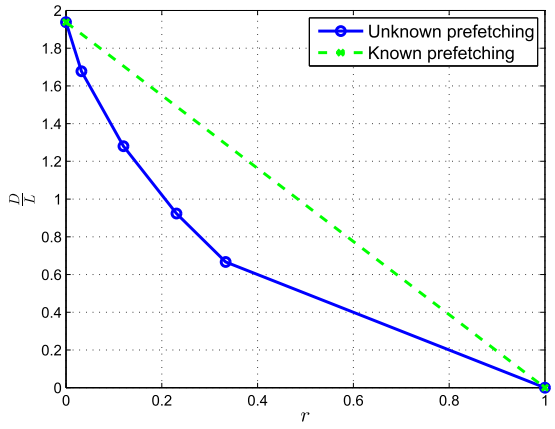

Fig. 1. Comparison between the optimal download cost for known prefetching (15) in [29] and the achievable download cost for unknown prefetching in (13) for $K = 5$ and $N = 2$.

is strictly larger than the achievable normalized download cost $\bar{D}(r)$ in (13) *for $0 < r < 1$, i.e., the databases' unawareness contributes to reducing the download cost beyond the memory-sharing scheme in [29].*

Proof: For $r = 0$, the achievable download cost $\bar{D}(r)$ in (13) is $\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right)$, which is the same as (15). For $r = 1$, the achievable download cost $\bar{D}(r)$ in (13) is 0, which is the same as (15). To show that $\hat{D}(r)$ in (15) is larger than $\bar{D}(r)$ in (13) for $0 < r < 1$, it suffices to show that there exists a caching ratio r such that $\bar{D}(r) < \hat{D}(r)$, since the other caching ratios can be achieved by the memory-sharing scheme. Taking $s = K - 1$ in (12), we have $r_{K-1} = \frac{1}{1+N}$. For $r = \frac{1}{1+N}$, we have $\bar{D}(r) = \frac{N}{1+N}$, and $\hat{D}(r) = \frac{N}{1+N}\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right)$. Therefore, for r_{K-1}, we have $\bar{D}(r) < \hat{D}(r)$, which shows the sub-optimality of $\hat{D}(r)$ in (15) for the case of known prefetching. ∎

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

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

$$\begin{aligned} D^*(r) &\geq \tilde{D}(r) \\ &= \max_{i\in\{2,\cdots,K+1\}} (1-r)\sum_{j=0}^{K+1-i}\frac{1}{N^j} \\ &\quad - r\sum_{j=0}^{K-i}\frac{K+1-i-j}{N^j}, \end{aligned} \tag{16}$$

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,

$$\tilde{r}_i = \frac{1}{1+N+N^2+\cdots+N^{K-i}}, \quad i = 1, \cdots, K-1. \tag{17}$$

We note that r_i in (12) and $\tilde{r}_i$ in (17) 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. Here, by very low caching ratios we mean $0 \leq r \leq r_1 = \tilde{r}_1 = \frac{1}{1+N+N^2+\cdots+N^{K-1}}$, and by very high caching ratios we mean $r_{K-2} = \frac{K-2}{(N+1)K+N^2-2N-2} \leq r \leq 1$. Note that, in the very high caching ratios, we have two segments, one in $r_{K-2} \leq r \leq r_{K-1}$ and the other in $r_{K-1} \leq r \leq 1$. Therefore, in the inner and outer bounds, each composed of K line segments, the first (very low r) and the last two (very high r) segments match giving exact result. This is stated and proved in the next corollary.

Corollary 2 (Optimal Tradeoff for Very Low and Very High Caching Ratios): 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,

$$\begin{aligned} D^*(r) = (1-r)\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right) \\ -r\left(K - 1 + \frac{K-2}{N} + \cdots + \frac{1}{N^{K-2}}\right) \end{aligned} \tag{18}$$

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,

$$D^*(r) = \begin{cases} (1-r)\left(1+\frac{1}{N}\right) - r, & \\ & \frac{K-2}{(N+1)K+N^2-2N-2} \leq r \leq \frac{1}{1+N} \\ 1-r, & \frac{1}{1+N} \leq r \leq 1. \end{cases} \tag{19}$$

Proof: First, from (12) and (17), let us note that

$$r_1 = \tilde{r}_1 = \frac{1}{1+N+N^2+\cdots+N^{K-1}}, \tag{20}$$

$$r_{K-2} = \frac{K-2}{(N+1)K+N^2-2N-2}, \tag{21}$$

$$r_{K-1} = \tilde{r}_{K-1} = \frac{1}{1+N}. \tag{22}$$

Then, we note from (13) that

$$\bar{D}(r_1) = \frac{\sum_{i=0}^{K-2}\binom{K}{2+i}(N-1)^i N}{\binom{K-2}{0} + \sum_{i=0}^{K-2}\binom{K-1}{1+i}(N-1)^i N} \tag{23}$$

$$= \frac{\frac{N}{(N-1)^2}\left[N^K - \sum_{i=0}^{1}\binom{K}{i}(N-1)^i\right]}{\binom{K-2}{0} + \frac{N}{(N-1)^1}\left[N^{K-1} - \sum_{i=0}^{0}\binom{K-1}{i}(N-1)^i\right]} \tag{24}$$

$$= \frac{N\left[N^K - 1 - K(N-1)\right]}{(N-1)^2 + N(N-1)\left[N^{K-1}-1\right]} \tag{25}$$

$$= \frac{N^{K+1} - KN^2 + (K-1)N}{N^{K+1} - N^K - N + 1} \tag{26}$$

Further, we note from (16), by choosing $i = 2$ and using $r = r_1$, that

$$\tilde{D}(r_1) \geq (1-r_1)\sum_{j=0}^{K+1-2}\frac{1}{N^j} - r_1\sum_{j=0}^{K-2}\frac{K-1-j}{N^j} \tag{27}$$

$$= \left(1-\frac{N-1}{N^K-1}\right)\frac{N^K-1}{N^K-N^{K-1}} - \frac{N-1}{N^K-1}\frac{N}{1-N}\left(-K+\frac{N^K-1}{N^K-N^{K-1}}\right) \tag{28}$$

$$= \frac{N^K-N}{N^K-1}\frac{N^K-1}{N^K-N^{K-1}} + \frac{N}{N^K-1}\left(-K+\frac{N^K-1}{N^K-N^{K-1}}\right) \tag{29}$$

$$= \frac{N^K-N}{N^K-N^{K-1}} + N\left(\frac{-K}{N^K-1}+\frac{1}{N^K-N^{K-1}}\right) \tag{30}$$

$$= \frac{N^{K+1}-KN^2+(K-1)N}{N^{K+1}-N^K-N+1} \tag{31}$$

$$= \bar{D}(r_1) \tag{32}$$

Thus, since $\tilde{D}(r_1) \leq \bar{D}(r_1)$ by definition, (32) implies $\tilde{D}(r_1) = \bar{D}(r_1)$.

Similarly, from (13),

$$\bar{D}(r_{K-2}) = \frac{\sum_{i=0}^{1}\binom{K}{K-1+i}(N-1)^i N}{\binom{K-2}{K-3}+\sum_{i=0}^{1}\binom{K-1}{K-2+i}(N-1)^i N} \tag{33}$$

$$= \frac{N^2+(K-1)N}{N^2+(K-2)N+(K-2)}, \tag{34}$$

and from (16) by choosing $i = K$ and using $r = r_{K-2}$,

$$\tilde{D}(r_{K-2}) \geq (1-r_{K-2})\sum_{j=0}^{1}\frac{1}{N^j} - r_{K-2}\sum_{j=0}^{0}\frac{1-j}{N^j} \tag{35}$$

$$= \left(\frac{N^2+(K-2)N}{N^2+(K-2)N+(K-2)}\right)\left(1+\frac{1}{N}\right) - \frac{K-2}{N^2+(K-2)N+(K-2)} \tag{36}$$

$$= \frac{N^2+(K-1)N}{N^2+(K-2)N+(K-2)} \tag{37}$$

$$= \bar{D}(r_{K-2}) \tag{38}$$

implying $\tilde{D}(r_{K-2}) = \bar{D}(r_{K-2})$.

Finally, from (13),

$$\bar{D}(r_{K-1}) = \frac{N}{1+N}, \tag{39}$$

and from (16) by choosing $i = K+1$ and using $r = r_{K-1}$,

$$\tilde{D}(r_{K-1}) \geq \frac{N}{1+N} = \bar{D}(r_{K-1}) \tag{40}$$

implying $\tilde{D}(r_{K-1}) = \bar{D}(r_{K-1})$.

Therefore, $\tilde{D}(r) = \bar{D}(r)$ at $r = r_1$, $r = r_{K-2}$ and $r = r_{K-1}$. We also note that $\tilde{D}(0) = \bar{D}(0)$ and $\tilde{D}(1) = \bar{D}(1)$. Since both $\bar{D}(r)$ and $\tilde{D}(r)$ are linear functions of r, and since $\tilde{D}(0) = \bar{D}(0)$ and $\tilde{D}(r_1) = \bar{D}(r_1)$, we have $\tilde{D}(r) = \bar{D}(r) = D^*(r)$ for $0 \leq r \leq r_1$. This is the very low caching ratio region. In addition, since $\tilde{D}(r_{K-2}) = \bar{D}(r_{K-2})$, $\tilde{D}(r_{K-1}) = \bar{D}(r_{K-1})$ and $\tilde{D}(1) = \bar{D}(1)$, we have $\tilde{D}(r) = \bar{D}(r) = D^*(r)$ for $r_{K-2} \leq r \leq 1$. This is the very high caching ratio region. ■

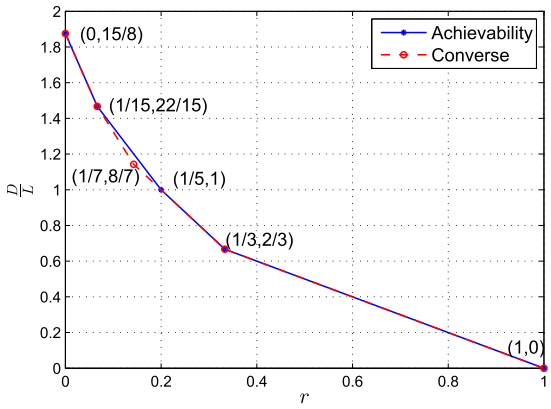

Fig. 2. Inner and outer bounds for $K = 4$ and $N = 2$. For the (x, y) points in this figure, x denotes the caching ratio r and y denotes the normalized download cost $\frac{D}{L}$.

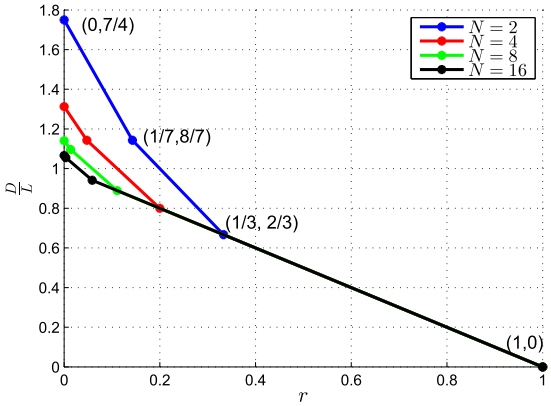

Fig. 3. Optimal download cost caching ratio tradeoff for the case of $K = 3$ messages.

As an example, the case of $K = 4$ and $N = 2$ is shown in Fig. 2. 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 $\frac{1}{15} \leq r \leq \frac{1}{5}$. More specifically, line segments connecting $(0, \frac{15}{8})$ and $(\frac{1}{15}, \frac{22}{15})$; connecting $(\frac{1}{5}, 1)$ and $(\frac{1}{3}, \frac{2}{3})$; and connecting $(\frac{1}{3}, \frac{2}{3})$ and $(1, 0)$ are tight.

Finally, we characterize the exact tradeoff curve for any N, r for the special case of $K = 3$ in the following corollary.

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. 3),

$$D^*(r) = \begin{cases} (1-r)\left(1+\frac{1}{N}+\frac{1}{N^2}\right) - r\left(2+\frac{1}{N}\right), \\ \qquad 0 \leq r \leq \frac{1}{1+N+N^2} \\ (1-r)\left(1+\frac{1}{N}\right) - r, \\ \qquad \frac{1}{1+N+N^2} \leq r \leq \frac{1}{1+N} \\ 1-r, \quad \frac{1}{1+N} \leq r \leq 1 \end{cases} \tag{41}$$

Proof: The proof follows from the proof of Corollary 2. Note that in this case, from (20) and (21), $r_1 = r_{K-2} = \frac{1}{1+N+N^2}$; and from (22), $r_2 = r_{K-1} = \frac{1}{1+N}$. Thus, we have a tight result for $0 \leq r \leq r_1 = \frac{1}{1+N+N^2}$ (very low caching ratios) and a tight result for $r_{K-2} = r_1 = \frac{1}{1+N+N^2} \leq r \leq 1$, i.e., a tight result for all $0 \leq r \leq 1$. We have three segments in this case: $[0, r_1]$, $[r_1, r_2]$ and $[r_2, 1]$ with three different line expressions for the exact result as given in (12)-(13) and written explicitly in (41). ∎

IV. Achievability Proof

Our achievability scheme is based on the PIR schemes in [9] and [29]. Similar to [9], 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 [9], we start the PIR scheme from the third principle due to the availability of pre-existing side information as a result of uncoded prefetching. These cached bits can be exploited right away as side information without compromising the privacy constraint as the databases do not know them. We begin the discussion by presenting the case of $K = 3$ and $N = 2$ as a motivating example to illustrate the main ideas of our achievability scheme.

A. Motivating Example: The Optimal Tradeoff Curve for $K = 3$ Messages and $N = 2$ Databases

In this example, we show the achievability for $K = 3$ and $N = 2$. We know from Corollary 3 that the inner and the outer bounds match for this case. The optimal download cost caching ratio tradeoff is shown in Fig. 3. 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 user has no cached bits and is forced to apply the achievable scheme in [9] that 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 $r_1 = \frac{1}{1+N+N^2} = \frac{K-2}{(N+1)K+N^2-2N-2} = \frac{1}{7}$, and $r_2 = \frac{1}{1+N} = \frac{1}{3}$. In the sequel, we show the achievability of these two 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. Using the notation in [20], we can say that the user starts downloading from round 2 that sums bits from every two messages together. We next show how $s = 1$ suffices to achieve $r_1 = \frac{1}{7}$, $\bar{D}(\frac{1}{7}) = \frac{8}{7}$ for $K = 3$ and $N = 2$; see Fig. 3.

We use a_i, b_i, and c_i to denote the bits of messages W_1, W_2 and W_3, respectively. 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. The steps of the retrieval can be followed in Table I. The user has already cached one bit from each message, i.e., a_1, b_1, c_1 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, we apply the first principle, and the user downloads $a_4 + b_1$ and $a_5 + c_1$ from the second database (DB2) to satisfy the database symmetry. Next, we apply the second principle to ensure the message symmetry within the queries. The user downloads $b_2 + c_2$ from DB1, and $b_3 + c_3$ from DB2. At this point, all side information corresponding to the cached bits have been exploited. Next, we apply the third principle, since undesired message mixes are available in the form of $b_2 + c_2$ and $b_3 + c_3$. 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. Now, the iterations stop, since all the undesired side information is used and the symmetry across databases and symmetry within the queries is attained. We summarize the process in the query table in Table I.

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

s	DB1	DB2
$s = 1$	$a_2 + b_1$	$a_4 + b_1$
	$a_3 + c_1$	$a_5 + c_1$
	$b_2 + c_2$	$b_3 + c_3$
	$a_6 + b_3 + c_3$	$a_7 + b_2 + c_2$
$Z = (a_1, b_1, c_1)$		

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 (7) is satisfied. The decodability can be easily checked as the user can cancel out b_1, c_1 which it has previously cached, and also cancel $b_2 + c_2$ and $b_3 + c_3$ which are previously downloaded, to obtain $a_2, \cdots, a_7$. Since a_1 is already cached, the user has $a_1, \cdots, a_7$. 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 next show how $s = 2$ suffices to achieve $r_2 = \frac{1}{3}$, $\bar{D}(\frac{1}{3}) = \frac{2}{3}$ for $K = 3$ and $N = 2$; see Fig. 3.

Let $[a_1, a_2, a_3]$, $[b_1, b_2, b_3]$, and $[c_1, c_2, c_3]$ denote a random permutation of the 3 bits of messages W_1, W_2 and W_3, respectively. Suppose the user caches a_1, b_1, c_1 in advance and wants to retrieve message W_1 privately. We start from the third principle. The user downloads $a_2 + b_1 + c_1$ from the first database (DB1). Then, we apply the first principle, and the user downloads $a_3 + b_1 + c_1$ from the second database (DB2). Now, the iterations stop, since all the undesired side information is used and the symmetry across databases and messages is attained. We summarize the process in the query table in Table II. In this case $L = 3$, hence $r = \frac{1}{3}$, and the normalized download cost is $\bar{D}(\frac{1}{3}) = \frac{2}{3}$.

3) Caching Ratio $r = \frac{1}{5}$: So far, we have characterized all the corner points by varying $s = 1, 2$ and achieved the points corresponding to caching ratios r_s in addition

TABLE II

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

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

$Z = (a_1, b_1, c_1)$

TABLE III

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

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

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

to the degenerate caching ratios $r = 0$ and $r = 1$; see Fig. 3. An achievable scheme for any other caching ratio can be obtained by memory-sharing between the two nearest corner points. As an example, we next consider the caching ratio $r = \frac{1}{5}$.

The achievability scheme for this case is a combination of the achievability schemes in Sections IV-A1 and IV-A2. Observe that by choosing $L = 10$, the achievable schemes in Sections IV-A1 and IV-A2 can be concatenated to achieve the caching ratio $r = \frac{1}{5}$. In this case, the user caches $a_1, a_2, b_1, b_2, c_1, c_2$ and wants to retrieve message W_1 privately. For cached bits a_1, b_1, c_1, we apply the same process as in Section IV-A1, i.e., we use $s = 1$ and use every cached bit as individual side information equation. For cached bits a_2, b_2, c_2, we apply the same process as in Section IV-A2, and choose $s = 2$, which implies that we use the mixture of two cached bits as a side information equation. We summarize the process in the query table in Table III.

Here, we have $L = 10$, therefore $r = \frac{1}{5}$, and $\bar{D}(\frac{1}{5}) = \frac{10}{10} = 1$. In fact, by applying [29, 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-A1 and IV-A2, as $\bar{D}(\frac{1}{5}) = \bar{D}(\frac{1}{7} \cdot \frac{7}{10} + \frac{1}{3} \cdot \frac{3}{10}) = \frac{7}{10}\bar{D}(\frac{1}{7}) + \frac{3}{10}\bar{D}(\frac{1}{3}) = \frac{7}{10} \cdot \frac{8}{7} + \frac{3}{10} \cdot \frac{2}{3} = 1$.

B. Achievable Scheme for the Corner Points for Arbitrary K, N

For fixed N and K, there are $K - 1$ non-degenerate corner points (in addition to degenerate caching ratios $r = 0$, $r = 1$). The caching ratios corresponding to these non-degenerate corner points are indexed by s, which enumerate the number of cached bits that are involved in the side information mixture. Hence, r_s is given by

$$r_s = \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N}, \tag{42}$$

where $s \in \{1, 2, \ldots, K-1\}$. We choose the length of the message to be $L(s)$ for the corner point indexed by s, where

$$L(s) = \binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N \tag{43}$$

bits per message. The details of the achievable scheme are as follows:

1) *Initialization:* The user permutes each message randomly and independently. The user caches randomly and privately $\binom{K-2}{s-1}$ bits from each message. Set the round index to $i = s + 1$, where the ith round involves downloading sums of every i combinations of the K messages.
2) *Exploiting side information:* If $i = s + 1$, the user mixes s bits from the cache bits to form one side information equation. Each side information equation is added to one bit from the uncached portion of the desired message. Therefore, the user downloads $\binom{K-1}{s}$ equations in the form of a desired bit added to a mixture of s cached bits from other messages. On the other hand, if $i > s + 1$, the user exploits the $\binom{K-1}{i-1}(N-1)^{i-s-1}$ side information equations generated from the remaining $(N-1)$ databases in the $(i-1)$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 downloads $\binom{K-1}{i-1}(N-1)^{i-s-1}$ bits from every database, which are done either using the cached bits as side information if $i = s + 1$, or the side information generated in the $(i-1)$th round if $i > s + 1$.
4) *Message symmetry:* To satisfy the privacy constraint, the user should download equal amount of bits from all other messages. Therefore, the user downloads $\binom{K-1}{i}(N-1)^{i-s-1}$ undesired equations from each database in the form of sum of i bits from the uncached portion of the undesired messages.
5) *Repeat* steps 2, 3, 4 after setting $i = i + 1$ until $i = K$.
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. This guarantees the privacy.

1) Decodability, Privacy, and the Achievable Normalized Download Cost:

a) Decodability: It is clear that the side information in each round is either constructed from the cached bits (if $i = s+1$) or obtained from the remaining $(N-1)$ databases in the $(i-1)$th round. Consequently, the user can cancel out these side information bits in order to decode the uncached portion of the desired message (the remaining $L(1-r)$ bits).

b) Privacy: The randomized mapping of the cached and the uncached portions of the messages and the randomization of the order of queries guarantees privacy as in [9].

c) Normalized Download Cost: We now calculate the total number of downloaded bits for the caching ratio r in (42). First, we exploit s bits of side information. Therefore, each download is a sum of $s + 1$ bits. Since the second principle enforces symmetry across K messages, we download $\binom{K}{s+1}$

bits from a database. Due to the first principle enforcing symmetry across databases, in total, we download $\binom{K}{s+1}N$ bits. Since we utilize s bits of side information of undesired messages for each download, for each undesired message we use $\frac{\binom{K-1}{s}s}{K-1} = \binom{K-2}{s-1}$ bits, which is the amount of bits we cached in advance for each message. Next, each download is a sum of $s+2$ bits since the available side information is in the form of sums of $s+1$ bits. Due to message symmetry and $(N-1)$ available side information from other $(N-1)$ databases, we download $\binom{K}{s+2}(N-1)$ bits from each database. Due to the first principle enforcing symmetry across databases, in total, we download $\binom{K}{s+2}(N-1)N$ bits. Next, each download is the sum of $s+3$ bits since the available side information is in the form of sums of $s+2$ bits. Note that in the previous iteration, each database provides $(N-1)$ sets of side information, and each database exploits the side information from the other $(N-1)$ databases. Therefore, we download $\binom{K}{s+3}(N-1)^2$ bits from each database. Due to the first principle enforcing symmetry across databases, in total, we download $\binom{K}{s+3}(N-1)^2N$ bits. By continuing in this manner, the total number of downloaded bits is,

$$D(r_s) = \sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^i N. \tag{44}$$

Now, we calculate the number of desired bits we have downloaded in this process. At the beginning of the iteration, each download is a sum of $s+1$ bits. If the download includes a desired bit, the other s bits are from the local cache memory. Therefore, we download $\binom{K-1}{s}$ desired bits from each database, and thus we download a total of $\binom{K-1}{s}N$ desired bits. Next, each download is a sum of $s+2$ bits. If the download includes a desired bit, the other $s+1$ bits are from the side information of undesired bits. For each database, there are $(N-1)$ sets of side information obtained from the previous iteration with one set from each database. Therefore, we download $\binom{K-1}{s+1}(N-1)$ bits from each database, and thus we download a total of $\binom{K-1}{s+1}(N-1)N$ desired bits. Next, each download is a sum of $s+3$ bits. If the download includes a desired bit, the other $s+2$ bits are from the side information of undesired bits. For each database, there are $(N-1)^2$ sets of side information obtained from the previous iteration with $(N-1)$ sets from one database. Therefore, we download $\binom{K-1}{s+2}(N-1)^2N$ desired bits from this iteration. In the end, the number of desired bits we downloaded is $L(s) - \binom{K-2}{s-1}$, where $L(s)$ is given in (43). Finally, the normalized download cost is,

$$\bar{D}(r_s) = \frac{D(r_s)}{L(s)} = \frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i N}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i N}. \tag{45}$$

C. Achievable Scheme for Non-Corner Points for Arbitrary K, N

For caching ratios r which are not exactly equal to (42) for some s, we first find an s such that $r_s < r < r_{s+1}$, and combine the achievability schemes of r_s and r_{s+1}. Then, we can write the achievable normalized download cost as a convex combination of $\bar{D}(r_s)$ and $\bar{D}(r_{s+1})$ using [29, Lemma 1] as follows,

$$\bar{D}(r) = \alpha \bar{D}(r_s) + (1-\alpha)\bar{D}(r_{s+1}), \tag{46}$$

where $r = \alpha r_s + (1-\alpha)r_{s+1}$ and r_s is defined in (42), and $\bar{D}(r)$ is given in (45).

V. Converse Proof

In this section, we derive an inner bound for the cache-aided PIR with uncoded and unknown prefetching. The inner bound is tight in general for very high and very low caching ratios, and in particular, the inner bound is tight everywhere for $K = 3$. We extend the techniques presented in [9] and [29] to our problem. We first need the following lemma, which characterizes a lower bound on the length of the undesired portion of the answer strings as a consequence of the privacy constraint.

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,

$$D(r) - L(1-r) + o(L) \geq I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \tag{47}$$

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 $D(r)$. The difference $D(r) - L(1-r)$ corresponds to the undesired portion of the answer strings. Lemma 1 shows that this portion is lower bounded by the mutual information between the answer strings and the messages $W_{k:K}$ after knowing the first $W_{1:k-1}$ messages and the cached bits. 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 is an extension to [9, Lemma 5] if $k = 2$, $r = 0$.

Proof: We start with the right hand side of (47),

$$\begin{aligned} &I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \\ &\quad = I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, W_{k-1} | W_{1:k-2}, Z, \mathbb{H}\right) \\ &\qquad - I\left(W_{k:K}; W_{k-1} | W_{1:k-2}, Z, \mathbb{H}\right) \end{aligned} \tag{48}$$

For the first term on the right hand side of (48), we have

$$\begin{aligned} &I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, W_{k-1} | W_{1:k-2}, Z, \mathbb{H}\right) \\ &= I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-2}, Z, \mathbb{H}\right) \\ &\quad + I\left(W_{k:K}; W_{k-1} | Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, W_{1:k-2}, Z, \mathbb{H}\right) && (49) \\ &\overset{(9)}{=} I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-2}, Z, \mathbb{H}\right) + o(L) && (50) \\ &\overset{(5),(6)}{=} I\left(W_{k:K}; A_{1:N}^{[k-1]} | W_{1:k-2}, Z, \mathbb{H}, Q_{1:N}^{[k-1]}\right) + o(L) && (51) \\ &= H\left(A_{1:N}^{[k-1]} | W_{1:k-2}, Z, \mathbb{H}, Q_{1:N}^{[k-1]}\right) \\ &\quad - H\left(A_{1:N}^{[k-1]} | W_{1:k-2}, Z, \mathbb{H}, Q_{1:N}^{[k-1]}, W_{k:K}\right) + o(L) && (52) \end{aligned}$$

$$\overset{(9)}{=} H\left(A_{1:N}^{[k-1]}|W_{1:k-2}, Z, \mathbb{H}, Q_{1:N}^{[k-1]}\right) - H\left(W_{k-1}, A_{1:N}^{[k-1]}|W_{1:k-2}, Z, \mathbb{H}, Q_{1:N}^{[k-1]}, W_{k:K}\right)+o(L) \tag{53}$$

$$\leq H\left(A_{1:N}^{[k-1]}|W_{1:k-2}, Z, \mathbb{H}, Q_{1:N}^{[k-1]}\right) - H\left(W_{k-1}|W_{1:k-2}, Z, \mathbb{H}, Q_{1:N}^{[k-1]}, W_{k:K}\right) + o(L) \tag{54}$$

$$\overset{(5),(6)}{=} H\left(A_{1:N}^{[k-1]}|W_{1:k-2}, Z, \mathbb{H}, Q_{1:N}^{[k-1]}\right) - H\left(W_{k-1}|Z, \mathbb{H}\right) + o(L) \tag{55}$$

$$= H\left(A_{1:N}^{[k-1]}|W_{1:k-2}, Z, \mathbb{H}, Q_{1:N}^{[k-1]}\right) - L(1-r) + o(L) \tag{56}$$

$$\leq D(r) - L(1-r) + o(L) \tag{57}$$

where (50), (53) follow from the reliability constraint of W_{k-1}, (51) follows from the independence of the queries $Q_{1:N}^{[k-1]}$ and the messages $W_{k:K}$ given Z and $\mathbb{H}$, (54) follows from the chain rule and the non-negativity of the entropy function, (55) is due to the fact that given Z and $\mathbb{H}$, W_{k-1} is statistically independent of $(W_{1:k-2}, W_{k:K}, Q_{1:N}^{[k-1]})$, (56) follows from the uncoded nature of the cache, and (57) follows from conditioning reduces entropy.

For the second term on the right hand side of (48), we have

$$I\left(W_{k:K}; W_{k-1}|W_{1:k-2}, Z, \mathbb{H}\right) = H\left(W_{k-1}|W_{1:k-2}, Z, \mathbb{H}\right) - H\left(W_{k-1}|W_{1:k-2}, W_{k:K}, Z, \mathbb{H}\right) \tag{58}$$

$$= (L - Lr) - (L - Lr) \tag{59}$$

$$= 0 \tag{60}$$

Combining (48), (57), and (60) yields (47). ∎

In the following lemma, we prove an inductive relation for the mutual information term on the right hand side of (47).

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

$$I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}|W_{1:k-1}, Z, \mathbb{H}\right) \geq \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]}|W_{1:k}, Z, \mathbb{H}\right) + \frac{L(1-r) - o(L)}{N} - (K-k+1)Lr. \tag{61}$$

Lemma 2 relates the mutual information between $W_{k:K}$ and the answer strings to the same mutual information term with $W_{k+1:K}$, i.e., it shifts the term by one message. Since the two terms have the same structure, Lemma 2 constructs an inductive relation. We obtain an explicit lower bound for $I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}|W_{1:k-1}, Z, \mathbb{H}\right)$ by applying this lemma $K-k+1$ times, and therefore characterize an explicit lower bound on $D(r) - L(1-r)$. We do this in Lemma 3 by combining Lemma 1 and Lemma 2. Lemma 2 reduces to [9, Lemma 6] if $r = 0$.

Proof: We start with the left hand side of (61),

$$I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}|W_{1:k-1}, Z, \mathbb{H}\right) = I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, Z, \mathbb{H}|W_{1:k-1}\right) - I\left(W_{k:K}; Z, \mathbb{H}|W_{1:k-1}\right) \tag{62}$$

$$= I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}|W_{1:k-1}\right) + I\left(W_{k:K}; Z, \mathbb{H}|W_{1:k-1}, Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}\right) - I\left(W_{k:K}; Z, \mathbb{H}|W_{1:k-1}\right) \tag{63}$$

$$\geq I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}|W_{1:k-1}\right) - I\left(W_{k:K}; Z, \mathbb{H}|W_{1:k-1}\right) \tag{64}$$

where (64) follows from the non-negativity of mutual information.

For the first term in (64), we have

$$NI\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}|W_{1:k-1}\right) \geq \sum_{n=1}^{N} I\left(W_{k:K}; Q_n^{[k-1]}, A_n^{[k-1]}|W_{1:k-1}\right) \tag{65}$$

$$\overset{(7)}{=} \sum_{n=1}^{N} I\left(W_{k:K}; Q_n^{[k]}, A_n^{[k]}|W_{1:k-1}\right) \tag{66}$$

$$\geq \sum_{n=1}^{N} I\left(W_{k:K}; A_n^{[k]}|W_{1:k-1}, Q_n^{[k]}\right) \tag{67}$$

$$\overset{(8)}{=} \sum_{n=1}^{N} H\left(A_n^{[k]}|W_{1:k-1}, Q_n^{[k]}\right) \tag{68}$$

$$\geq \sum_{n=1}^{N} H\left(A_n^{[k]}|W_{1:k-1}, \mathbb{H}, Q_{1:N}^{[k]}, A_{1:n-1}^{[k]}, Z\right) \tag{69}$$

$$\overset{(8)}{=} \sum_{n=1}^{N} I\left(W_{k:K}; A_n^{[k]}|W_{1:k-1}, \mathbb{H}, Q_{1:N}^{[k]}, A_{1:n-1}^{[k]}, Z\right) \tag{70}$$

$$= I\left(W_{k:K}; A_{1:N}^{[k]}|W_{1:k-1}, \mathbb{H}, Q_{1:N}^{[k]}, Z\right) \tag{71}$$

$$\overset{(5),(6)}{=} I\left(W_{k:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]}|W_{1:k-1}, Z, \mathbb{H}\right) \tag{72}$$

$$\overset{(9)}{=} I\left(W_{k:K}; W_k, Q_{1:N}^{[k]}, A_{1:N}^{[k]}|W_{1:k-1}, Z, \mathbb{H}\right) - o(L) \tag{73}$$

$$= I\left(W_{k:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]}|W_{1:k}, Z, \mathbb{H}\right) + I\left(W_{k:K}; W_k|W_{1:k-1}, Z, \mathbb{H}\right) - o(L) \tag{74}$$

$$= I\left(W_{k:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]}|W_{1:k}, Z, \mathbb{H}\right) + L(1-r) - o(L) \tag{75}$$

$$= I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]}|W_{1:k}, Z, \mathbb{H}\right) + L(1-r) - o(L) \tag{76}$$

where (65), (69) follow from the non-negativity of mutual information, (66) follows from the privacy constraint, (67) follows from the chain rule and the non-negativity of the mutual information, (68), (70) follow from the fact that the answer string $A_n^{[k]}$ is a deterministic function of $(Q_n^{[k]}, W_{1:K})$, (71) follows from the chain rule, (72) follows from the statistical independence of $(Q_{1:N}^{[k]}, W_{k:K})$ given $(Z, \mathbb{H})$, (73) is consequence of the decodability of W_k from $(Q_{1:N}^{[k]}, A_{1:N}^{[k]})$, and (75) is due to the uncoded assumption of the cached bits.

For the second term in (64), we have

$$\begin{aligned} &I(W_{k:K}; Z, \mathbb{H}|W_{1:k-1}) \\ &\quad = H(W_{k:K}|W_{1:k-1}) - H(W_{k:K}|W_{1:k-1}, Z, \mathbb{H}) \quad (77) \\ &\quad = (K-k+1)L - (K-k+1)L(1-r) \quad (78) \\ &\quad = (K-k+1)Lr \quad (79) \end{aligned}$$

where (79) follows from the uncoded nature of the cached bits.

Combining (64), (76), and (79) yields (61). ∎

Now we are ready to derive the general inner bound for arbitrary K, N, r. To obtain this 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. This is stated in the following lemma.

Lemma 3: For N and K, we have

$$D(r) \geq L(1-r) \sum_{j=0}^{K+1-k} \frac{1}{N^j} - Lr \sum_{j=0}^{K-k} \frac{K+1-k-j}{N^j} - o(L), \quad (80)$$

where $k = 2, \ldots, K+1$.

Proof: We have

$$\begin{aligned} &D(r) + o(L) \\ &\overset{(47)}{\geq} L(1-r) + I\left(W_{k:K};, Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}|W_{1:k-1}, Z, \mathbb{H}\right) \quad (81) \\ &\overset{(61)}{\geq} L(1-r) + \frac{L(1-r) - o(L)}{N} - (K-k+1)Lr \\ &\quad + \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]}|W_{1:k}, Z, \mathbb{H}\right) \quad (82) \\ &\overset{(61)}{\geq} L(1-r)\left[1 + \frac{1}{N} + \frac{1}{N^2} + o(L)\right] \\ &\quad - Lr\left[(K-k+1) + \frac{(K-k)}{N}\right] \\ &\quad + \frac{1}{N^2} I\left(W_{k+2:K}; Q_{1:N}^{[k+1]}, A_{1:N}^{[k+1]}|W_{1:k+1}, Z, \mathbb{H}\right) \quad (83) \\ &\overset{(61)}{\geq} \cdots \\ &\overset{(61)}{\geq} L(1-r) \sum_{j=0}^{K+1-k} \frac{1}{N^j} - Lr \sum_{j=0}^{K-k} \frac{K+1-k-j}{N^j} + o(L), \quad (84) \end{aligned}$$

where (81) follows from Lemma 1 starting from general index k, and the remaining bounding steps correspond to successive application of Lemma 2. ∎

We conclude the converse proof by dividing by L and taking the limit as $L \to \infty$, then for $k = 2, \cdots, K+1$, we have

$$D^*(r) \geq (1-r) \sum_{j=0}^{K+1-k} \frac{1}{N^j} - r \sum_{j=0}^{K-k} \frac{K+1-k-j}{N^j} \quad (85)$$

TABLE IV

QUERY TABLE FOR $K = 4$, $N = 2$ AND $r_1 = \frac{1}{15}$

s	DB1	DB2
$s = 1$	$a_2 + b_1$	$a_5 + b_1$
	$a_3 + c_1$	$a_6 + c_1$
	$a_4 + d_1$	$a_7 + d_1$
	$b_2 + c_2$	$b_4 + c_4$
	$b_3 + d_2$	$b_5 + d_4$
	$c_3 + d_3$	$c_5 + d_5$
	$a_8 + b_4 + c_4$	$a_{11} + b_2 + c_2$
	$a_9 + b_5 + d_4$	$a_{12} + b_3 + d_2$
	$a_{10} + c_5 + d_5$	$a_{13} + c_3 + d_3$
	$b_6 + c_6 + d_6$	$b_7 + c_7 + d_7$
	$a_{14} + b_7 + c_7 + d_7$	$a_{15} + b_6 + c_6 + d_6$

$Z = (a_1, b_1, c_1, d_1)$

TABLE V

QUERY TABLE FOR $K = 4$, $N = 2$ AND $r_2 = \frac{1}{5}$

s	DB1	DB2
$s = 2$	$a_3 + b_1 + c_1$	$a_6 + b_1 + c_1$
	$a_4 + d_1 + b_2$	$a_7 + d_1 + b_2$
	$a_5 + c_2 + d_2$	$a_8 + c_2 + d_2$
	$b_3 + c_3 + d_3$	$b_4 + c_4 + d_4$
	$a_9 + b_4 + c_4 + d_4$	$a_{10} + b_3 + c_3 + d_3$

$Z = (a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2)$

TABLE VI

QUERY TABLE FOR $K = 4$, $N = 2$ AND $r_3 = \frac{1}{3}$

s	DB1	DB2
$s = 3$	$a_2 + b_1 + c_1 + d_1$	$a_3 + b_1 + c_1 + d_1$

$Z = (a_1, b_1, c_1, d_1)$

Finally, (85) gives K intersecting line segments, therefore, the normalized download cost is lower bounded by their maximum value

$$D^*(r) \geq \max_{i \in \{2, \cdots, K+1\}} (1-r) \sum_{j=0}^{K+1-i} \frac{1}{N^j} - r \sum_{j=0}^{K-i} \frac{K+1-i-j}{N^j}. \quad (86)$$

VI. FURTHER EXAMPLES

A. $K = 4$ Messages, $N = 2$ Databases

For $K = 4$ and $N = 2$, we show the achievable PIR schemes for caching ratios $r_1 = \frac{1}{15}$ in Table IV, $r_2 = \frac{1}{5}$ in Table V, and $r_3 = \frac{1}{3}$ in Table VI. The achievable normalized download costs for these caching ratios are $\frac{22}{15}$, 1 and $\frac{2}{3}$, respectively. We show the normalized download cost and caching ratio trade off curve in Fig. 2.

B. $K = 4$ Messages, $N = 3$ Databases

For $K = 4$ and $N = 3$, we show the achievable PIR schemes for caching ratios $r_1 = \frac{1}{40}$ in Table VII, $r_2 = \frac{2}{17}$ in Table VIII, and $r_3 = \frac{1}{4}$ in Table IX. We show the normalized download cost and caching ratio trade off in Fig. 4. The achievable normalized download costs for these caching ratios are $\frac{27}{20}$,

TABLE VII

QUERY TABLE FOR $K = 4$, $N = 3$ AND $r_1 = \frac{1}{40}$

s	DB1	DB2	DB3
$s = 1$	$a_2 + b_1$	$a_5 + b_1$	$a_8 + b_1$
	$a_3 + c_1$	$a_6 + c_1$	$a_9 + c_1$
	$a_4 + d_1$	$a_7 + d_1$	$a_{10} + d_1$
	$b_2 + c_2$	$b_4 + c_4$	$b_6 + c_6$
	$b_3 + d_2$	$b_5 + d_4$	$b_7 + d_6$
	$c_3 + d_3$	$c_5 + d_5$	$c_7 + d_7$
	$a_{11} + b_4 + d_4$	$a_{17} + b_2 + c_2$	$a_{23} + b_2 + c_2$
	$a_{12} + b_5 + d_4$	$a_{18} + b_3 + d_2$	$a_{24} + b_3 + d_2$
	$a_{13} + c_5 + d_5$	$a_{19} + c_3 + d_3$	$a_{25} + c_3 + d_3$
	$a_{14} + b_6 + c_6$	$a_{20} + b_6 + c_6$	$a_{26} + b_4 + c_4$
	$a_{15} + b_7 + d_6$	$a_{21} + b_7 + d_6$	$a_{27} + b_5 + d_4$
	$a_{16} + c_7 + d_7$	$a_{22} + c_7 + d_7$	$a_{28} + c_5 + d_5$
	$b_8 + c_8 + d_8$	$b_{10} + c_{10} + d_{10}$	$b_{12} + c_{12} + d_{12}$
	$b_9 + c_9 + d_9$	$b_{11} + c_{11} + d_{11}$	$b_{13} + c_{13} + d_{13}$
	$a_{29} + b_{10} + c_{10} + d_{10}$	$a_{33} + b_8 + c_8 + d_8$	$a_{37} + b_8 + c_8 + d_8$
	$a_{30} + b_{11} + c_{11} + d_{11}$	$a_{34} + b_9 + c_9 + d_9$	$a_{38} + b_9 + c_9 + d_9$
	$a_{31} + b_{12} + c_{12} + d_{12}$	$a_{35} + b_{12} + c_{12} + d_{12}$	$a_{39} + b_{10} + c_{10} + d_{10}$
	$a_{32} + b_{13} + c_{13} + d_{13}$	$a_{36} + b_{13} + c_{13} + d_{13}$	$a_{40} + b_{11} + c_{11} + d_{11}$

$$Z = (a_1, b_1, c_1, d_1)$$

TABLE VIII

QUERY TABLE FOR $K = 4$, $N = 3$ AND $r_2 = \frac{2}{17}$

s	DB1	DB2	DB3
$s = 2$	$a_3 + b_1 + c_1$	$a_6 + b_1 + c_1$	$a_9 + b_1 + c_1$
	$a_4 + d_1 + b_2$	$a_7 + d_1 + b_2$	$a_{10} + d_1 + b_2$
	$a_5 + c_2 + d_2$	$a_8 + c_2 + d_2$	$a_{11} + c_2 + d_2$
	$b_3 + c_3 + d_3$	$b_4 + c_4 + d_4$	$b_5 + c_5 + d_5$
	$a_{12} + b_4 + c_4 + d_4$	$a_{14} + b_3 + c_3 + d_3$	$a_{16} + b_3 + c_3 + d_3$
	$a_{13} + b_5 + c_5 + d_5$	$a_{15} + b_5 + c_5 + d_5$	$a_{17} + b_4 + c_4 + d_4$

$$Z = (a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2)$$

TABLE IX

QUERY TABLE FOR $K = 4$, $N = 3$ AND $r_3 = \frac{1}{4}$

s	DB1	DB2	DB3
$s = 3$	$a_2 + b_1 + c_1 + d_1$	$a_3 + b_1 + c_1 + d_1$	$a_4 + b_1 + c_1 + d_1$

$$Z = (a_1, b_1, c_1, d_1)$$

$\frac{18}{17}$ and $\frac{3}{4}$, respectively. By comparing Fig. 4 with Fig. 2, we observe that, for fixed K, as N grows, the gap between the achievable bound and the converse bound shrinks. This observation will be specified in Section VII.

C. $K = 5$, $K = 10$ and $K = 100$ Messages, $N = 2$ Databases

For $N = 2$, we show the numerical results for the inner and outer bounds for $K = 5$, $K = 10$ and $K = 100$ in Figs. 5, 6 and 7. For fixed N as K grows, the gap between the achievable bound and converse bound increases. This observation will be elaborated in Section VII.

D. $K = 5$, $K = 10$ and $K = 100$ Messages, $N = 3$ Databases

For $N = 3$, we show the numerical results for the inner and outer bounds for $K = 5$, $K = 10$ and $K = 100$ in Figs. 8, 9 and 10. For fixed N as K grows, the gap between the achievable bound and converse bound increases. This observation will be further clarified in Section VII.

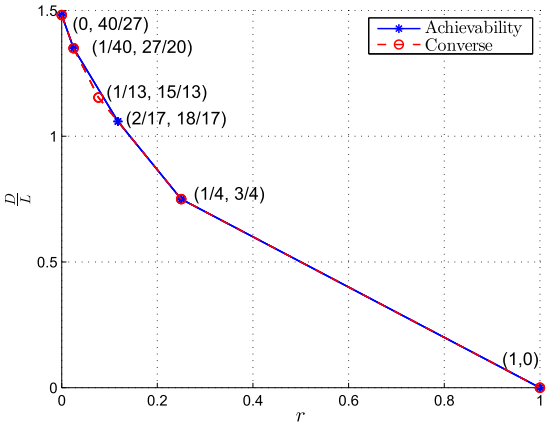

Fig. 4. Inner and outer bounds for $K = 4$ and $N = 3$.

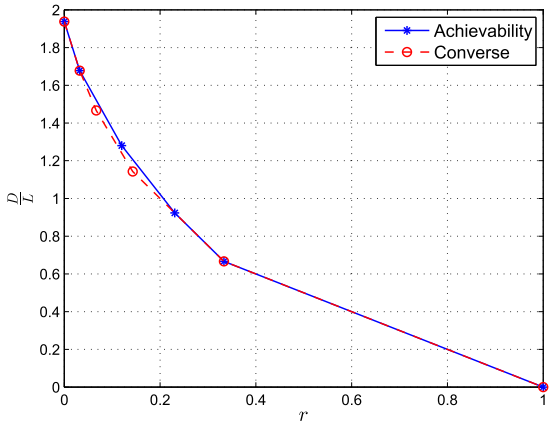

Fig. 5. Inner and outer bounds for $K = 5$ and $N = 2$.

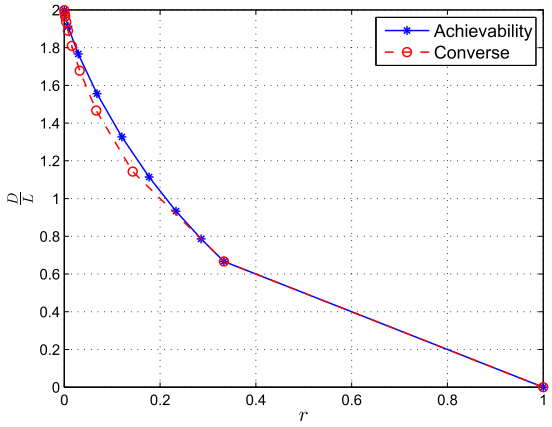

Fig. 6. Inner and outer bounds for $K = 10$ and $N = 2$.

VII. Gap Analysis

In this section, we analyze the gap between the achievability and converse bounds for general N, K, and r, and show that the worst-case gap, which happens when $N = 2$ and $K \to \infty$, is at most $\frac{1}{6}$. We start this section with an interesting property for the monotonicity of the achievable bounds. We first see an example. For $N = 2$, $K = 4$, $K = 5$ and $K = 6$, the achievable bounds are shown in Fig. 11.

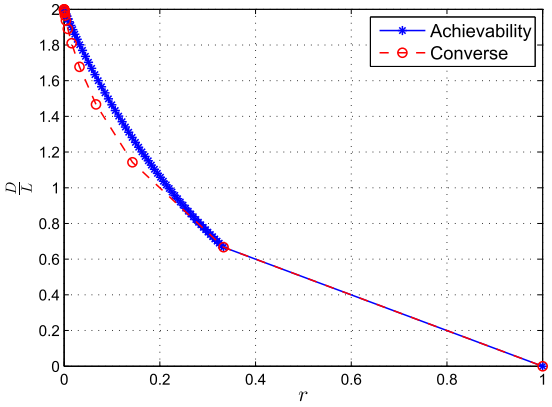

Fig. 7. Inner and outer bounds for $K = 100$ and $N = 2$.

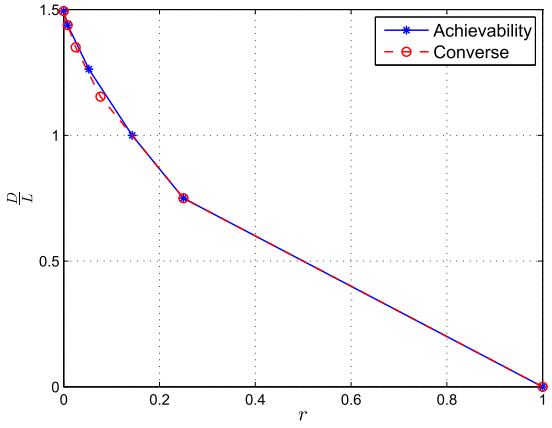

Fig. 8. Inner and outer bounds for $K = 5$ and $N = 3$.

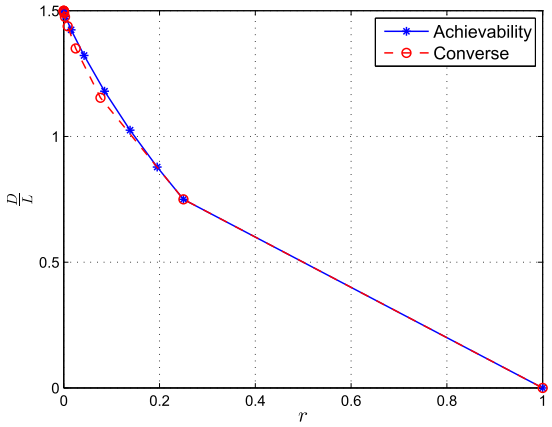

Fig. 9. Inner and outer bounds for $K = 10$ and $N = 3$.

The achievable bound for $K = 6$ is above the achievable bound for $K = 5$, and the achievable bound for $K = 5$ is above the achievable bound for $K = 4$. By denoting $r_s^{(K)}$ as the caching ratio with total K messages and parameter s (see (12)), we observe that $(r_1^{(5)}, \bar{D}(r_1^{(5)}))$ falls on the line connecting $(r_0^{(4)}, \bar{D}(r_0^{(4)}))$ and $(r_1^{(4)}, \bar{D}(r_1^{(4)}))$. This observation is general, $(r_s^{(K+1)}, \bar{D}(r_s^{(K+1)}))$ falls on the line connecting $(r_{s-1}^{(K)}, \bar{D}(r_{s-1}^{(K)}))$ and $(r_s^{(K)}, \bar{D}(r_s^{(K)}))$. We state and prove this observation in the following lemma.

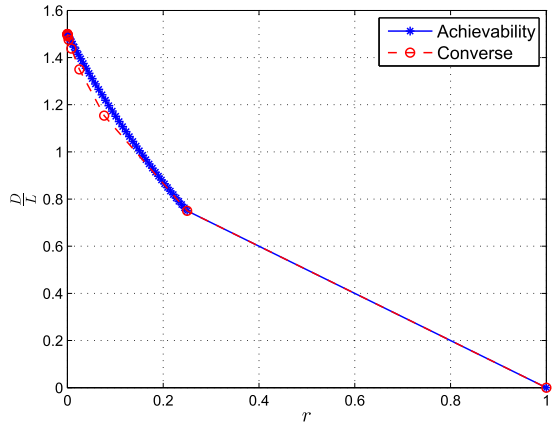

Fig. 10. Inner and outer bounds for $K = 100$ and $N = 3$.

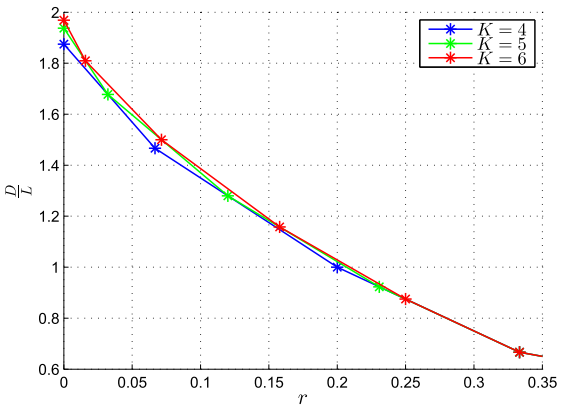

Fig. 11. Outer bounds for $N = 2$, $K = 4$, $K = 5$ and $K = 6$.

Lemma 4 (Monotonicity of the Achievable Bounds): In cache-aided PIR with uncoded and unknown prefetching, for fixed number of databases N, if the number of messages K increases, then the achievable normalized download cost increases. Furthermore, we have

$$r_s^{(K+1)} = \alpha r_{s-1}^{(K)} + (1-\alpha) r_s^{(K)}, \tag{87}$$

$$\bar{D}(r_s^{(K+1)}) = \alpha \bar{D}(r_{s-1}^{(K)}) + (1-\alpha)\bar{D}(r_s^{(K)}), \tag{88}$$

where $0 \le \alpha \le 1$.

Proof: To show (88) is equivalent to show

$$\bar{D}(r_s^{(K+1)}) - \bar{D}(r_s^{(K)}) = \alpha \left(\bar{D}(r_{s-1}^{(K)}) - \bar{D}(r_s^{(K)}) \right), \tag{89}$$

where $\bar{D}\left(r_{s-1}^{(K)}\right) > \bar{D}\left(r_s^{(K)}\right)$. From (87), we have

$$\alpha = \frac{r_s^{(K)} - r_s^{(K+1)}}{r_s^{(K)} - r_{s-1}^{(K)}}. \tag{90}$$

Therefore, to show (89) is equivalent to show

$$\left(r_s^{(K)} - r_{s-1}^{(K)}\right)\left(\bar{D}(r_s^{(K+1)}) - \bar{D}(r_s^{(K)})\right) = \left(r_s^{(K)} - r_s^{(K+1)}\right)\left(\bar{D}(r_{s-1}^{(K)}) - \bar{D}(r_s^{(K)})\right). \tag{91}$$

Let $\bar{D}(r_s^{(K)}) = \frac{D_s^{(K)}}{L_s^{(K)}}$, where

$$L_s^{(K)} = \binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N, \tag{92}$$

$$D_s^{(K)} = \sum_{i=0}^{K-1-s} \binom{K}{s+1+i}(N-1)^i N. \tag{93}$$

To show (91) is equivalent to show

$$\left[\frac{\binom{K-2}{s-1}}{L_s^{(K)}} - \frac{\binom{K-2}{s-2}}{L_{s-1}^{(K)}}\right]\left[\frac{D_s^{(K+1)}}{L_s^{(K+1)}} - \frac{D_s^{(K)}}{L_s^{(K)}}\right] = \left[\frac{\binom{K-2}{s-1}}{L_s^{(K)}} - \frac{\binom{K-1}{s-1}}{L_s^{(K+1)}}\right]\left[\frac{D_{s-1}^{(K)}}{L_{s-1}^{(K)}} - \frac{D_s^{(K)}}{L_s^{(K)}}\right], \tag{94}$$

which is obtained by using (12), (13), (92) and (93). Expanding (94), we have

$$\frac{\binom{K-2}{s-1}}{L_s^{(K)}}\frac{D_s^{(K+1)}}{L_s^{(K+1)}} - \frac{\binom{K-2}{s-2}}{L_{s-1}^{(K)}}\frac{D_s^{(K+1)}}{L_s^{(K+1)}} + \frac{\binom{K-2}{s-2}}{L_{s-1}^{(K)}}\frac{D_s^{(K)}}{L_s^{(K)}} = \frac{\binom{K-2}{s-1}}{L_s^{(K)}}\frac{D_{s-1}^{(K)}}{L_{s-1}^{(K)}} - \frac{\binom{K-1}{s-1}}{L_s^{(K+1)}}\frac{D_{s-1}^{(K)}}{L_{s-1}^{(K)}} + \frac{\binom{K-1}{s-1}}{L_s^{(K+1)}}\frac{D_s^{(K)}}{L_s^{(K)}}. \tag{95}$$

Multiplying $L_s^{(K)} L_{s-1}^{(K)} L_s^{(K+1)}$ to both side of (95), we have

$$\begin{aligned}
&\binom{K-2}{s-1} D_s^{(K+1)} L_{s-1}^{(K)} + \binom{K-1}{s-1} D_{s-1}^{(K)} L_s^{(K)} + \binom{K-2}{s-2} D_s^{(K)} L_s^{(K+1)} \\
&= \binom{K-2}{s-1} D_{s-1}^{(K)} L_s^{(K+1)} + \binom{K-2}{s-2} D_s^{(K+1)} L_s^{(K)} + \binom{K-1}{s-1} D_s^{(K)} L_{s-1}^{(K)}.
\end{aligned} \tag{96}$$

By using (92) and (93), we further have

$$\begin{aligned}
&\binom{K-2}{s-1}\left[\sum_{i=0}^{K-s}\binom{K+1}{s+1+i}(N-1)^i N\right] \times \left[\binom{K-2}{s-2} + \sum_{i=0}^{K-s}\binom{K-1}{s-1+i}(N-1)^i N\right] \\
&+ \binom{K-1}{s-1}\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i N\right] \times \left[\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i N\right] \\
&+ \binom{K-2}{s-2}\left[\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i N\right] \times \left[\binom{K-1}{s-1} + \sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i N\right] \\
&= \binom{K-2}{s-1}\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i N\right]
\end{aligned}$$

$$\times \left[\binom{K-1}{s-1}+\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i N\right]$$
$$+\binom{K-2}{s-2}\left[\sum_{i=0}^{K-s}\binom{K+1}{s+1+i}(N-1)^i N\right]$$
$$\times \left[\binom{K-2}{s-1}+\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i N\right]$$
$$+\binom{K-1}{s-1}\left[\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i N\right]$$
$$\times \left[\binom{K-2}{s-2}+\sum_{i=0}^{K-s}\binom{K-1}{s-1+i}(N-1)^i N\right]. \quad (97)$$

By canceling same terms on both sides, we have

$$\binom{K-2}{s-1}\left[\sum_{i=0}^{K-s}\binom{K+1}{s+1+i}(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-s}\binom{K-1}{s-1+i}(N-1)^i\right]$$
$$+\binom{K-1}{s-1}\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i\right]$$
$$+\binom{K-2}{s-2}\left[\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right]$$
$$=\binom{K-2}{s-1}\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right]$$
$$+\binom{K-2}{s-2}\left[\sum_{i=0}^{K-s}\binom{K+1}{s+1+i}(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i\right]$$
$$+\binom{K-1}{s-1}\left[\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-s}\binom{K-1}{s-1+i}(N-1)^i\right]. \quad (98)$$

By using the fact that $\binom{K}{s}=\binom{K-1}{s}+\binom{K-1}{s-1}$, we have

$$\binom{K-2}{s-1}\left[\sum_{i=0}^{K-s}\left(\binom{K}{s+1+i}+\binom{K}{s+i}\right)(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-s}\binom{K-1}{s-1+i}(N-1)^i\right]$$
$$+\left(\binom{K-2}{s-1}+\binom{K-2}{s-2}\right)\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i\right]$$
$$+\binom{K-2}{s-2}\left[\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-s}\left(\binom{K-1}{s+i}+\binom{K-1}{s+i-1}\right)(N-1)^i\right]$$
$$=\binom{K-2}{s-1}\left[\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-s}\left(\binom{K-1}{s+i}+\binom{K-1}{s+i-1}\right)(N-1)^i\right]$$
$$+\binom{K-2}{s-2}\left[\sum_{i=0}^{K-s}\left(\binom{K}{s+1+i}+\binom{K}{s+i}\right)(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i\right]$$
$$+\left(\binom{K-2}{s-1}+\binom{K-2}{s-2}\right)$$
$$\times\left[\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i\right]$$
$$\times\left[\sum_{i=0}^{K-s}\binom{K-1}{s-1+i}(N-1)^i\right]. \quad (99)$$

Since the left hand side of (99) is equal to the right hand side of (99), (88) holds.

To show $\alpha \geq 0$, since $r_s^{(K)} > r_{s-1}^{(K)}$ in (90), it suffices to show that $r_s^{(K)} \geq r_s^{(K+1)}$. From (12), it is equivalent to show that

$$\frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1}+\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i N} \geq \frac{\binom{K-1}{s-1}}{\binom{K-1}{s-1}+\sum_{i=0}^{K-s}\binom{K}{s+i}(N-1)^i N}. \quad (100)$$

By using the fact that $\binom{K}{s}=\binom{K-1}{s}+\binom{K-1}{s-1}$, we have

$$\frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1}+\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i N} \geq \frac{\binom{K-2}{s-1}+\binom{K-2}{s-2}}{\binom{K-2}{s-1}+\binom{K-2}{s-2}+\sum_{i=0}^{K-s}\left[\binom{K-1}{s+i}+\binom{K-1}{s+i-1}\right](N-1)^i N} \quad (101)$$

which is equivalent to

$$\frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1}+\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i N} \geq \frac{\binom{K-2}{s-2}}{\binom{K-2}{s-2}+\sum_{i=0}^{K-s}\binom{K-1}{s+i-1}(N-1)^i N}. \quad (102)$$

By using (12), (102) is equivalent to

$$r_s^{(K)} \geq r_{s-1}^{(K)}. \quad (103)$$

Since (103) holds, we have $\alpha \geq 0$. Furthermore, $\alpha \leq 1$ can be proved similarly. For fixed N, since $\bar{D}(r_0^{(K+1)}) > \bar{D}(r_0^{(K)})$, the achievable normalized download cost monotonically increases. ∎

The following lemma provides an asymptotic upper bound for the achievable normalized download cost as a smooth function in (r, N). From this expression, we characterize the worst-case gap between the outer and the inner bounds to be $\frac{1}{6}$.

Lemma 5 (Asymptotics and the Worst-Case Gap): In cache-aided PIR with uncoded and unknown prefetching, as $K \to \infty$, the outer bound is tightly upper bounded by,

$$\bar{D}(r) \leq \frac{N(1-r)^2}{(N-1)+r} \quad (104)$$

Hence, the worst-case gap is $\frac{1}{6}$. The asymptotic unawareness multiplicative gain over memory-sharing in [29] is $\frac{1-r}{1+\frac{r}{N-1}} \leq 1$.

Proof: We write the outer bound $\bar{D}(r_s)$ as

$$\bar{D}(r_s) = \frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i N}{\binom{K-2}{s-1}+\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i N} \quad (105)$$

$$= \frac{\frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i}{\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i}}{\frac{\binom{K-2}{s-1}}{\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i N}+1} \quad (106)$$

$$= \frac{\psi_1(N,K,s)}{\psi_2(N,K,s)+1}. \quad (107)$$

Denote $\lambda = \frac{s}{K}$. To upper bound $\psi_1(N,K,s)$,

$$\psi_1(N,K,s) = \frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^i}{\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i} \quad (108)$$

$$= \frac{\sum_{i=0}^{K-1-s}\frac{K}{s+1+i}\binom{K-1}{s+i}(N-1)^i}{\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i} \quad (109)$$

$$\leq \frac{\sum_{i=0}^{K-1-s}\frac{K}{s}\binom{K-1}{s+i}(N-1)^i}{\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^i} = \frac{1}{\lambda}. \quad (110)$$

We upper bound the reciprocal of $\psi_2(N,K,s)$ as,

$$\frac{1}{\psi_2(N,K,s)} = \sum_{i=0}^{K-1-s}\frac{\binom{K-1}{s+i}(N-1)^i}{\binom{K-2}{s-1}}N \quad (111)$$

$$= \sum_{i=0}^{K-1-s}\frac{(K-1)(K-1-s)(K-2-s)\cdots(K-i-s)}{s(s+1)(s+2)\cdots(s+i)} \times N(N-1)^i \quad (112)$$

$$\leq \sum_{i=0}^{K-1-s}\frac{K(K-s)^i}{s^{i+1}}N(N-1)^i \quad (113)$$

$$= \sum_{i=0}^{(1-\lambda)K-1}\frac{(1-\lambda)^i}{\lambda^{i+1}}N(N-1)^i \quad (114)$$

$$= \frac{N}{\lambda}\sum_{i=0}^{(1-\lambda)K-1}\left(\frac{(1-\lambda)(N-1)}{\lambda}\right)^i. \quad (115)$$

Now, if $\lambda > 1-\frac{1}{N}$, then $\frac{(1-\lambda)(N-1)}{\lambda} < 1$. Hence, as $K \to \infty$, $\frac{1}{\psi_2(N,K,s)}$ converges to

$$\lim_{K\to\infty}\frac{1}{\psi_2(N,K,s)} \leq \frac{N}{\lambda}\sum_{i=0}^{\infty}\left(\frac{(1-\lambda)(N-1)}{\lambda}\right)^i \quad (116)$$

$$= \frac{N}{\lambda}\cdot\frac{1}{1-\frac{(1-\lambda)(N-1)}{\lambda}} \quad (117)$$

$$= \frac{N}{N\lambda-(N-1)}. \quad (118)$$

Moreover, (112) can be lower bounded by keeping the first ϵK terms in the sum for any ϵ such that $0 < \epsilon < 1-\lambda$,

$$\frac{1}{\psi_2(N,K,s)} \geq \sum_{i=0}^{\epsilon K}\frac{(K-1)(K-1-s)(K-2-s)\cdots(K-i-s)}{s(s+1)(s+2)\cdots(s+i)} \times N(N-1)^i \quad (119)$$

$$\geq \sum_{i=0}^{\epsilon K}\frac{(K-1)(K-\epsilon K-s)^i}{(s+\epsilon K)^{i+1}}N(N-1)^i \quad (120)$$

$$= \sum_{i=0}^{\epsilon K}\frac{(1-\frac{1}{K})((1-(\lambda+\epsilon))^i}{(\lambda+\epsilon)^{i+1}}N(N-1)^i. \quad (121)$$

Similarly, by taking $K \to \infty$, for any $0 < \epsilon < 1-\lambda$, we have

$$\lim_{K\to\infty}\frac{1}{\psi_2(N,K,s)} \geq \frac{N}{\lambda+\epsilon}\sum_{i=0}^{\infty}\left(\frac{(1-(\lambda+\epsilon))(N-1)}{\lambda+\epsilon}\right)^i \quad (122)$$

$$= \frac{N}{N(\lambda+\epsilon)-(N-1)}. \quad (123)$$

Since ϵ is arbitrarily chosen, then as $K \to \infty$, $\epsilon \to 0$, we have $\psi_2(N,K,s) \to \frac{N\lambda-(N-1)}{N}$.

Consequently, as $K \to \infty$, r_s converges to

$$r_s \to r = \lim_{K\to\infty} \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^i N} \quad (124)$$

$$= \lim_{K\to\infty} \frac{\psi_2(N,K,s)}{\psi_2(N,K,s)+1} \quad (125)$$

$$= \frac{N\lambda - (N-1)}{N\lambda + 1}. \quad (126)$$

Note that if $\lambda = 1 - \frac{1}{N}$, then $r = 0$, while if $\lambda = 1$, then $r = \frac{1}{1+N}$. This means that the restriction in the limit to have $\lambda > 1 - \frac{1}{N}$ is without loss of generality as $\lambda > 1 - \frac{1}{N}$ corresponds to the entire range of r other than the $1 - r$ matching bound. We can write λ as

$$\lambda = \frac{r + (N-1)}{N(1-r)}. \quad (127)$$

Substituting in (107), we have the following upper bound on $\bar{D}(r)$

$$\bar{D}(r) \le \frac{\frac{1}{\lambda}}{\frac{N\lambda-(N-1)}{N} + 1} \quad (128)$$

$$= \frac{N}{\lambda(N\lambda+1)} \quad (129)$$

$$= \frac{N}{\frac{r+(N-1)}{N(1-r)}\left(\frac{r+(N-1)}{(1-r)} + 1\right)} \quad (130)$$

$$= \frac{N^2(1-r)^2}{(r+(N-1))^2 + (1-r)(r+(N-1))} \quad (131)$$

$$= \frac{N^2(1-r)^2}{Nr + N(N-1)} \quad (132)$$

$$= \frac{N(1-r)^2}{(N-1)+r}. \quad (133)$$

The memory-sharing scheme in [29] achieves $\frac{N}{N-1}(1-r)$ if $K \to \infty$, hence the asymptotic unawareness gain is given by the multiplicative factor $\frac{1-r}{1+\frac{r}{N-1}} \le 1$.

For the inner bound, we note that the ith corner point is given by,

$$\tilde{r}_i = \frac{1}{1+N+\cdots+N^i}, \quad i = 1, \cdots, K-1. \quad (134)$$

Therefore, although there exist K linear bounds, it suffices to consider only a small number of them, as the remaining bounds are concentrated around $r = 0$. Denote the gap between the inner and the outer bounds by $\Delta(N,K,r)$. We note that the gap $\Delta(N,\infty,r)$ is a piece-wise convex function for $0 \le r \le 1$ since it is the difference between a convex function $\bar{D}(r)$ and a piece-wise linear function. Hence, the maximizing caching ratio for the gap exists exactly at the corner points $\tilde{r}_i$ and it suffices to examine the gap at these corner points.

For the outer bound, we have

$$\bar{D}(\tilde{r}_i) \le \frac{N\left(1 - \frac{1}{1+N+\cdots+N^i}\right)^2}{(N-1) + \frac{1}{1+N+\cdots+N^i}} \quad (135)$$

$$= \frac{N(1+N+N^2+\cdots+N^i-1)^2}{(N-1)(1+N+\cdots+N^i)^2 + (1+N+\cdots+N^i)} \quad (136)$$

$$= \frac{N^2(1+N+\cdots+N^{i-1})^2}{N^i(1+N+\cdots+N^i)}. \quad (137)$$

Furthermore, for the inner bound, we have

$$\tilde{D}(\tilde{r}_i) = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^i}\right) - \frac{1}{1+N+\cdots+N^i}\left(i+1+\frac{i}{N}+\cdots+\frac{1}{N^i}\right) \quad (138)$$

$$= \frac{1+N+\cdots+N^i}{N^i} - \frac{(i+1)N^i + iN^{i-1} + \cdots + 1}{N^i(1+N+\cdots+N^i)} \quad (139)$$

$$= \frac{(1+N+\cdots+N^i)^2}{N^i(1+N+\cdots+N^i)} - \frac{(1+2N+3N^2+\cdots+(i+1)N^i)}{N^i(1+N+\cdots+N^i)} \quad (140)$$

Consequently, we can upper bound the asymptotic gap at the corner point $\tilde{r}_i$ as

$$\Delta(N,\infty,\tilde{r}_i) = \bar{D}(\tilde{r}_i) - \tilde{D}(\tilde{r}_i) \quad (141)$$

$$\le \frac{N^2(1+N+\cdots+N^{i-1})^2 - (1+N+\cdots+N^i)^2}{N^i(1+N+\cdots+N^i)} + \frac{(1+2N+3N^2+\cdots+(i+1)N^i)}{N^i(1+N+\cdots+N^i)} \quad (142)$$

$$= \frac{-1-2N(1+N+\cdots+N^{i-1}))}{N^i(1+N+\cdots+N^i)} + \frac{(1+2N+3N^2+\cdots+(i+1)N^i)}{N^i(1+N+\cdots+N^i)} \quad (143)$$

$$= \frac{N^2+2N^3+\cdots+(i-1)N^i}{N^i(1+N+\cdots+N^i)} \quad (144)$$

$$= \frac{\frac{1}{N^{i-2}} + \frac{2}{N^{i-3}} + \cdots + (i-1)}{1+N+\cdots+N^i} \quad (145)$$

$$= \frac{\frac{1}{N^{i-2}} + \frac{2}{N^{i-3}} + \cdots + (i-1)}{1+N+\cdots+N^i} \quad (146)$$

Hence, $\Delta(N,\infty,\tilde{r}_i)$ is monotonically decreasing in N. Therefore,

$$\Delta(N,K,r) \le \Delta(2,\infty,r) \le \max_i \frac{(2)^2 + 2(2)^3 + \cdots + (i-1)(2)^i}{2^i(1+2+\cdots+2^i)} \quad (147)$$

For the case $N = 2$, we note that all the inner bounds after the 6th corner point are concentrated around $r = 0$ since $\tilde{r}_i \le \frac{1}{127}$ for $i \ge 6$. Therefore, it suffices to characterize the gap only for the first 6 corner points. Considering the 6th corner point which corresponds to $\tilde{r}_6 = \frac{1}{127} = 0.0078$, and $\bar{D}(r) \le 2$

trivially for all r, and $\tilde{D}(\frac{1}{127}) = 1.8898$. Hence, $\Delta(2, \infty, r) \leq 0.11$, for $r \leq \frac{1}{127}$. Now, we focus on calculating the gap at $\tilde{r}_i$, $i = 1, \cdots, 6$. Examining all the corner points, we see that $r = \frac{1}{15}$ is the maximizing caching ratio for the gap (corresponding to $i = 3$), and $\Delta(2, \infty, \frac{1}{15}) \leq \frac{1}{6}$, which is the worst-case gap. ■

VIII. 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 unknown 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 piece-wise 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., $r \leq \frac{1}{1+N+N^2+\cdots+N^{K-1}}$, where $D^*(r) = (1-r)\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}\right) - r\left((K-1)+\frac{K-2}{N}+\cdots+\frac{1}{N^{K-2}}\right)$; and the very high caching ratio regime, where $D^*(r) = (1-r)(1+\frac{1}{N}) - r$, for $\frac{K-2}{(N+1)K+N^2-2N-2} \leq r \leq \frac{1}{1+N}$ and $D^*(r) = 1-r$, for $r \geq \frac{1}{1+N}$. 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 $\frac{1}{6}$. The outer bound shows significant reduction in the download cost with respect to the case when the cache content is fully known at all databases [29], which achieves $D^*(r) = (1-r)(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}})$ by memory-sharing.

The achievable scheme extends the greedy scheme in [9] so that it starts with exploiting the cache bits as side information. For fixed K, N, there are $K-1$ non-degenerate corner points. These points differ in the number of cached bits that contribute in generating one side information equation. The achievability for the remaining caching ratios is done by memory-sharing between the two adjacent corner points that enclose that caching ratio r. For the converse, we extend the induction-based techniques in [9] and [29] to account for the availability of uncoded and unknown prefetching. The converse proof hinges on developing $K-1$ lower bounds on the length of the undesired portion of the answer string. By applying induction on each bound separately, we obtain the piece-wise linear inner bound.

References

[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," *J. ACM*, vol. 45, no. 6, pp. 965–981, 1998.

[2] W. Gasarch, "A survey on private information retrieval," *Bull. EATCS*, vol. 82, pp. 72–107, 2004.

[3] C. Cachin, S. Micali, and M. Stadler, "Computationally private information retrieval with polylogarithmic communication," in *International Conference on the Theory and Applications of Cryptographic Techniques*. Prague, Czech Republic: Springer, 1999.

[4] R. Ostrovsky and W. Skeith III, "A survey of single-database private information retrieval: Techniques and applications," in *International Workshop on Public Key Cryptography*. Beijing, China: Springer, 2007, pp. 393–411.

[5] S. Yekhanin, "Private information retrieval," *Commun. ACM*, vol. 53, no. 4, pp. 68–73, Apr. 2010.

[6] N. B. Shah, K. V. Rashmi, and K. Ramchandran, "One extra bit of download ensures perfectly private information retrieval," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2014, pp. 856–860.

[7] R. Tajeddine and S. El Rouayheb, "Private information retrieval from MDS coded data in distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7081–7093, Nov. 2016.

[8] T. H. Chan, S.-W. Ho, and H. Yamamoto, "Private information retrieval for coded storage," in *Proc. IEEE ISIT*, Jun. 2015, pp. 2842–2846.

[9] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4075–4088, Jul. 2017.

[10] S. A. Jafar, "Blind interference alignment," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 3, pp. 216–227, Jun. 2012.

[11] H. Sun and S. A. Jafar, "Blind interference alignment for private information retrieval," in *Proc. IEEE ISIT*, Jul. 2016, pp. 560–564.

[12] H. Sun and S. A. Jafar, "The capacity of robust private information retrieval with colluding databases," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2361–2370, Apr. 2018.

[13] 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," in *Proc. IEEE ISIT*, Jun. 2017, pp. 1908–1912.

[14] R. Tajeddine and S. El Rouayheb, "Robust private information retrieval on coded data," in *Proc. IEEE ISIT*, Jun. 2017, pp. 1903–1907.

[15] C. Devet, I. Goldberg, and N. Heninger, "Optimally robust private information retrieval," in *Proc. USENIX Secur. Symp.*, Aug. 2012, pp. 269–283.

[16] I. Goldberg, "Improving the robustness of private information retrieval," in *Proc. IEEE Symp. Secur. Privacy*, May 2007, pp. 131–148.

[17] Y. Zhang and G. Ge. (Oct. 2017), "Private information retrieval from MDS coded databases with colluding servers under several variant models." [Online]. Available: https://arxiv.org/abs/1705.03186ss

[18] H. Sun and S. A. Jafar, "The capacity of symmetric private information retrieval," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 322–329, Jan. 2019.

[19] K. Banawan and S. Ulukus, "The capacity of private information retrieval from coded databases," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1945–1956, Mar. 2018.

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

[21] K. Banawan and S. Ulukus, "The capacity of private information retrieval from Byzantine and colluding databases," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 1206–1219, Feb. 2019.

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

[23] H. Sun and S. A. Jafar, "Multiround private information retrieval: Capacity and storage overhead," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5743–5754, Aug. 2018.

[24] Q. Wang and M. Skoglund. (Oct. 2016). "Symmetric private information retrieval for MDS coded distributed storage." [Online]. Available: https://arxiv.org/abs/1610.04530

[25] R. Freij-Hollanti, O. W. Gnilke, C. Hollanti, and D. A. Karpuk, "Private information retrieval from coded databases with colluding servers," *SIAM J. Appl. Algebra Geometry*, vol. 1, no. 1, pp. 647–664, Aug. 2017.

[26] H. Sun and S. A. Jafar, "Private information retrieval from mds coded data with colluding servers: Settling a conjecture by Freij-Hollanti*et al.*," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1000–1022, Feb. 2018.

[27] Y. Zhang and G. Ge. (Apr. 2017). "A general private information retrieval scheme for MDS coded databases with colluding servers." [Online]. Available: https://arxiv.org/abs/1704.06785

[28] Q. Wang and M. Skoglund. (Aug. 2017). "Linear symmetric private information retrieval for MDS coded distributed storage with colluding servers." [Online]. Available: https://arxiv.org/abs/1708.05673

[29] R. Tandon, "The capacity of cache aided private information retrieval," in *Proc. IEEE Allerton*, Oct. 2017, pp. 1078–1082.

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

[31] M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," *IEEE/ACM Trans. Netw.*, vol. 23, no. 4, pp. 1029–1040, Aug. 2015.

[32] M. Karmoose, L. Song, M. Cardone, and C. Fragouli. (Jan. 2017). "Private broadcasting: An index coding approach." [Online]. Available: https://arxiv.org/abs/1701.04958

[33] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1281–1296, Feb. 2018.

Yi-Peng Wei (S'15) received his B.Sc. in Electrical Engineering from National Tsing Hua University, Taiwan, in June 2009, and M.Sc. in Graduate Institute of Communication Engineering from National Taiwan University, Taiwan, in July 2012. His thesis work was on the low density graph code design. Since August 2013, he has been a graduate student in the Electrical and Computer Engineering Department at the University of Maryland, College Park, working towards his Ph.D. degree. His research focuses on information theory.

Karim Banawan (S'13–M'18) received the B.Sc. and M.Sc. degrees (Highest Hons.) in electrical engineering from Alexandria University, Alexandria, Egypt, in 2008, 2012, respectively, M.Sc. degree in electrical engineering from University of Maryland, College Park, MD, USA in 2017. He is currently pursuing the Ph.D. degree at the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA. He was the recipient of the Distinguished Dissertation Fellowship from the Department of Electrical and Computer Engineering, at the University of Maryland College Park, for his Ph.D. thesis work. His research interests include information theory, wireless communications, physical layer security and private information retrieval.

Sennur Ulukus (S'90–M'98–SM'15–F'16) is the Anthony Ephremides Professor in Information Sciences and Systems in the Department of Electrical and Computer Engineering at the University of Maryland at College Park, where she also holds a joint appointment with the Institute for Systems Research (ISR). Prior to joining UMD, she was a Senior Technical Staff Member at AT&T Labs-Research. She received her Ph.D. degree in Electrical and Computer Engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University, and B.S. and M.S. degrees in Electrical and Electronics Engineering from Bilkent University. Her research interests are in communication theory, information theory, networks and signal processing, with recent focus on private information retrieval, age of information, energy harvesting communications, physical layer security, and wireless energy and information transfer.

Dr. Ulukus is a fellow of the IEEE, and a Distinguished Scholar-Teacher of the University of Maryland. She received the 2003 IEEE Marconi Prize Paper Award in Wireless Communications, an 2005 NSF CAREER Award, the 2010-2011 ISR Outstanding Systems Engineering Faculty Award, and the 2012 ECE George Corcoran Education Award. She is a Distinguished Lecturer of the IEEE Information Theory Society for 2018-2019. She is on the Editorial Board of the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING since 2016. She was an Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Series on Green Communications and Networking (2015-2016), IEEE TRANSACTIONS ON INFORMATION THEORY (2007-2010), and IEEE TRANSACTIONS ON COMMUNICATIONS (2003-2007). She was a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (2015 and 2008), *Journal of Communications and Networks* (2012), and IEEE TRANSACTIONS ON INFORMATION THEORY (2011). She is a TPC co-chair of 2019 ITW, 2017 IEEE ISIT, 2016 IEEE Globecom, 2014 IEEE PIMRC, and 2011 IEEE CTW.