Cache-Aided Private Information Retrieval With Partially Known Uncoded Prefetching: Fundamental Limits

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

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

***Index Terms*— Private information retrieval, caching, side information, distributed databases, uncoded prefetching.**

I. Introduction

IN TODAY'S communication networks, the end-users are equipped with large memories, and the data transmitted in the network has shifted from real-time generated data like voice to pre-generated content like movies. These two factors together have enabled caching techniques, which store data in user cache a priori in order to reduce the peak-hour network traffic load. In the meanwhile, privacy has become an important consideration for users, who wish to download data from publicly accessable databases as privately and as efficiently as possible. This is studied under the subject of private information retrieval (PIR). In this paper, we combine the caching and PIR approaches, and consider the problem of PIR for a cache-enabled end-user.

The PIR problem considers the privacy of the requested message by a user from distributed databases. In the classical setting of PIR [1], there are N non-communicating databases, each storing the same set of K messages. The user wishes to download one of these K messages without letting the databases know the identity of the desired message. A feasible scheme is to download all the K messages from a database. However, this results in excessive download cost since it results in a download that is K times the size of the desired message. The goal of the PIR problem is to construct an efficient retrieval scheme such that no database knows which message is retrieved. The PIR problem has originated in computer science [1]–[5] and has drawn significant attention in information theory [6]–[11] in recent years.

Recently, Sun and Jafar [12] have characterized the optimal normalized download cost for the classical PIR problem to be $\frac{D}{L} = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right)$, where L is the message size and D is the total number of downloaded bits from the N databases. Since the work of Sun and Jafar [12], many interesting variants of the classical PIR problem have been investigated, such as, PIR from colluding databases, robust PIR, symmetric PIR, PIR from MDS-coded databases, PIR for arbitrary message lengths, multi-round PIR, multi-message PIR, PIR from Byzantine databases, secure symmetric PIR with adversaries, cache-aided PIR, PIR with private side information (PSI), PIR for functions, storage constrained PIR, and their several combinations [13]–[35].

Also recently, Maddah-Ali and Niesen [36] have proposed a theoretical framework to study the tradeoff between the cache memory size of users and the network traffic load for a two-phase scheme. In the prefetching phase, when the network traffic is low, the server allocates data to the user's cache memory. In the retrieval phase, when the network traffic is high, the server delivers the messages according to the users' requests. By jointly designing the prefetching of the cache content during the low traffic period and the delivery of the requested content during the high traffic period, the coded-caching technique proposed in [36] achieves a global caching gain significantly reducing the peak-time traffic load. The concept of coded-caching has been applied to many different scenarios, such as, decentralized networks, device

Manuscript received December 11, 2017; revised April 17, 2018; accepted April 18, 2018. Date of publication June 7, 2018; date of current version September 12, 2018. This work was supported by NSF under Grants CNS 13-14733, CCF 14-22111, CNS 15-26608, and CCF 17-13977. This paper is presented in part at the IEEE ICC, Kansas City, MO, USA, May 2018. *(Corresponding author: Sennur Ulukus.)*

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

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/JSAC.2018.2844940

to device networks, random demands, online settings, general cache networks, security constraints, finite file size constraints, broadcast channels and their several combinations [37]–[55].

The caching technique not only reduces the traffic load but can also help the user to privately retrieve the desired file more efficiently by providing additional side information. The interplay between side information and the PIR problem has been studied recently in [27] and [29]-[32]. We first recall that the achievability scheme proposed in [12] is based on three principles: database symmetry, message symmetry, and side information utilization. The side information in [12] comes from the undesired bits downloaded from the other $(N-1)$ databases. Side information plays an important role in the PIR problem; for instance, when $N=1$ (single database), i.e., when no side information is available, the normalized download cost is K, which is the largest possible. Caching can improve PIR download cost by providing useful side information.

Reference [27] is the first to study the cache-aided PIR problem. In [27], the user has a memory of size KLr bits and can store an arbitrary function of the K messages, where $0 \leq r \leq 1$ is the caching ratio. Reference [27] considers the case when the cached content is fully known to all N databases, and determines the optimal normalized download cost to be $(1-r)\left(1+\frac{1}{N}+\cdots+\frac{1}{N^{K-1}}\right)$. Although the result is pessimistic since it implies that the user cannot utilize the cached content to further reduce the download cost, [27] reveals two new dimensions for the cache-aided PIR problem. The first one is the databases' awareness of the side information at its initial acquisition. Different from [27] and [29]–[31] study the case when the databases are unaware of the cached side information, and [32] studies the case when the databases are partially aware of the cached side information. The second one is the structure of the side information. Instead of storing an arbitrary function of the K messages, [29], [31], and [32] consider caching M full messages out of total K messages, and [30] considers storing an r fraction of each message in uncoded form.

This paper is closely related to [30]. In [30], the databases are assumed to be completely unaware of the side information. However, this may be practically challenging to implement. Here, we consider a more natural model which uses the same set of databases for both prefetching and retrieval phases. Therefore, different from [30], here each database gains partial knowledge about the side information, that is the part it provides during the prefetching phase. Our aim is to determine if there is a rate loss due to this partial knowledge with respect to the fully unknown case in [30], and characterize this rate loss as a function of K, N and r.

In this work, we consider PIR with *partially known uncoded prefetching*. We consider the PIR problem with a two-phase scheme, namely, prefetching phase and retrieval phase. In the prefetching phase, the user caches an uncoded $\frac{r}{N}$ fraction of each message from the nth database. The nth database is aware of these $\frac{KLr}{N}$ bit side information, while it has no knowledge about the cached bits from the other $(N-1)$ databases. We aim at characterizing the optimal tradeoff between the normalized download cost $\frac{D(r)}{L}$ and the caching ratio r. For the outer bound, we explicitly determine the achievable download rates for specific $K+1$ caching ratios. Download rates for any other caching ratio can be achieved by memory-sharing between the nearest explicit points. Hence, the outer bound is a piece-wise linear curve which consists of K line segments. For the inner bound, we extend the techniques of [12] and [30] to obtain a piece-wise linear curve which also consists of K line segments. We show that the inner and the outer bounds match exactly at three line segments for any K. Consequently, we characterize the optimal tradeoff for the very low ($r \leq \frac{1}{N^{K-1}}$) and the very high ($r \geq \frac{K-2}{N^2-3N+KN}$) caching ratios. As a direct corollary, we fully characterize the optimal download cost caching ratio tradeoff for $K=3$ messages. For general K, N and r, we show that the worst-case additive gap between the inner and the outer bounds is $\frac{5}{32}$.

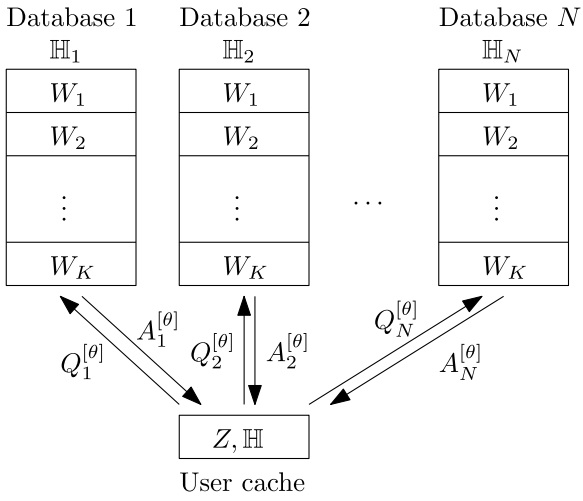


Fig. 1. System model.

II. System Model

We consider a PIR problem with N non-communicating databases; see Fig. 1. Each database stores an identical copy of K statistically independent messages, $W_1, \ldots, W_K$. Each message is L bits long,

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

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

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

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

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

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

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

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

$$I(W_1, \ldots, W_K; Q_1^{[\theta]}, \ldots, Q_N^{[\theta]}) = 0. \tag{4}$$

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

$$\begin{aligned} &(Q_n^{[1]}, A_n^{[1]}, W_1, \ldots, W_K, \mathbb{H}_n) \\ &\qquad \sim (Q_n^{[\theta]}, A_n^{[\theta]}, W_1, \ldots, W_K, \mathbb{H}_n), \end{aligned} \tag{5}$$

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

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

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

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

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

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

For a fixed N, K, and caching ratio r, a pair $(D(r), L)$ is achievable if there exists a PIR scheme for message of size L bits long with partially known uncoded prefetching satisfying the privacy constraint (5) and the reliability constraint (7), where $D(r)$ represents the expected number of downloaded bits (over all the queries) from the N databases via the answering strings $A_{1:N}^{[\theta]}$, where $A_{1:N}^{[\theta]} = (A_1^{[\theta]}, \ldots, A_N^{[\theta]})$, i.e.,

$$D(r) = \sum_{n=1}^{N} H\left(A_n^{[\theta]}\right). \tag{8}$$

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

$$D^*(r) = \inf\left\{ \frac{D(r)}{L} : (D(r), L) \text{ is achievable} \right\}, \tag{9}$$

which is a function of the caching ratio r.

III. Main Results

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

Theorem 1 (Outer Bound): In the cache-aided PIR with partially known uncoded prefetching, for the caching ratio

$$r_s = \frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s} \binom{K-1}{s+i}(N-1)^{i+1}}, \tag{10}$$

and length of the message

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

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

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

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

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

The proof of Theorem 1 is provided in Section IV. The outer bound in Theorem 1 is a piece-wise linear curve, which consists of K line segments. These K line segments intersect at the points r_s.

We characterize an inner bound (converse bound), which is denoted by $\tilde{D}(r)$, for the optimal normalized download cost $D^*(r)$ for general K, N, r.

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

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

$$= \max_{i \in \{2, \cdots, K+1\}} \sum_{j=0}^{K+1-i} \frac{1}{N^j} - (K+2-i)r. \tag{15}$$

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

$$\tilde{r}_i = \frac{1}{N^{K-i}}, \quad i = 1, \cdots, K-1. \tag{16}$$

The outer bounds provided in Theorem 1 and the inner bounds provided in Theorem 2 match for some caching ratios r as summarized in the following corollary.

Corollary 1 (Optimal Tradeoff for Very Low and Very High Caching Ratios): In the cache-aided PIR with partially known uncoded prefetching, for very low caching ratios, i.e., for $r \leq \frac{1}{N^{K-1}}$, the optimal normalized download cost is given by,

$$D^*(r) = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1}}\right) - Kr. \tag{17}$$

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

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

Proof: From (10) and (16), we have

$$r_1 = \tilde{r}_1 = \frac{1}{N^{K-1}}, \tag{19}$$

$$r_{K-2} = \frac{K-2}{N^2-3N+KN}, \tag{20}$$

$$r_{K-1} = \tilde{r}_{K-1} = \frac{1}{N}. \tag{21}$$

For the outer bound of the case of very low caching ratios, from (12), we have

$$\bar{D}(r_1) = \frac{\sum_{i=0}^{K-2} \binom{K}{2+i}(N-1)^{i+1}}{\binom{K-2}{0} + \sum_{i=0}^{K-2} \binom{K-1}{1+i}(N-1)^{i+1}} \tag{22}$$

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

$$= \frac{\frac{1}{(N-1)} \left(N^K - 1 - K(N-1)\right)}{N^{K-1}} \tag{24}$$

$$= \frac{N^K - KN + K - 1}{N^K - N^{K-1}}. \tag{25}$$

For the inner bound of the case of very low caching ratios, from (14), by choosing $i = 2$ and using $r = r_1$, we have

$$\tilde{D}(r_1) \geq (1 - r_1) \sum_{j=0}^{K-1} \frac{1}{N^j} - r_1 \left(1 - \frac{1}{N}\right) \sum_{j=0}^{K-2} \frac{K-1-j}{N^j} \tag{26}$$

$$= \left(1 - \frac{1}{N^{K-1}}\right) \frac{1 - \frac{1}{N^K}}{1 - \frac{1}{N}} - \frac{1}{N^{K-1}} \left(1 - \frac{1}{N}\right) \frac{K - \frac{K}{N} - 1 + \frac{1}{N^K}}{\left(1 - \frac{1}{N}\right)^2} \tag{27}$$

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

$$= \frac{N^K - KN + K - 1}{N^K - N^{K-1}} = \bar{D}(r_1). \tag{29}$$

Thus, since $\tilde{D}(r_1) \leq \bar{D}(r_1)$ by definition, (29) implies $\tilde{D}(r_1) = \bar{D}(r_1)$.

For the outer bound of the case of very high caching ratios, from (12), we have

$$\bar{D}(r_{K-2}) = \frac{\sum_{i=0}^{1} \binom{K}{K-1+i}(N-1)^{i+1}N}{N\binom{K-2}{K-3} + \sum_{i=0}^{1} \binom{K-1}{K-2+i}(N-1)^{i+1}N} \tag{30}$$

$$= \frac{N^2 + KN - 2N - K + 1}{N^2 - 3N + KN}, \tag{31}$$

and for the inner bound of the case of very high caching ratios, from (14) by choosing $i = K$ and using $r = r_{K-2}$,

$$\tilde{D}(r_{K-2}) \geq (1 - r_{K-2}) \sum_{j=0}^{1} \frac{1}{N^j} - r_{K-2} \left(1 - \frac{1}{N}\right) \sum_{j=0}^{0} \frac{1-j}{N^j} \tag{32}$$

$$= 1 + \frac{1}{N} - 2r_{K-2} \tag{33}$$

$$= \frac{N^2 + KN - 2N - K + 1}{N^2 - 3N + KN} = \bar{D}(r_{K-2}) \tag{34}$$

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

Finally, from (12), $\bar{D}(r_{K-1}) = \frac{N-1}{N}$, and from (14) by choosing $i = K+1$ and using $r = r_{K-1}$,

$$\tilde{D}(r_{K-1}) \geq \frac{N-1}{N} = \bar{D}(r_{K-1}) \tag{35}$$

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 $r = r_{K-1}$. In addition to 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. ∎

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

For the case $K = 3$, we have exact tradeoff curve for any N, r as shown in the following corollary.

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

$$D^*(r) = \begin{cases} 1 + \frac{1}{N} + \frac{1}{N^2} - 3r, & 0 \leq r \leq \frac{1}{N^2} \\ 1 + \frac{1}{N} - 2r, & \frac{1}{N^2} \leq r \leq \frac{1}{N} \\ 1 - r, & \frac{1}{N} \leq r \leq 1. \end{cases} \tag{36}$$

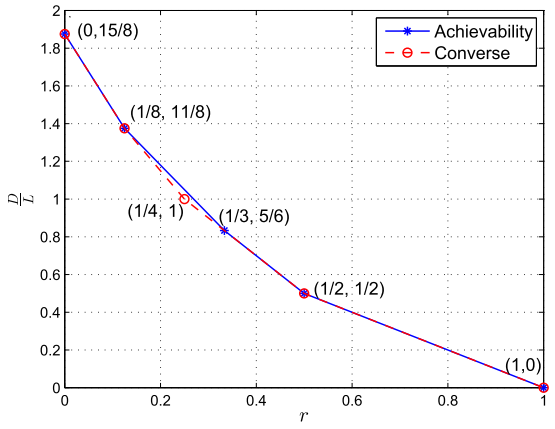


Fig. 2. Inner and outer bounds for $K = 4$, $N = 2$.

TABLE I

QUERY TABLE FOR $K = 3$, $N = 2$, $r_1 = \frac{1}{4}$

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

TABLE II

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

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

Proof: The proof follows from the proof of Corollary 1. Note that in this case, from (19) and (20), $r_1 = r_{K-2} = \frac{1}{N^2}$; and from (21), $r_2 = r_{K-1} = \frac{1}{N}$. Thus, we have a tight result for $0 \leq r \leq r_1 = \frac{1}{N^2}$ (very low caching ratios) and a tight result for $r_{K-2} = r_1 = \frac{1}{N^2} \leq r \leq 1$, 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 (10)-(12) and written explicitly in (36). ■

IV. ACHIEVABLE SCHEME

In this section, we present an achievable scheme for the outer bounds provided in Theorem 1. Our achievable scheme is based on [12], [27], and [30]. We first provide achievable schemes for the caching ratios r_s in (10) by applying the principles in [12]: 1) database symmetry, 2) message symmetry within each database, and 3) exploiting undesired messages as side information. For an arbitrary caching ratio $r \neq r_s$, we apply the memory-sharing scheme in [27]. Since the cached content is partially known by the databases, the achievable scheme is different from that in [30]. We first use the case of $K = 3$, $N = 2$ to illustrate the main ideas of our achievability scheme.

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

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

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

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

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

3) Caching Ratio $r = \frac{1}{3}$: We choose the message size as 12 bits. In the prefetching phase, for caching ratio $r = \frac{1}{3}$, the user caches 4 bits from each message. Therefore, the user caches 2 bits from each database for each message. Therefore, $Z_1 = (a_1, a_2, b_1, b_2, c_1, c_2)$ and $Z_2 = (a_3, a_4, b_3, b_4, c_3, c_4)$.

TABLE III

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

s	DB1	DB2
$s = 1$	$a_5 + b_3$	$a_7 + b_1$
	$a_6 + c_3$	$a_8 + c_1$
	$b_5 + c_5$	$b_6 + c_6$
	$a_9 + b_6 + c_6$	$a_{10} + b_5 + c_5$
$s = 2$	$a_{11} + b_4 + c_4$	$a_{12} + b_2 + c_2$

$Z_1 = (a_1, a_2, b_1, b_2, c_1, c_2)$	$Z_2 = (a_3, a_4, b_3, b_4, c_3, c_4)$

In the retrieval phase, we combine the achievable schemes in Section IV-A.1 and IV-A.2 as shown in Table III. The normalized download cost is $\frac{5}{6}$. By applying [27, Lemma 1] and taking $\alpha = \frac{2}{3}$, we can show that $\bar{D}(\frac{1}{3}) = \bar{D}(\frac{2}{3}\cdot\frac{1}{4}+\frac{1}{3}\cdot\frac{1}{2}) = \frac{2}{3}\bar{D}(\frac{1}{4}) + \frac{1}{3}\bar{D}(\frac{1}{2}) = \frac{2}{3}\cdot 1 + \frac{1}{3}\cdot\frac{1}{2} = \frac{5}{6}$.

B. Achievable Scheme

We first present the achievable scheme for the caching ratios r_s given in (10). Then, we apply the memory-sharing scheme provided in [27] for the intermediate caching ratios.

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

$$L(s) = N\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}N. \quad (37)$$

In the prefetching phase, for each message the user randomly and independently chooses $N\binom{K-2}{s-1}$ bits to cache, and caches $\binom{K-2}{s-1}$ bits from each database for each message. Therefore, the caching ratio r_s is equal to

$$r_s = \frac{N\binom{K-2}{s-1}}{N\binom{K-2}{s-1} + \sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}N}. \quad (38)$$

In the retrieval phase, the user applies the PIR scheme in Algorithm 1.

Since the desired bits are added to the side information which is either obtained from the cached bits (if $t = s+1$) or from the remaining $(N-1)$ databases in the $(t-1)$th round when $t > s+1$, the user can decode the uncached portion of the desired message by canceling out the side information bits. In addition, for each database, each message is queried equally likely with the same set of equations, which guarantees privacy as in [12]. Therefore, the privacy constraint in (5) and the reliability constraint in (7) are satisfied.

We now calculate the total number of downloaded bits for the caching ratio r_s in (38). For the round $t = s+1$, we exploit s cached bits to form the side information equation. Therefore, each download is a sum of $s+1$ bits. For each database, we utilize the side information cached from other $N-1$ databases. In addition to the message symmetry step enforcing symmetry across K messages, we download $\binom{K}{s+1}(N-1)$ bits from a database. Due to the database symmetry step, in total, we download $\binom{K}{s+1}(N-1)N$ bits. For the round $t = s+i > s+1$, we exploit $s+i-1$ undesired bits downloaded from the $(t-1)$th round to form the side information equation. Due to message symmetry and database symmetry, we download $\binom{K}{s+1+i}(N-1)^{i+1}N$ bits. Overall, the total number of downloaded bits is,

$$D(r_s) = \sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^{i+1}N. \quad (39)$$

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

Algorithm 1 PIR Scheme

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

cost is,

$$\bar{D}(r_s) = \frac{D(r_s)}{L(s)} = \frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^{i+1}N}{N\binom{K-2}{s-1}+\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}N}. \quad (40)$$

2) Achievable Scheme for the Caching Ratios Not Equal to r_s: For caching ratios r which are not exactly equal to (38) for some s, we first find an s such that $r_s < r < r_{s+1}$. We choose $0 < \alpha < 1$ such that $r = \alpha r_s + (1-\alpha) r_{s+1}$. By using the memory-sharing scheme in [27, Lemma 1], we achieve the following normalized download cost,

$$\bar{D}(r) = \alpha \bar{D}(r_s) + (1-\alpha)\bar{D}(r_{s+1}). \quad (41)$$

V. Converse Proof

In this section, we derive an inner bound for the cache-aided PIR with partially known uncoded prefetching. We extend the techniques in [12] and [30] to our problem. The main difference between this proof and that in [30] is the usage of privacy constraint given in (5).

Lemma 1 (Interference Lower Bound [30, Lemma 1]): For the cache-aided PIR with partially known uncoded prefetching, the interference from undesired messages within the answering strings $D(r) - L(1-r)$ is lower bounded by,

$$D(r) - L(1-r) + o(L) \geq I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \quad (42)$$

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

The proof of Lemma 1 is similar to [30, Lemma 1]. In the following lemma, we prove an inductive relation for the mutual information term on the right hand side of (42).

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

$$I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) \geq \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) + \frac{L(1-r)}{N} + \frac{1-N}{N}(K-k+1)Lr - o(L). \quad (43)$$

Lemma 2 is a generalization of [12, Lemma 6] and [30, Lemma 2], and it reduces to [12, Lemma 6] when $r = 0$. Compared to [30, Lemma 2], the lower bound in (43) is increased by $\frac{(K-k+1)Lr}{N}$, since the cached content is partially known by the databases.

Proof: We start with the left hand side of (43),

$$I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]} | W_{1:k-1}, Z, \mathbb{H}\right) = I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, Z, \mathbb{H} | W_{1:k-1}\right) - I(W_{k:K}; Z, \mathbb{H} | W_{1:k-1}). \quad (44)$$

For the first term on the right hand side of (44), we have

$$\begin{aligned} & I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, Z, \mathbb{H} | W_{1:k-1}\right) \\ &= \frac{1}{N} N I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}, Z, \mathbb{H} | W_{1:k-1}\right) && (45) \\ &\geq \frac{1}{N}\sum_{n=1}^{N} I\left(W_{k:K}; Q_n^{[k-1]}, A_n^{[k-1]}, Z_n, \mathbb{H}_n | W_{1:k-1}\right) && (46) \\ &= \frac{1}{N}\left[\sum_{n=1}^{N} I\left(W_{k:K}; Q_n^{[k-1]}, A_n^{[k-1]} | W_{1:k-1}, Z_n, \mathbb{H}_n\right) + \sum_{n=1}^{N} I\left(W_{k:K}; Z_n, \mathbb{H}_n | W_{1:k-1}\right)\right] && (47) \\ &= \frac{1}{N}\left[\sum_{n=1}^{N} I\left(W_{k:K}; Q_n^{[k-1]}, A_n^{[k-1]} | W_{1:k-1}, Z_n, \mathbb{H}_n\right) + N \times \frac{(K-k+1)Lr}{N}\right] && (48) \\ &\overset{(5)}{=} \frac{1}{N}\sum_{n=1}^{N} I\left(W_{k:K}; Q_n^{[k]}, A_n^{[k]} | W_{1:k-1}, Z_n, \mathbb{H}_n\right) + \frac{(K-k+1)Lr}{N} && (49) \\ &\overset{(3),(4)}{=} \frac{1}{N}\sum_{n=1}^{N} I\left(W_{k:K}; A_n^{[k]} | W_{1:k-1}, Z_n, \mathbb{H}_n, Q_n^{[k]}\right) + \frac{(K-k+1)Lr}{N} && (50) \\ &\overset{(6)}{=} \frac{1}{N}\sum_{n=1}^{N} H\left(A_n^{[k]} | W_{1:k-1}, Z_n, \mathbb{H}_n, Q_n^{[k]}\right) + \frac{(K-k+1)Lr}{N} && (51) \\ &\geq \frac{1}{N}\sum_{n=1}^{N} H\left(A_n^{[k]} | W_{1:k-1}, Z, \mathbb{H}, Q_{1:N}^{[k]}, A_{1:n-1}^{[k]}\right) + \frac{(K-k+1)Lr}{N} && (52) \\ &\overset{(6)}{=} \frac{1}{N}\sum_{n=1}^{N} I\left(W_{k:K}; A_n^{[k]} | W_{1:k-1}, Z, \mathbb{H}, Q_{1:N}^{[k]}, A_{1:n-1}^{[k]}\right) + \frac{(K-k+1)Lr}{N} && (53) \\ &= \frac{1}{N} I\left(W_{k:K}; A_{1:N}^{[k]} | W_{1:k-1}, Z, \mathbb{H}, Q_{1:N}^{[k]}\right) + \frac{(K-k+1)Lr}{N} && (54) \\ &\overset{(3),(4)}{=} \frac{1}{N} I\left(W_{k:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k-1}, Z, \mathbb{H}\right) + \frac{(K-k+1)Lr}{N} && (55) \\ &\overset{(7)}{=} \frac{1}{N} I\left(W_{k:K}; W_k, Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k-1}, Z, \mathbb{H}\right) + \frac{(K-k+1)Lr}{N} - o(L) && (56) \\ &= \frac{(K-k+1)Lr}{N} + \frac{1}{N}\left[I\left(W_{k:K}; W_k | W_{1:k-1}, Z, \mathbb{H}\right) + I\left(W_{k:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right)\right] - o(L) && (57) \\ &= \frac{(K-k+1)Lr}{N} + \frac{L(1-r)}{N} + \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]} | W_{1:k}, Z, \mathbb{H}\right) - o(L), && (58) \end{aligned}$$

where (46) follows from the non-negativity of mutual information, (48) is due to the fact that from the nth database, the user prefetches $\frac{KLr}{N}$ bits, (49) follows from the privacy constraint, (50) and (55) follow from the independence of $W_{k:K}$ and $Q_n^{[k]}$, (51) and (53) follow from the fact that the answering string $A_n^{[k]}$ is a deterministic function of $(W_{1:K}, Q_n^{[k]})$, (52) follows from conditioning reduces entropy, and (56) follows from the reliability constraint.

For the second term on the right hand side of (44), we have

$$\begin{aligned} &I(W_{k:K}; Z, \mathbb{H}|W_{1:k-1}) \\ &\quad = H\left(W_{k:K}|W_{1:k-1}\right) - H(W_{k:K}|W_{1:k-1}, Z, \mathbb{H}) \quad (59) \\ &\quad = (K-k+1)\,Lr \quad (60) \end{aligned}$$

where (60) follows from the uncoded nature of the cached bits.

Combining (44), (58) and (60) yields (43). ∎

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 by varying the index k in the lemma from $k=2$ to $k=K$, and by using the non-negativity of mutual information for the Kth bound. Next, we inductively lower bound each term of Lemma 1 by using Lemma 2 $(K-k+1)$ times to get K explicit lower bounds.

Lemma 3: For fixed N, K and r, we have

$$\begin{aligned} D(r) \geq\; & L(1-r)\sum_{j=0}^{K+1-k}\frac{1}{N^j} \\ & - Lr\left(1-\frac{1}{N}\right)\sum_{j=0}^{K-k}\frac{K+1-k-j}{N^j} + o(L), \quad (61) \end{aligned}$$

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

Proof: We have

$$\begin{aligned} D(r) &\overset{(42)}{\geq} I\left(W_{k:K}; Q_{1:N}^{[k-1]}, A_{1:N}^{[k-1]}|W_{1:k-1}, Z, \mathbb{H}\right) \\ &\quad + L(1-r) - o(L) \quad (62) \\ &\overset{(43)}{\geq} \frac{1}{N} I\left(W_{k+1:K}; Q_{1:N}^{[k]}, A_{1:N}^{[k]}|W_{1:k}, Z, \mathbb{H}\right) \\ &\quad + L(1-r)\left(1+\frac{1}{N}\right) \\ &\quad - Lr\left(1-\frac{1}{N}\right)(K-k+1) - o(L) \quad (63) \\ &\overset{(43)}{\geq} \frac{1}{N^2} I\left(W_{k+2:K}; Q_{1:N}^{[k+1]}, A_{1:N}^{[k+1]}|W_{1:k+1}, Z, \mathbb{H}\right) \\ &\quad + L(1-r)\left[1+\frac{1}{N}+\frac{1}{N^2}\right] - Lr\left(1-\frac{1}{N}\right) \\ &\quad \times\left[(K-k+1)+\frac{(K-k)}{N}\right] - o(L) \quad (64) \\ &\overset{(43)}{\geq} \cdots \quad (65) \\ &\overset{(43)}{\geq} L(1-r)\sum_{j=0}^{K+1-k}\frac{1}{N^j} - Lr\left(1-\frac{1}{N}\right) \\ &\quad \times\sum_{j=0}^{K-k}\frac{K+1-k-j}{N^j} + o(L), \quad (66) \end{aligned}$$

where (62) follows from Lemma 1, and the remaining steps follow from the successive application of Lemma 2. ∎

TABLE IV

QUERY TABLE FOR $K=4$, $N=2$ AND $r_1=\frac{1}{8}$

s	DB1	DB2
$s=1$	a_3+b_2	a_6+b_1
	a_4+c_2	a_7+c_1
	a_5+d_2	a_8+d_1
	b_3+c_3	b_5+c_5
	b_4+d_3	b_6+d_5
	c_4+d_4	c_6+d_6
	$a_9+b_5+c_5$	$a_{12}+b_3+c_3$
	$a_{10}+b_6+d_5$	$a_{13}+b_4+d_3$
	$a_{11}+c_6+d_6$	$a_{14}+c_4+d_4$
	$b_7+c_7+d_7$	$b_8+c_8+d_8$
	$a_{15}+b_8+c_8+d_8$	$a_{16}+b_7+c_7+d_7$
	$Z_1=(a_1,b_1,c_1,d_1)$	$Z_2=(a_2,b_2,c_2,d_2)$

TABLE V

QUERY TABLE FOR $K=4$, $N=2$, $r_2=\frac{1}{3}$

s	DB1	DB2
$s=2$	$a_5+b_3+c_3$	$a_8+b_1+c_1$
	$a_6+d_3+b_4$	$a_9+d_1+b_2$
	$a_7+c_4+d_4$	$a_{10}+c_2+d_2$
	$b_5+c_5+d_5$	$b_6+c_6+d_6$
	$a_{11}+b_6+c_6+d_6$	$a_{12}+b_5+c_5+d_5$
	$Z_1=\begin{pmatrix}a_1,a_2,b_1,b_2,\\ c_1,c_2,d_1,d_2\end{pmatrix}$	$Z_2=\begin{pmatrix}a_3,a_4,b_3,b_4,\\ c_3,c_4,d_3,d_4\end{pmatrix}$

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

$$\begin{aligned} D^*(r) \geq\; & (1-r)\sum_{j=0}^{K+1-k}\frac{1}{N^j} \\ & - r\left(1-\frac{1}{N}\right)\sum_{j=0}^{K-k}\frac{K+1-k-j}{N^j}. \quad (67) \end{aligned}$$

Since (67) gives K intersecting line segments, the normalized download cost is lower bounded by their maximum value as follows

$$\begin{aligned} D^*(r) \geq\; & \max_{i\in\{2,\cdots,K+1\}}(1-r)\sum_{j=0}^{K+1-i}\frac{1}{N^j} \\ & - r\left(1-\frac{1}{N}\right)\sum_{j=0}^{K-i}\frac{K+1-i-j}{N^j}. \quad (68) \end{aligned}$$

VI. FURTHER EXAMPLES

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

For $K=4$ and $N=2$, we present achievable PIR schemes for caching ratios $r_1=\frac{1}{8}$ in Table IV, $r_2=\frac{1}{3}$ in Table V, and $r_3=\frac{1}{2}$ in Table VI. The PIR schemes aim to retrieve message W_1, where we use a_i to denote its bits. The achievable normalized download costs for these caching ratios are $\frac{11}{8}$, $\frac{5}{6}$ and $\frac{1}{2}$, respectively. The plot of the inner and outer bounds can be found in Fig. 2.

TABLE VI

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

s	DB1	DB2
$s = 3$	$a_3 + b_2 + c_2 + d_2$	$a_4 + b_1 + c_1 + d_1$
	$Z_1 = (a_1, b_1, c_1, d_1)$	$Z_2 = (a_2, b_2, c_2, d_2)$

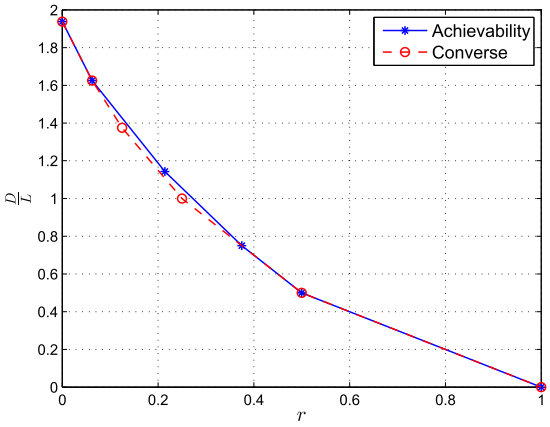


Fig. 3. Inner and outer bounds for $K = 5$, $N = 2$.

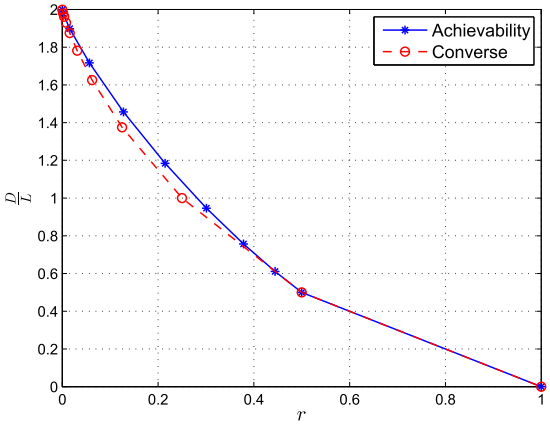


Fig. 4. Inner and outer bounds for $K = 10$, $N = 2$.

B. $K = 5$, $K = 10$ 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. 3, 4 and 5. For fixed N as K grows, the gap between the achievable bound and converse bound increases. This observation will be made specific in Section VII.

VII. GAP ANALYSIS

In this section, we analyze the gap between the achievable bounds given in (12) and the converse bounds given in (14). We first observe that for fixed number of databases N, as the number of messages K increases, the achievable normalized download cost increases, and for large enough caching ratios $r \geq \frac{1}{N}$, the PIR schemes for different number of messages share the same normalized download cost $1 - r$. In addition to the monotonicity, the achievable normalized download cost for

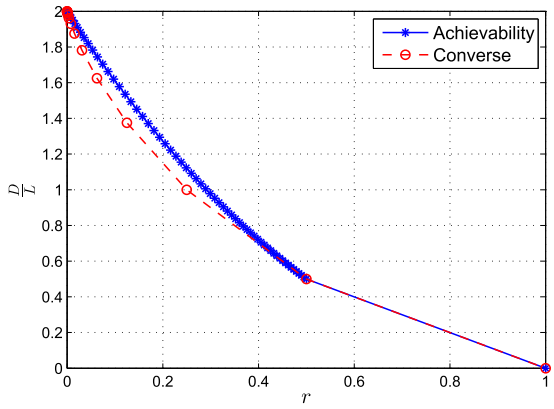


Fig. 5. Inner and outer bounds for $K = 100$, $N = 2$.

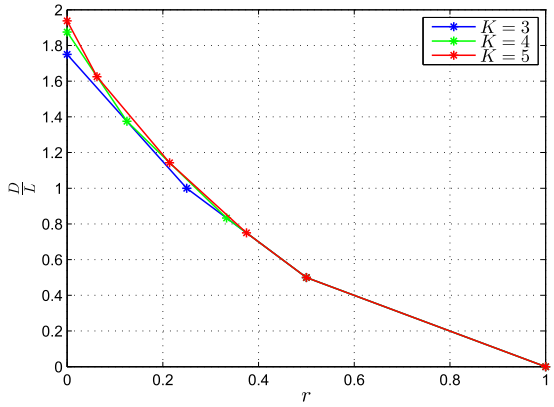


Fig. 6. Outer bounds for $N = 2$, $K = 3$, $K = 4$ and $K = 5$.

$K + 1$ messages has a special relationship with the achievable normalized download cost for K messages. We first use an example to illustrate this property. For $N = 2$, $K = 3$, $K = 4$, and $K = 5$, the achievable bounds are shown in Fig. 6. The achievable bound for $K = 5$ is above the achievable bound for $K = 4$, and the achievable bound for $K = 4$ is above the achievable bound for $K = 3$. By denoting $r_s^{(K)}$ as the caching ratio with total K messages and parameter s (see (10)), 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 summarize this result in the following lemma.

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

$$r_s^{(K+1)} = \alpha r_{s-1}^{(K)} + (1 - \alpha) r_s^{(K)}, \tag{69}$$

$$\bar{D}(r_s^{(K+1)}) = \alpha \bar{D}(r_{s-1}^{(K)}) + (1 - \alpha) \bar{D}(r_s^{(K)}), \tag{70}$$

where $0 \leq \alpha \leq 1$.

The proof of Lemma 4 is similar to [30, Lemma 4].

After showing the monotonicity of the achievable bounds, we show that as $K \to \infty$, the asymptotic upper bound for the

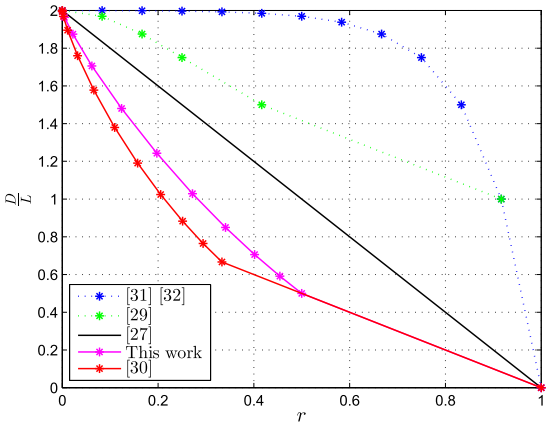


Fig. 7. Outer bounds for $N = 2$, $K = 12$ for different cache-aided PIR models.

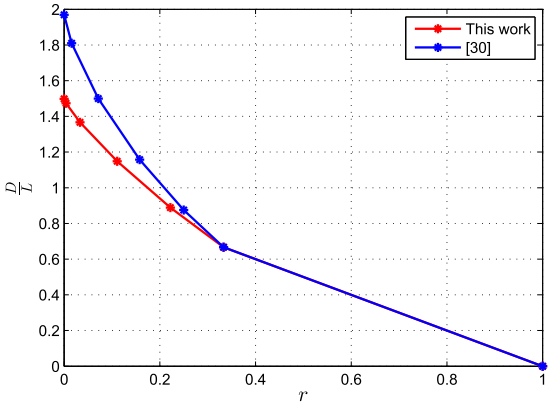


Fig. 8. Comparison between this work and [30] for $N = 3$ and $K = 6$.

achievable bounds is given as in the following lemma. With this asymptotic upper bound, we conclude that the worst-case additive gap is $\frac{5}{32}$.

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

$$\bar{D}(r) \leq \frac{N}{N-1}(1-r)^2 \tag{71}$$

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

The detailed proof of Lemma 5 is provided in Section IX. We note that the outer bound is monotonically increasing in K. Therefore, we first derive an asymptotic upper bound as $K \to \infty$ for the outer bound as in (71). Then, we show that most of the K inner bounds concentrate around $r = 0$. Therefore, we only need to consider a small number of the inner bounds for the worst-case gap analysis.

VIII. Comparisons With Other Cache-Aided PIR Models

In this section, we compare the normalized download costs between different cache-aided PIR models subjected to same memory size constraint. We first use an example of $N = 2$ and $K = 12$ (see Fig. 7) to show the relative normalized download costs for different models. In [29], [31], and [32], the user caches M full messages out of total K messages. In order to compare with other cache-aided PIR schemes, we use $\frac{M}{K}$ as the caching ratio. Since the PIR schemes are only reported for the corner points in [29], [31], and [32], we use dotted lines to connect the corner points. For [30], [27], and this work, since we can apply memory-sharing to achieve the download costs between the corner points, we use solid lines to connect the corner points.

We first compare [29], [31], and [32], in which the user caches M full messages out of K messages and the databases are (partially) unaware. In [31] and [32], the user not only wishes to protect the privacy of the desired messages but also wishes to protect the privacy of the cached messages. Note that the other works ([27], [29], [30], and this work) only consider to protect the privacy of the desired messages. Since the message privacy constraint is less restricted, [29] achieves lower normalized download cost than [31] and [32]. The main difference between [31] and [32] is that the databases are totally unaware of the cached M messages as in [31] or the nth database is aware of some of M messages cached from the nth database as in [32]. Interestingly, these two models result in the same normalized download costs. Although the nth database's awareness of some cached messages might increase the download cost, at the same time the user does not need to protect the privacy of these known messages from the nth database, which might reduce the download cost.

We then compare [27], [30], and this work. The main difference between these three works is the different level of awareness of the side information the user cached. Reference [27] considers that all the databases are aware of the side information the user cached. In contrast, [30] considers that all the databases are unaware of the side information. This work considers that the nth database is aware of the side information cached from the nth database. Reference [30, Corollary 1] shows the unawareness gain. Therefore, [30] achieves lower normalized download cost than [27]. The same proof technique in [30, Corollary 1] can also show the partially unawareness gain. Therefore, this work also achieves lower normalized download cost than [27]. Since these three works consider only the privacy of the desired message, different from [31] and [32], [30] achieves lower normalized download cost than this work. For high caching ratios $\frac{1}{N} \leq r \leq 1$, the proposed scheme in this work and that in [30] share the same normalized download cost $1 - r$.

We further compare [30] and this work in the following scenario. To apply the scheme in [30], for N databases, we choose one database for prefeching and use the remaining $N - 1$ databases for retrieval. Therefore, the cached side information is completely unknown to the $N-1$ databases. We also apply the scheme in this work for comparison. For a fixed caching ratio, we compare the normalized download costs. For caching ratios $\frac{1}{N} \leq r \leq 1$, the normalized download cost is $1 - r$ for both schemes. For caching ratios $\frac{K-2}{N^2+KN-4N+1} < r < \frac{1}{N}$, we can show analytically that the normalized download cost in this work is lower than that in [30]. For caching ratios $0 < r < \frac{1}{N}$, from numerical results, we observe that the scheme in this paper achieves lower normalized

download cost. For $N = 3$ and $K = 6$, numerical results are shown in Fig. 8.

IX. Conclusion

In this paper, we studied the cache-aided PIR problem from N non-communicating and replicated databases, when the cache stores uncoded bits that are partially known to the databases. We determined inner and outer bounds for the optimal normalized download cost $D^*(r)$ as a function of the total number of messages K, the number of databases N, and the caching ratio r. Both inner and outer bounds are 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., $r \leq \frac{1}{N^{K-1}}$, and the very high caching ratio regime, where $r \geq \frac{K-2}{N^2-3N+KN}$. As a direct corollary for this result, we characterized the exact tradeoff between the download cost and the caching ratio for $K = 3$. For general K, N, and r, we showed that the largest additive gap between the achievability and the converse bounds is $\frac{5}{32}$. The achievable scheme extends the greedy scheme in [12] 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 [12] and [30] to account for the availability of uncoded and partially prefetched side information at the retriever. The converse proof hinges on developing K 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.

Appendix

Proof: From (12), we rewrite $\bar{D}(r_s)$ as

$$\bar{D}(r_s) = \frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^{i+1}}{\binom{K-2}{s-1}+\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}} \tag{72}$$

$$= \frac{\frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^{i+1}}{\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}}}{\frac{\binom{K-2}{s-1}}{\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}}+1} = \frac{\psi_1(N,K,s)}{\psi_2(N,K,s)+1}. \tag{73}$$

Let $\lambda = \frac{s}{K}$. We first upper bound $\psi_1(N,K,s)$,

$$\psi_1(N,K,s) = \frac{\sum_{i=0}^{K-1-s}\binom{K}{s+1+i}(N-1)^{i+1}}{\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}} \tag{74}$$

$$= \frac{\sum_{i=0}^{K-1-s}\frac{K}{s+1+i}\binom{K-1}{s+i}(N-1)^{i+1}}{\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}} \tag{75}$$

$$\leq \frac{\sum_{i=0}^{K-1-s}\frac{K}{s}\binom{K-1}{s+i}(N-1)^{i+1}}{\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}} = \frac{1}{\lambda}. \tag{76}$$

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

$$\frac{1}{\psi_2(N,K,s)} = \sum_{i=0}^{K-1-s}\frac{\binom{K-1}{s+i}(N-1)^{i+1}}{\binom{K-2}{s-1}} \tag{77}$$

$$= \sum_{i=0}^{K-1-s}\frac{(K-1)(K-1-s)\cdots(K-i-s)}{s(s+1)\cdots(s+i)}(N-1)^{i+1} \tag{78}$$

$$\leq (N-1)\sum_{i=0}^{K-1-s}\frac{K(K-s)^i}{s^{i+1}}(N-1)^i \tag{79}$$

$$= \frac{(N-1)}{\lambda}\sum_{i=0}^{(1-\lambda)K-1}\left(\frac{(1-\lambda)(N-1)}{\lambda}\right)^i. \tag{80}$$

When $\lambda > 1-\frac{1}{N}$, $\frac{(1-\lambda)(N-1)}{\lambda} < 1$. As $K \to \infty$, $\frac{1}{\psi_2(N,K,s)}$ is upper bounded by

$$\lim_{K\to\infty}\frac{1}{\psi_2(N,K,s)} \leq \frac{N-1}{\lambda}\sum_{i=0}^{\infty}\left(\frac{(1-\lambda)(N-1)}{\lambda}\right)^i \tag{81}$$

$$= \frac{N-1}{N\lambda-(N-1)}. \tag{82}$$

Now, we lower bound (78) by keeping the first ϵK terms in the sum for any ϵ such that $0 < \epsilon < 1-\lambda$,

$$\frac{1}{\psi_2(N,K,s)} \geq \sum_{i=0}^{\epsilon K}\frac{(K-1)(K-1-s)\cdots(K-i-s)}{s(s+1)\cdots(s+i)}(N-1)^{i+1} \tag{83}$$

$$\geq (N-1)\sum_{i=0}^{\epsilon K}\frac{(K-1)(K-\epsilon K-s)^i}{(s+\epsilon K)^{i+1}}(N-1)^i \tag{84}$$

$$= (N-1)\sum_{i=0}^{\epsilon K}\frac{(1-\frac{1}{K})((1-(\lambda+\epsilon))^i}{(\lambda+\epsilon)^{i+1}}(N-1)^i. \tag{85}$$

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

$$\lim_{K\to\infty}\frac{1}{\psi_2(N,K,s)} \geq \frac{N-1}{\lambda+\epsilon}\sum_{i=0}^{\infty}\left(\frac{(1-(\lambda+\epsilon))(N-1)}{\lambda+\epsilon}\right)^i \tag{86}$$

$$= \frac{N-1}{N(\lambda+\epsilon)-(N-1)}. \tag{87}$$

From (87) and (82), as $K \to \infty$, by picking $\epsilon \to 0$, we have

$$\psi_2(N,K,s) \to \frac{N}{N-1}\lambda - 1. \tag{88}$$

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

$$r_s \to r = \lim_{K\to\infty}\frac{\binom{K-2}{s-1}}{\binom{K-2}{s-1}+\sum_{i=0}^{K-1-s}\binom{K-1}{s+i}(N-1)^{i+1}} \tag{89}$$

$$= \lim_{K\to\infty}\frac{\psi_2(N,K,s)}{\psi_2(N,K,s)+1} \tag{90}$$

$$= \frac{N\lambda-(N-1)}{N\lambda} = 1-\left(1-\frac{1}{N}\right)\frac{1}{\lambda}. \tag{91}$$

Note that if $\lambda = 1 - \frac{1}{N}$, then $r = 0$, while if $\lambda = 1$, then $r = \frac{1}{N}$. Since we now consider the gap in the region of $0 \leq r \leq \frac{1}{N}$, without loss of generality, we consider $\lambda > 1 - \frac{1}{N}$. We express λ as

$$\lambda = \frac{1 - \frac{1}{N}}{1 - r}. \tag{92}$$

Continuing (73), by using (76), (88) and (92), we have the following upper bound on $\bar{D}(r)$

$$\bar{D}(r) \leq \frac{\frac{1}{\lambda}}{\frac{N}{N-1}\lambda} = \frac{1}{\lambda^2}\left(1 - \frac{1}{N}\right) = \frac{N}{N-1}(1-r)^2. \tag{93}$$

Now, we compare the inner bound in (14) with the outer bound derived in (93). Note that the inner bound in (14) consists of K line segments, and these K line segments intersect at the following $K-1$ points given by,

$$\tilde{r}_i = \frac{1}{N^i}, \quad i = 1, \cdots, K-1. \tag{94}$$

As i increases, $\tilde{r}_i$ concentrates to $r = 0$. Therefore, for these K line segments, we only need to consider small number of them for the worst-gap analysis. 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 $0 \leq r \leq 1$ 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, by plugging (94) into (93), we have

$$\bar{D}(\tilde{r}_i) \leq \frac{N}{N-1}\left(1 - \frac{1}{N^i}\right)^2 = \frac{1 - (\frac{1}{N})^i}{1 - \frac{1}{N}}\left(1 - \frac{1}{N^i}\right). \tag{95}$$

Furthermore, for the inner bound, we have

$$\begin{aligned}
\tilde{D}(\tilde{r}_i) &= (1 - r_i)\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^i}\right) \\
&\quad - r_i\left(1 - \frac{1}{N}\right)\left(i + \frac{(i-1)}{N} + \cdots + \frac{1}{N^{i-1}}\right) && (96) \\
&= -r_i\Bigg[\left(1 + \frac{1}{N} + \cdots + \frac{1}{N^i}\right) \\
&\quad + \left(1 - \frac{1}{N}\right)\left(i + \frac{(i-1)}{N} + \cdots + \frac{1}{N^{i-1}}\right)\Bigg] \\
&\quad + \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^i}\right) && (97) \\
&= -r_i(i+1) + \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^i}\right) && (98) \\
&= \frac{1 - (\frac{1}{N})^{i+1}}{1 - \frac{1}{N}} - r_i(i+1) = \frac{1 - (\frac{1}{N})^{i+1}}{1 - \frac{1}{N}} - \frac{i+1}{N^i} && (99)
\end{aligned}$$

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

$$\Delta(N, \infty, \tilde{r}_i) = \bar{D}(\tilde{r}_i) - \tilde{D}(\tilde{r}_i) \leq \frac{1}{N^i}\left[i - \frac{1 - (\frac{1}{N})^i}{1 - \frac{1}{N}}\right] \tag{100}$$

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

$$\Delta(N, K, r) \leq \Delta(2, \infty, r) \leq \max_i \ \frac{1}{2^i}\left[i - \frac{1 - (\frac{1}{2})^i}{1 - \frac{1}{2}}\right] \tag{101}$$

For the case $N = 2$, we note that all the inner bounds after the 7th corner point are concentrated around $r = 0$ since $\tilde{r}_i \leq \frac{1}{128}$ for $i \geq 7$. Therefore, it suffices to characterize the gap only for the first 7 corner points. Considering the 7th corner point which corresponds to $\tilde{r}_6 = \frac{1}{128}$, and $\bar{D}(r) \leq 2$ trivially for all r, and $\tilde{D}(\frac{1}{128}) = 1.9297$. Hence, $\Delta(2, \infty, r) \leq 0.07$, for $r \leq \frac{1}{127}$. Now, we focus on calculating the gap at $\tilde{r}_i$, $i = 1, \cdots, 7$. Examining all the corner points, we see that $r = \frac{1}{8}$ is the maximizing caching ratio for the gap (corresponding to $i = 3$), and $\Delta(2, \infty, \frac{1}{8}) \leq \frac{5}{32}$, which is the worst-case additive gap. ■

REFERENCES

[1] B. Chor, O. Goldreich, E. Kushilevitz, 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. Eur. Assoc. Theor. Comput. Sci.*, vol. 82, pp. 72–107, Feb. 2004.

[3] C. Cachin, S. Micali, and M. Stadler, "Computationally private information retrieval with polylogarithmic communication," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.*, Springer, 1999, pp. 402–414.

[4] R. Ostrovsky and W. E. Skeith, III, "A survey of single-database private information retrieval: Techniques and applications," in *Proc. 10th Int. Conf. Pract. Theory Public-Key Cryptogr.*, Springer, 2007, pp. 393–411.

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

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

[7] G. Fanti and K. Ramchandran, "Efficient private information retrieval over unsynchronized databases," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 7, pp. 1229–1239, Oct. 2015.

[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] A. Fazeli, A. Vardy, and E. Yaakobi, "Codes for distributed PIR with low storage overhead," in *Proc. IEEE ISIT*, Jun. 2015, pp. 2852–2856.

[10] R. Tajeddine, O. W. Gnilke, and S. El Rouayheb, "Private information retrieval from MDS coded data in distributed storage systems," in *Proc. IEEE ISIT*, Jul. 2016.

[11] H. Sun and S. A. Jafar, "The capacity of private information retrieval," in *Proc. IEEE Globecom*, Dec. 2016, pp. 1–6.

[12] 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.

[13] 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.

[14] H. Sun and S. A. Jafar. (2016). "The capacity of symmetric private information retrieval." [Online]. Available: https://arxiv.org/abs/1606.08828

[15] 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.

[16] 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.

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

[18] H. Sun and S. A. Jafar, "Multiround private information retrieval: Capacity and storage overhead," *IEEE Trans. Inf. Theory*, to be published.

[19] 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, 2017.

[20] 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.

[21] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, C. Hollanti, and S. El Rouayheb. (2017). "Private information retrieval schemes for coded data with arbitrary collusion patterns." [Online]. Available: https://arxiv.org/abs/1701.07636

[22] K. Banawan and S. Ulukus. (2017). "Multi-message private information retrieval: Capacity results and near-optimal schemes." [Online]. Available: https://arxiv.org/abs/1702.01739

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

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

[25] K. Banawan and S. Ulukus. (2017). "The capacity of private information retrieval from Byzantine and colluding databases." [Online]. Available: https://arxiv.org/abs/1706.01442

[26] Q. Wang and M. Skoglund. (2017). "Secure symmetric private information retrieval from colluding databases with adversaries." [Online]. Available: https://arxiv.org/abs/1707.02152

[27] R. Tandon. (2017). "The capacity of cache aided private information retrieval." [Online]. Available: https://arxiv.org/abs/1706.07035

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

[29] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson. (2017). "Private information retrieval with side information." [Online]. Available: https://arxiv.org/abs/1709.00112

[30] Y.-P. Wei, K. Banawan, and S. Ulukus. (2017). "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching." [Online]. Available: https://arxiv.org/abs/1709.01056

[31] Z. Chen, Z. Wang, and S. Jafar. (2017). "The capacity of private information retrieval with private side information." [Online]. Available: https://arxiv.org/abs/1709.03022

[32] Y.-P. Wei, K. Banawan, and S. Ulukus. (2017). "The capacity of private information retrieval with partially known private side information." [Online]. Available: https://arxiv.org/abs/1710.00809

[33] H. Sun and S. A. Jafar. (2017). "The capacity of private computation." [Online]. Available: https://arxiv.org/abs/1710.11098

[34] M. Mirmohseni and M. A. Maddah-Ali. (2017). "Private function retrieval." [Online]. Available: https://arxiv.org/abs/1711.04677

[35] M. Abdul-Wahid, F. Almoualem, D. Kumar, and R. Tandon. (2017). "Private information retrieval from storage constrained databases–coded caching meets PIR." [Online]. Available: https://arxiv.org/abs/1711.05244

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

[37] 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.

[38] R. Pedarsani, M. A. Maddah-Ali, and U. Niesen, "Online coded caching," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 836–845, Apr. 2016.

[39] A. Sengupta, R. Tandon, and T. C. Clancy, "Fundamental limits of caching with secure delivery," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 2, pp. 355–370, Feb. 2015.

[40] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of caching in wireless D2D networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 849–869, Feb. 2016.

[41] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4388–4413, Jul. 2017.

[42] M. Ji, A. M. Tulino, J. Llorca, and G. Caire, "Order-optimal rate of caching and coded multicasting with random demands," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3923–3949, Jun. 2017.

[43] R. Timo and M. Wigger, "Joint cache-channel coding over erasure broadcast channels," in *Proc. IEEE ISWCS*, Aug. 2015, pp. 201–205.

[44] K. Shanmugam, M. Ji, A. M. Tulino, J. Llorca, and A. G. Dimakis, "Finite-length analysis of caching-aided coded multicasting," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5524–5537, Oct. 2016.

[45] J. Zhang and P. Elia, "Fundamental limits of cache-aided wireless BC: Interplay of coded-caching and CSIT feedback," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3142–3160, May 2017.

[46] M. Gregori, J. Gómez-Vilardebó, J. Matamoros, and D. Gündüz, "Wireless content caching for small cell and D2D networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1222–1234, May 2016.

[47] M. M. Amiri and D. Gündüz, "Fundamental limits of coded caching: Improved delivery rate-cache capacity tradeoff," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 806–815, Feb. 2017.

[48] C. Tian and J. Chen, "Caching and delivery via interference elimination," in *Proc. IEEE ISIT*, Jul. 2016, pp. 830–834.

[49] K. Wan, D. Tuninetti, and P. Piantanida, "On the optimality of uncoded cache placement," in *Proc. IEEE ITW*, Sep. 2016, pp. 161–165.

[50] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," in *Proc. IEEE ISIT*, Jun. 2017, pp. 1613–1617.

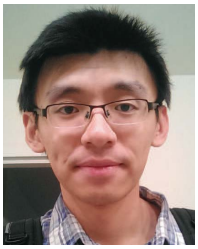
[51] A. A. Zewail and A. Yener, "Fundamental limits of secure device-to-device coded caching," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Nov. 2016, pp. 1414–1418.

[52] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr. (2017). "On the optimality of separation between caching and delivery in general cache networks." [Online]. Available: https://arxiv.org/abs/1701.05881

[53] S. S. Bidokhti, M. Wigger, and A. Yener. (2017). "Benefits of cache assignment on degraded broadcast channels." [Online]. Available: https://arxiv.org/abs/1702.08044

[54] L. Tang and A. Ramamoorthy. (2017). "Coded caching schemes with reduced subpacketization from linear block codes." [Online]. Available: https://arxiv.org/abs/1706.00101

[55] L. Xiang, D. W. K. Ng, R. Schober, and V. W. S. Wong, "Cache-enabled physical layer security for video streaming in backhaul-limited cellular networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 736–751, Feb. 2018.



Yi-Peng Wei (S'15) received the B.Sc. degree in electrical engineering from National Tsing Hua University, Taiwan, in 2009, and the M.Sc. degree from the Graduate Institute of Communication Engineering, National Taiwan University, Taiwan, in 2012. He is currently pursuing the Ph.D. degree with the Electrical and Computer Engineering Department, University of Maryland at College Park, College Park, MD, USA. His M.Sc. thesis was on the low-density graph code design. His research focuses on information theory.

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

Sennur Ulukus (S'90–M'98–SM'15–F'16) received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University and the Ph.D. degree in electrical and computer engineering from the Wireless Information Network Laboratory, Rutgers University. She was a Senior Technical Staff Member at AT&T Labs-Research. She is currently a Professor 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). Her research interests are in wireless communications, information theory, signal processing, and networks, with recent focus on private information retrieval, timely status updates over networks, energy harvesting communications, information theoretic physical layer security, and wireless energy and information transfer.

She is a Distinguished Scholar-Teacher with the University of Maryland at College Park. She received the 2003 IEEE Marconi Prize Paper Award in wireless communications, the 2005 NSF CAREER Award, the 2010–2011 ISR Outstanding Systems Engineering Faculty Award, and the 2012 ECE George Corcoran Education Award. She was a General TPC Co-Chair of the 2017 IEEE ISIT, the 2016 IEEE Globecom, the 2014 IEEE PIMRC, and the 2011 IEEE CTW. She was an Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS—Series on Green Communications and Networking (2015–2016), the IEEE TRANSACTIONS ON INFORMATION THEORY (2007–2010), and the IEEE TRANSACTIONS ON COMMUNICATIONS (2003–2007). She was a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (2015 and 2008), the *Journal of Communications and Networks* (2012), and the IEEE TRANSACTIONS ON INFORMATION THEORY (2011). She has been a Distinguished Lecturer of the Information Theory Society since 2018. She has been on the Editorial Board of the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING since 2016.