Private Information Retrieval with
Private Side Information Under Storage Constraints

Yi-Peng Wei   Sennur Ulukus

Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
ypwei@umd.edu   *ulukus@umd.edu*

Abstract—We consider the problem of private information retrieval (PIR) of a single message out of K messages from N replicated and non-colluding databases where a cache-enabled user (retriever) of cache-size S possesses side information in the form of uncoded portions of the messages that are unknown to the databases. The identities of these side information messages need to be kept private from the databases, i.e., we consider PIR with private side information (PSI). We characterize the optimal normalized download cost for this PIR-PSI problem under the storage constraint S as $D^* = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}} + \frac{1-r_M}{N^{K-M}} + \frac{1-r_{M-1}}{N^{K-M+1}} + \cdots + \frac{1-r_1}{N^{K-1}}$, where r_i is the portion of the ith side information message that is cached with $\sum_{i=1}^{M} r_i = S$. Based on this capacity result, we prove two facts: First, for a fixed memory size S and a fixed number of accessible messages M, uniform caching achieves the lowest normalized download cost, i.e., $r_i = \frac{S}{M}$, for $i = 1, \ldots, M$, is optimum. Second, for a fixed memory size S, among all possible $K - \lceil S \rceil + 1$ uniform caching schemes, the uniform caching scheme which caches $M = K$ messages achieves the lowest normalized download cost.

I. Introduction

We consider the private information retrieval (PIR) problem with private side information (PSI) for a cache-enabled user (retriever) under a cache storage size constraint. PIR refers to the problem where a user wishes to download a desired message from distributed replicated databases while keeping the identity of the desired message private against the databases. PSI refers to the setting where the user possesses cached messages in its local storage, which it wants to utilize to decrease the download cost during PIR, but at the same time, keep their identities private against the databases. The goal of the PIR-PSI problem is to devise the most efficient retrieval scheme under the joint desired message and side information privacy constraints. The efficiency of a PIR-PSI scheme is measured by the normalized download cost which is the ratio of the number of total downloaded bits to the number of desired bits.

We introduce the PIR-PSI problem under a storage constraint using the example shown in Fig. 1. Consider a user wanting to download a message from $N = 3$ non-communicating databases, each storing the same set of $K = 5$ messages. Assume that the user is already in possession of $M = 3$ messages through some unspecified means. The databases do not know the identities of these messages, but

they know that the user has access to $M = 3$ messages. (For this example, these messages are W_2, W_4 and W_5.) However, the user has limited local storage with size $S = 1$ message. What should the user keep in order to minimize the download cost of the desired message during the PIR phase? Should the user keep 1 full message in its cache, shown as caching option 1 in Fig. 1? Should the user choose 2 messages, store half of each chosen message, shown as caching option 2 in Fig. 1? Or, should the user keep all 3 messages and store a portion of each? In that case, what portions of messages should the user store? E.g., caching option 3 or caching option 4 in Fig. 1. Different caching schemes result in different download costs for the PIR-PSI problem. In this work, we characterize the optimal normalized download cost for any given caching strategy, and then determine the optimal caching strategy.

Related Work: The PIR problem has originated in the computer science community [1]–[5] and has drawn significant attention in the information theory society [6]–[10] in recent years. In the classical setting of PIR, there are N non-communicating databases, each storing the same set of K messages, and the user wishes to download one of these K messages privately. Sun and Jafar [11] 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. After [11], many interesting variants of the classical PIR problem have been investigated in [12]–[45]. The most closely related branch of PIR to our setting is cache-aided PIR in [25], [29]–[32], [35], [38].

Cache-aided PIR is first considered in [25], where the user has a local cache, and the cache content of the user is perfectly known to all the N databases. This has motivated subsequent works which have considered the case where the databases are completely unaware or partially unaware of the cache content [29]–[32], [35], [38]. Within this sub-branch of literature, references [29], [31], [32] have considered PIR with PSI. In [29], the authors considered the case where the user randomly chooses M full messages out of K messages to cache, and none of the databases is aware of the identities of the M chosen messages. The user wishes to keep the identities of the M chosen messages and the desired message private, which is coined as PIR with PSI. For the case of a single database, the optimal normalized download cost is settled in [29]. For general number of databases, the
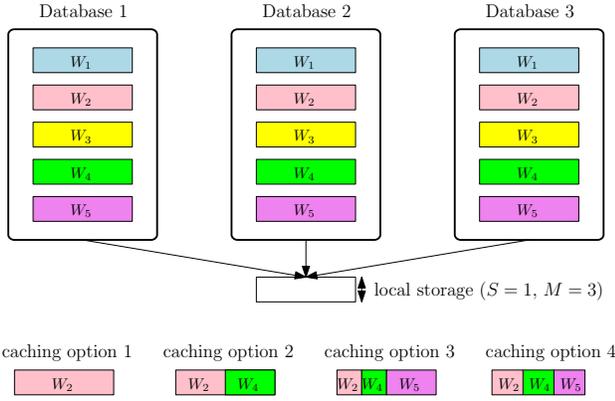
Fig. 1. PIR-PSI under a storage constraint: $N = 3$, $K = 5$, $S = 1$, and $M = 3$.

optimal normalized download cost is characterized in [31] as $D^*(M) = \left(1 + \frac{1}{N} + \cdots + \frac{1}{N^{K-1-M}}\right)$.

Coming back to our paper, in this work, we consider PIR-PSI under a storage constraint. In the prefetching phase, the user can access M messages, and has a local cache storage of S messages (SL symbols), where $S \leq M$. For each of these M messages, the user caches the first Lr_i symbols out of the total L symbols for $i = 1, \ldots, M$. Note that in [30], [35], [38], for each message, the user randomly chooses Lr symbols out of the total L symbols to cache. In [30], [35], [38], to reliably reconstruct the desired message, the user should record the indices of the cached symbols within each message, which results in extra storage overhead. In contrast, here, we consider the case where the user caches the first Lr_i symbols for $i = 1, \ldots, M$. The caching scheme is subject to a memory size constraint, i.e., $\sum_{i=1}^{M} r_i = S$. The databases are aware of the caching scheme, but do not know the identities of the cached messages, i.e., the databases know M and r_i for $i = 1, \ldots, M$, but do not know the identities of the cached messages. In the retrieval phase, the user wishes to jointly keep the identities of the cached messages and the desired message private. We call this model as PIR-PSI under a storage constraint.

For any given caching scheme, i.e., for given M and $(r_1, r_2, \ldots, r_M)$, we characterize the optimal normalized download cost to be $D^* = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}} + \frac{1-r_M}{N^{K-M}} + \frac{1-r_{M-1}}{N^{K-M+1}} + \cdots + \frac{1-r_1}{N^{K-1}}$, where without loss of generality $r_1 \geq r_2 \geq \cdots \geq r_M$. Based on this capacity result, we prove two important results: First, for a fixed memory size S and fixed number of accessible messages M, uniform caching achieves the lowest normalized download cost, where uniform caching means $r_i = \frac{S}{M}$, for $i = 1, \ldots, M$. Second, for a fixed memory size S, among all the $K - \lceil S \rceil + 1$ uniform caching schemes, the uniform caching scheme which caches K messages achieves the lowest normalized download cost. That is, in order to optimally utilize the limited user cache memory, if the user has access to M files, it should keep SL/M bits (equal amounts) from each message in its cache memory; and second, if possible, the user should aim to have access to all K messages, i.e., $M = K$ yields the lowest download cost.

II. System Model

We consider a system consisting of N non-communicating databases and a user (retriever). Each database stores the same set of K independent messages $W_1, \ldots, W_K$, and each message is of size L symbols, i.e.,

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

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

The user has a local cache memory which is of size SL symbols, where $S \in [0, K]$. There are two phases in the system: the prefetching phase and the retrieval phase.

In the prefetching phase, the user can randomly access M messages out of total K messages, where $M \geq S$. For each of the M accessed messages, the user caches the first Lr_i symbols out of the total L symbols for $i = 1, \ldots, M$. The caching scheme is subject to a memory size constraint of S, i.e., $\sum_{i=1}^{M} r_i = S$. We denote the indices (identities) of the cached M messages as $\mathbb{H}$, and denote $\mathcal{W}_\mathbb{H}$ as the cached messages. Therefore, $|\mathbb{H}| = M$, and $H(\mathcal{W}_\mathbb{H}) = SL$. Note that M and $(r_1, \ldots, r_M)$ specify a caching scheme. If $r_1 = \cdots = r_M$, we call this a uniform caching scheme. For fixed S, there are $K - \lceil S \rceil + 1$ uniform caching schemes depending on the number of accessible messages since $M \geq S$.

We assume that all the databases are aware of the caching scheme but are unaware of which messages are cached. For example, if $S = 2$, $M = 3$, and the user has applied a uniform caching scheme, the databases know that the user has chosen 3 messages out of the total K messages to cache, and for each chosen message, the user has cached the first $\frac{2}{3}L$ symbols out of the total L symbols. However, the databases do not know which messages are cached by the user.

In the retrieval phase, the user privately generates an index $\theta \in [K] = \{1, \ldots, K\}$, and wishes to retrieve message W_θ such that it is impossible for any individual database to identify θ. At the same time, the user also wishes to keep the indices of the M cached messages private, i.e., in the retrieval phase the databases cannot learn which messages are cached. For random variables θ, $\mathbb{H}$, and $W_1, \ldots, W_K$, we have

$$H(\theta, \mathbb{H}, W_1, \ldots, W_K)$$
$$= H(\theta) + H(\mathbb{H}) + H(W_1) + \cdots + H(W_K). \tag{3}$$

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

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

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

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

After receiving the answering strings $A_1^{[\theta, \mathbb{H}]}, \ldots, A_N^{[\theta, \mathbb{H}]}$ 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

$$H\left(W_\theta | \mathcal{W}_\mathbb{H}, \mathbb{H}, Q_1^{[\theta, \mathbb{H}]}, \ldots, Q_N^{[\theta, \mathbb{H}]}, A_1^{[\theta, \mathbb{H}]}, \ldots, A_N^{[\theta, \mathbb{H}]}\right) = o(L), \tag{6}$$

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

To ensure that individual databases do not know which message is retrieved and to keep the M cached messages private, we have the following privacy constraint, $\forall n \in [N]$, $\forall \theta, \theta' \in [K]$, $\forall \mathbb{H}, \mathbb{H}' \subset [K]$ such that $|\mathbb{H}| = |\mathbb{H}'| = M$,

$$\begin{aligned}
&(Q_n^{[\theta, \mathbb{H}]}, A_n^{[\theta, \mathbb{H}]}, W_1, \ldots, W_K) \\
&\sim (Q_n^{[\theta', \mathbb{H}']}, A_n^{[\theta', \mathbb{H}']}, W_1, \ldots, W_K),
\end{aligned} \tag{7}$$

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

For a fixed N, K, S and caching scheme $(r_1, \ldots, r_M)$, a pair (D, L) is achievable if there exists a PIR scheme for the message which is of size L symbols satisfying the reliability constraint (6) and the privacy constraint (7), where D represents the expected number of downloaded bits (over all the queries) from the N databases via the answering strings $A_{1:N}^{[\theta, \mathbb{H}]}$, where $A_{1:N}^{[\theta, \mathbb{H}]} = (A_1^{[\theta, \mathbb{H}]}, \ldots, A_N^{[\theta, \mathbb{H}]})$, i.e., $D = \sum_{n=1}^N H\left(A_n^{[\theta, \mathbb{H}]}\right)$. In this work, we aim at characterizing the optimal normalized download cost D^*, where

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

III. MAIN RESULTS AND DISCUSSIONS

We characterize the exact normalized download cost for PIR-PSI under a storage constraint in the following theorem.

Theorem 1 *In PIR-PSI under a storage constraint, the optimal normalized download cost is*

$$\begin{aligned}
D^* =& 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}} \\
&+ \frac{1-r_M}{N^{K-M}} + \frac{1-r_{M-1}}{N^{K-M+1}} + \cdots + \frac{1-r_1}{N^{K-1}}.
\end{aligned} \tag{9}$$

where $r_1 \geq r_2 \geq \cdots \geq r_M$ without loss of generality.

The converse proof for Theorem 1 is given in Section IV, and the achievability scheme for Theorem 1 is given in Section V. We have the following corollaries. The proofs are provided in [46].

Corollary 1 *For fixed $M \geq S$, uniform caching scheme achieves the lowest normalized download cost.*

Corollary 2 *For fixed S, among all the $K - \lceil S \rceil + 1$ uniform caching schemes, the uniform caching scheme with $M = K$ achieves the lowest normalized download cost.*

Corollary 3 *For fixed S, among all possible caching schemes, the uniform caching scheme with $M = K$ achieves the lowest normalized download cost.*

We note that, for $S = 0$, by letting $r_i = 0$, for $i = 1, \ldots, M$, (9) reduces to

$$D^* = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1}}, \tag{10}$$

which is the optimal normalized download cost of the original PIR problem as shown in [11].

We also note that, for $S \in [K]$ and $M = S$, by letting $r_i = 1$ for $i = 1, \ldots, M$, (9) reduces to

$$D^* = 1 + \frac{1}{N} + \frac{1}{N^2} + \cdots + \frac{1}{N^{K-1-M}}, \tag{11}$$

which is the optimal normalized download cost of the PIR with PSI problem as shown in [31].

IV. CONVERSE PROOF

In this section, we provide a lower bound for PIR-PSI under a storage constraint. In the following, without loss of generality, we relabel the messages according to $\mathbb{H}$, such that $W_{1:M}$ are the messages accessed by the user in the prefetching phase, where $W_{1:M} = (W_1, W_2, \ldots, W_M)$. Here, W_i denotes the message whose first Lr_i symbols are cached by the user, for $i = 1, 2, \ldots, M$, and without loss of generality, $r_1 \geq r_2 \geq \cdots \geq r_M$. We first need the following lemma, which develops a lower bound on the length of the undesired portion of the answering strings.

Lemma 1 (Interference lower bound) *For PIR-PSI under a storage constraint, the interference from undesired messages within the answering strings, $D - L$, is lower bounded by,*

$$D - L + o(L) \geq I\left(W_{1:K-1}; Q_{1:N}^{[K, \mathbb{H}]}, A_{1:N}^{[K, \mathbb{H}]} | \mathcal{W}_\mathbb{H}, \mathbb{H}, W_K\right). \tag{12}$$

The proof of Lemma 1 is provided in [46]. Note that Lemma 1 is an extension of [11, Lemma 5], where $M = 0$. In addition, we note that a similar argument to Lemma 1 can be implied from [31] and [32]. The main difference between Lemma 1 and [31], [32] is that $\mathcal{W}_\mathbb{H}$ refers to parts of messages here, while in [31], [32], $\mathcal{W}_\mathbb{H}$ refers to full messages. For the right hand side of (12), with $p(h) = \frac{1}{K \times (K-1) \times \cdots \times (K-M+1)}$,

$$\begin{aligned}
&I\left(W_{1:K-1}; Q_{1:N}^{[K, \mathbb{H}]}, A_{1:N}^{[K, \mathbb{H}]} | \mathcal{W}_\mathbb{H}, \mathbb{H}, W_K\right) \\
&= \sum_h p(h) I\left(W_{1:K-1}; Q_{1:N}^{[K, h]}, A_{1:N}^{[K, h]} | \mathcal{W}_h, W_K\right). \tag{13}
\end{aligned}$$

In the following lemma, we develop an inductive relation for the mutual information term on the right hand side of (13).

Lemma 2 (Fractional induction lemma) *For $k \in [K - 1]$, the mutual information term in (13) can be inductively lower bounded as,*

$$\begin{aligned}
&I\left(W_{1:k}; Q_{1:N}^{[k+1, h]}, A_{1:N}^{[k+1, h]} | \mathcal{W}_h, W_{k+1:K}\right) \\
&\geq \frac{1}{N} I\left(W_{1:k-1}; Q_{1:N}^{[k, h]}, A_{1:N}^{[k, h]} | \mathcal{W}_h, W_{k:K}\right) \\
&\quad + \frac{L}{N}(1 - r_k) - o(L), \tag{14}
\end{aligned}$$

where $r_k = 0$ when $k > M$.

Lemma 2 is a generalization of [11, Lemma 6] to our setting. The main difference between Lemma 2 and [11, Lemma 6] is that the cached PSI results in a different induction relation. The proof of Lemma 2 is provided in [46]. By applying Lemma 2 recursively to the right hand side of (13)

$$
I\left(W_{1:K-1}; Q_{1:N}^{[K,h]}, A_{1:N}^{[K,h]} | \mathcal{W}_h, W_K\right)
$$

$$
\overset{(14)}{\geq} \frac{1}{N} I\left(W_{1:K-2}; Q_{1:N}^{[K-1,h]}, A_{1:N}^{[K-1,h]} | \mathcal{W}_h, W_{K-1:K}\right)
$$
$$
+ \frac{L}{N} - o(L) \tag{15}
$$

$$
\overset{(14)}{\geq} \cdots \tag{16}
$$

$$
\overset{(14)}{\geq} \frac{I\left(W_{1:M}; Q_{1:N}^{[M+1,h]}, A_{1:N}^{[M+1,h]} | \mathcal{W}_h, W_{M+1:K}\right)}{N^{K-1-M}}
$$
$$
+ \frac{L}{N^{K-1-M}} + \cdots + \frac{L}{N^2} + \frac{L}{N} - o(L) \tag{17}
$$

$$
\overset{(14)}{\geq} \frac{I\left(W_{1:M-1}; Q_{1:N}^{[M,h]}, A_{1:N}^{[M,h]} | \mathcal{W}_h, W_{M:K}\right)}{N^{K-M}}
$$
$$
+ \frac{L(1-r_M)}{N^{K-M}} + \frac{L}{N^{K-1-M}} + \cdots + \frac{L}{N} - o(L) \tag{18}
$$

$$
\overset{(14)}{\geq} \cdots \tag{19}
$$

$$
\overset{(14)}{\geq} \frac{L(1-r_1)}{N^{K-1}} + \cdots + \frac{L(1-r_M)}{N^{K-M}} + \cdots + \frac{L}{N} - o(L). \tag{20}
$$

Note that in (15) to (17), we apply the fractional induction lemma with $r = 0$, since $W_{M+1:K}$ are not cached in $\mathcal{W}_h$. In (18) to (20), $r_k > 0$ for the fractional induction lemma, since $W_{1:M}$ are cached in $\mathcal{W}_h$ partially.

By combining (12), (13), and (20), and dividing by L on both sides, we obtain a lower bound for the normalized download cost as (9).

V. ACHIEVABILITY SCHEME

Our achievability scheme is based on the PIR schemes in [11] and [31]. For the portion of the messages not cached by the user, we use the PIR scheme in [11]. For the portion of the messages cached by the user, we use the PIR scheme in [31], which is based on MDS codes and consists of two stages. The first stage determines the systematic part of the MDS code according to the queries generated in [11]. In the second stage, the user reduces the download cost by downloading the parity part of the MDS code only. By applying the two PIR schemes, the user retrieves the desired message privately while keeping the cached messages private. Here, we use two examples to illustrate the main ideas of our achievability scheme. The general achievability scheme can be found in [46].

A. $N = 2$ Databases, $K = 5$ Messages, $M = 2$ Accessed Messages, and $S = 1$ with Uniform Caching

In this example, in the prefetching phase, the user randomly chooses two messages to cache, say W_1 and W_4. Since $S = 1$
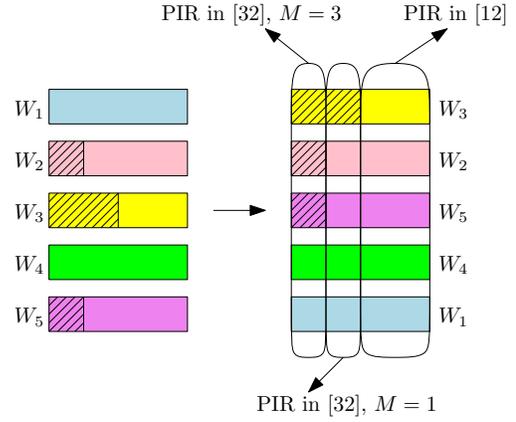

Fig. 2. $K = 5$, $S = 1$, and $M = 3$ with $r_1 = \frac{1}{2}$, and $r_2 = r_3 = \frac{1}{4}$.

and the user uses uniform caching scheme, the user caches the first half of W_1 and W_4. We note that the databases are aware of the caching scheme, i.e., the databases know that two out of five messages are chosen by the user, and the first halves of the chosen messages are cached. However, the databases do not know which are the two chosen messages.

In the retrieval phase, assume that the user wishes to retrieve message W_3. For the first half portion of the message, since the user has cached messages W_1 and W_4, the user applies the PIR scheme in [31]. The total download cost for the first half portion of the message as shown in (11) is $\frac{7}{8}L$. For the remaining half portion of the message, i.e., for the symbols in the interval $[\frac{L}{2}, L]$, since the user has not cached any messages, the user applies the PIR scheme in [11]. The total download cost for the remaining half portion of the message as shown in (10) is $\frac{31}{32}L$.

The overall normalized download cost is $\frac{59}{32}$, which can also be obtained through (9) by letting $r_1 = \frac{1}{2}$ and $r_2 = \frac{1}{2}$. Note that since we have applied the PIR scheme in [31] to retrieve the first half portion of the message, the databases cannot learn which messages are cached by the user. In addition, both PIR schemes in [11] and [31] keep the identity of the desired message private. Therefore, the combination of these two PIR schemes is a feasible PIR scheme for PIR-PSI under a storage constraint [15].

B. $N = 2$ Databases, $K = 5$ Messages, $S = 1$, $M = 3$ with $r_1 = \frac{1}{2}$, and $r_2 = r_3 = \frac{1}{4}$

In this example, see Fig. 2, in the prefetching phase, since $r_1 = \frac{1}{2}$, the user first randomly chooses one message to cache, say W_3, and the user caches the first half of W_3. Since $r_2 = r_3 = \frac{1}{4}$, the user then randomly chooses two other messages to cache, say W_2 and W_5, and the user caches the first $\frac{1}{4}$ portions of W_2 and W_5. Note that $S = 1$ and $\frac{1}{2} \times 1 + \frac{1}{4} \times 2 = 1$, and the local cache memory size constraint is satisfied. We note that the databases are aware of the caching strategy.

In the retrieval phase, assume that the user wishes to retrieve message W_1 privately. For the first $\frac{1}{4}$ portion of messages, see Fig. 2, since the user caches messages W_2, W_3 and W_5, the user applies the PIR scheme in [31] with $M = 3$. The total

download cost for the first $\frac{1}{4}$ portion of the message as shown in (11) is $\frac{3}{8}L$. For the following $\frac{1}{4}$ portion of messages, see Fig. 2, since the user caches message W_3, the user applies the PIR scheme in [31] with $M = 1$. The total download cost for the second $\frac{1}{4}$ portion of the message as shown in (11) is $\frac{15}{32}L$. For the last half portion of messages, see Fig. 2, since the user has not cached any messages, the user applies the PIR scheme in [11]. The total download cost for the last half portion of the message as shown in (10) is $\frac{31}{32}L$.

The overall normalized download cost is $\frac{29}{16}$, which can also be obtained through (9) by letting $r_1 = \frac{1}{2}$, and $r_2 = r_3 = \frac{1}{4}$. Note that the combination of the PIR schemes in [31] and [11] is a feasible PIR scheme for PIR-PSI under a storage constraint [15].

REFERENCES

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

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

[3] 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*. 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*, pages 393–411. Springer, 2007.

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

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

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

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

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

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

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

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

[13] H. Sun and S. A. Jafar. The capacity of symmetric private information retrieval. *IEEE Transactions on Information Theory*, 2018. To appear.

[14] K. Banawan and S. Ulukus. The capacity of private information retrieval from coded databases. *IEEE Transactions on Information Theory*, 64(3):1945–1956, March 2018.

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

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

[17] H. Sun and S. A. Jafar. Multiround private information retrieval: Capacity and storage overhead. *IEEE Transactions on Information Theory*, 64(8):5743 – 5754, August 2018.

[18] R. Freij-Hollanti, O. Gnilke, C. Hollanti, and D. Karpuk. Private information retrieval from coded databases with colluding servers. *SIAM Journal on Applied Algebra and Geometry*, 1(1):647–664, 2017.

[19] 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 Transactions on Information Theory*, 64(2):1000–1022, February 2018.

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

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

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

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

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

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

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

[27] R. Tajeddine and S. El Rouayheb. Robust private information retrieval on coded data. In *IEEE ISIT*, June 2017.

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

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

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

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

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

[33] Q. Wang and M. Skoglund. Secure private information retrieval from colluding databases with eavesdroppers. 2017. Available at arXiv:1710.01190.

[34] H. Sun and S. A. Jafar. The capacity of private computation. 2017. Available at arXiv:1710.11098.

[35] M. Kim, H. Yang, and J. Lee. Cache-aided private information retrieval. In *IEEE Asilomar*, October 2017.

[36] M. Mirmohseni and M. A. Maddah-Ali. Private function retrieval. 2017. Available at arXiv:1711.04677.

[37] M. Abdul-Wahid, F. Almoualem, D. Kumar, and R. Tandon. Private information retrieval from storage constrained databases – coded caching meets PIR. 2017. Available at arXiv:1711.05244.

[38] Y.-P. Wei, K. Banawan, and S. Ulukus. Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits. *IEEE Jour. on Selected Areas in Communications*. To appear. Also available at arXiv:1712.07021.

[39] S. Kumar, H.-Y. Lin, E. Rosnes, and A. G. i Amat. Achieving private information retrieval capacity in distributed storage using an arbitrary linear code. 2017. Available at arXiv:1712.03898.

[40] K. Banawan and S. Ulukus. Asymmetry hurts: Private information retrieval under asymmetric traffic constraints. 2018. Available at arXiv:1801.03079.

[41] Z. Chen, Z. Wang, and S. A. Jafar. The asymptotic capacity of private search. 2018. Available at arXiv:1801.05768.

[42] R. G.L. D'Oliveira and S. El Rouayheb. Lifting private information retrieval from two to any number of messages. *Available at arXiv:1802.06443*, 2018.

[43] K. Banawan and S. Ulukus. Private information retrieval through wiretap channel II: Privacy meets security. 2018. Available at arXiv:1801.06171.

[44] Q. Wang, H. Sun, and M. Skoglund. The capacity of private information retrieval with eavesdroppers. 2018. Available at arXiv:1804.10189.

[45] K. Banawan and S. Ulukus. Noisy private information retrieval: On separability of channel coding and information retrieval. 2018. Available at arXiv:1807.05997.

[46] Y.-P. Wei and S. Ulukus. The capacity of private information retrieval with private side information under storage constraints. 2018. Available on arXiv:1806.01253.