# Multi-Message Private Information Retrieval with Private Side Information

Seyed Pooya Shariatpanahi
School of Computer Science
Institute for Research
in Fundamental Sciences (IPM)
Email: pooya@ipm.ir

Mahdi Jafari Siavoshani
Department of Computer Engineering
Sharif University of Technology
Email: mjafari@sharif.edu

Mohammad Ali Maddah-Ali
Nokia Bell Labs
Holmdel, NJ, USA
Email: mohammad.maddahali@nokia-bell-labs.com

*Abstract*—We consider the problem of private information retrieval (PIR) where a single user with private side information aims to retrieve multiple files from a library stored (uncoded) at a number of servers. We assume the side information at the user includes a subset of files stored privately (i.e., the server does not know the indices of these files). In addition, we require that the identity of requests and side information at the user are not revealed to any of the servers. The problem involves finding the minimum load to be transmitted from the servers to the user such that the requested files can be decoded with the help of received and side information. By providing matching lower and upper bounds, for certain regimes, we characterize the minimum load imposed to all the servers (i.e., the capacity of this PIR problem). Our result shows that the capacity is the same as the capacity of a multi-message PIR problem without private side information, but with a library of reduced size. The effective size of the library is equal to the original library size minus the size of side information.

*Index Terms*—Private information retrieval, side information, MDS codes.

## I. INTRODUCTION

Private Information Retrieval (PIR) is the problem of downloading contents from a library stored at a number of servers, without revealing the indices of requested contents to any of the servers. This problem was first considered in the computer science society, from a computational complexity viewpoint [1], [2], which resulted in elegant cryptographic PIR schemes even for the simple scenario of a single user connected to a single server. Recently, in [3], the capacity of this problem has been characterized from an information theoretic perspective[1]. Here, the *capacity* refers to the supremum of the number of decoded bits per each downloaded bit. The setup considered in [3] consists of multiple non-colluding databases (servers) having access to a common library, from which a single user requests one file, via a shared link. Interestingly, as shown in [3], the user can wisely send queries to the non-cooperative servers so that the aggregate load imposed to servers is minimized, while no server gets any information about the requested file index. Following the success of [3] in characterizing the exact capacity of the problem, many works have investigated extensions of this setup, such as coded

databases [4], [5], colluding databases [6], multi-message PIR [7], adversarial PIR [8], PIR with asymmetric traffic at servers [9], and private function retrieval [10], [11].

An important extension to the PIR problem is the case when the user has access to some form of stored data. This data can be provided by caching data during network off-peak hours (known as the *Cache Content Placement* phase), to reduce the required load of servers when the actual request arrives (known as the *Content Delivery* Phase). The first work considering this line is [12], which characterizes the capacity of a public cache setup for a multi-server scenario[2]. In contrast, [13] considers a PIR problem with private cache where the servers do not know the cache contents. Moreover, [14] considers a scenario where the cache content placement is done via the same servers used in the content delivery phase, hence resulting in a partially known cache scenario, i.e., each server only knows which contents itself has sent during the cache content placement phase.

While in the cache aided PIR problem one can design the cache contents, in the PIR problem with side information it is assumed that the user has access to a given subset of the library files. Along this direction, the authors in [15] consider a single server scenario with private side information, and characterize the capacity by reducing the PIR setup to an index coding problem. The work [16] extends this result to the original multiple server setup with a user having a private side information.

In this paper, we consider a new PIR scenario, namely multi-message PIR with private side information. In this setup we assume that the user requires multiple messages and at the same time has access to a private side information, which is a subset of the files in the library. This generalized setup includes the scenarios considered in [7] (i.e., multi-message PIR without any side information) and [16] (i.e., single-message PIR with private side information) as its special cases. We propose lower and upper bounds on the required load, which match in some regimes of problem parameters, and thus, characterize the PIR capacity of those regimes. Specifically, we establish that if the user has access to a private side information of size $M$ files, then the capacity of the problem is the same

---

[1]This notion of information theoretic privacy is much stronger than the notion of cryptographic privacy used in the computer science society.

[2]By *public* we mean that all the servers know the cache contents.

as the capacity of the multi-message PIR problem (see [7]) without any side information, but with a library of size reduced by $M$. This result is a generalization of the same finding in [16] for the single-message PIR setting, and suggests the following conjecture: *The capacity of any PIR problem with private side information of size $M$ is the same as capacity of the same problem without side information, but with a library size reduced by $M$.*

The structure of paper is as follows. In Section II we describe the problem setup. In Section III our main result for the capacity of a multi-message PIR problem with private side information is stated. Sections IV and V provide the achievability and converse proofs, respectively. Finally, Section VI concludes the paper.

## II. PROBLEM SETUP

Consider a single user connected to $N$ servers having access to a library of $K$ files $\{W_1, \ldots, W_K\}$, where each file consists of $L$ symbols chosen independently and uniformly at random from a finite field $\mathbb{F}$, i.e., $W_i = (w_i(1), \ldots, w_i(L))$. We assume the user has a private side information which contains $M$ files from the library, denoted by $W_{\mathcal{S}} \triangleq \{W_i : i \in \mathcal{S}\}$, $\mathcal{S} \subseteq [1 : K], |\mathcal{S}| = M$, where the servers do not know the index set $\mathcal{S}$. The user wishes to retrieve $P$ new file $W_{\mathcal{P}} \triangleq \{W_i : i \in \mathcal{P}\}$, $\mathcal{P} \subseteq [1 : K], |\mathcal{P}| = P$, where $\mathcal{P} \cap \mathcal{S} = \varnothing$. To this end, the user sends a set of queries $Q^{\mathcal{S},\mathcal{P}} \triangleq \{Q_n^{\mathcal{S},\mathcal{P}}, n \in [1 : N]\}$, where server $n$ just receives $Q_n^{\mathcal{S},\mathcal{P}}$ without having any access to other queries (this is known as the non-colluding servers assumption). The queries should be designed such that the servers do not obtain any information about neither the requested file index set $\mathcal{P}$ nor the side information index set $\mathcal{S}$ as formally stated in the following

$$I(Q_n^{\mathcal{S},\mathcal{P}}; \mathcal{P}, \mathcal{S}) = 0, \quad \forall n \in [1 : N], \quad \textit{(privacy constraint)}.$$

After receiving the queries, each server $n \in [1 : N]$ sends the answer $A_n^{\mathcal{S},\mathcal{P}}$ which is a function of library contents $W_{1:K}$ and the query $Q_n^{\mathcal{S},\mathcal{P}}$ received at that server. The answers must be designed such that there exists a decoding function $\Psi$ which satisfies

$$\Psi(A_{[1:N]}^{\mathcal{S},\mathcal{P}}, Q_{[1:N]}^{\mathcal{S},\mathcal{P}}, W_{\mathcal{S}}) = W_{\mathcal{P}}, \quad \textit{(decodability constraint)}.$$

The objective is to characterize the minimum required download symbols defined as follows

$$D^{\mathsf{PSI}}(N, K, P, M) \triangleq \inf \frac{1}{PL} \sum_{n=1}^{N} H(A_n^{\mathcal{S},\mathcal{P}}),$$

so that the *privacy* and *decodability* constraints are satisfied, where the infimum is taken over all possible strategies. Equivalently, the PIR capacity can be defined as $C^{\mathsf{PSI}}(N, K, P, M) \triangleq \frac{1}{D^{\mathsf{PSI}}(N,K,P,M)}$.

Note that in the two special cases of $P = 1$ and $M = 0$ our setup reduces to those studied in [16] and [7], respectively.

## III. MAIN RESULT

In this paper, we characterize the capacity of a multi-message PIR problem with side information. Our main result is formally stated in Theorem 1.

**Theorem 1.** *For $P \geq \frac{K-M}{2}$ we have*

$$D^{\mathsf{PSI}}(N, K, P, M) = 1 + \frac{K - M - P}{PN}. \tag{1}$$

*Moreover, for $P \leq \frac{K-M}{2}$ and $\frac{K-M}{P} \in \mathbb{N}$ we have*

$$D^{\mathsf{PSI}}(N, K, P, M) = \frac{1 - \left(\frac{1}{N}\right)^{(K-M)/P}}{1 - \left(\frac{1}{N}\right)}. \tag{2}$$

It is worth mentioning that Theorem 1 can be equivalently stated as

$$D^{\mathsf{PSI}}(N, K, P, M) = D^{\mathsf{PSI}}(N, K - M, P, 0),$$

under the assumptions of theorem, where $D^{\mathsf{PSI}}(N, K - M, P, 0)$ is characterized in [7]. This implies that for the multi-message PIR problem, introducing private side information of size $M$ into the problem setup reduces the problem's capacity to the capacity of a multi-message PIR problem without side information but with a library of size $K - M$. Note that our result generalizes the same finding reported in [16] for the single message PIR problem with side information.

The above library size reduction effect would be trivial if the privacy constraint was not required for side information. However, we prove that the same performance can be achieved even with preserving the privacy constraint for side information. The main ingredient of the proposed achievable scheme is to use an outer layer of an MDS code to leverage the private side information available at the user.

## IV. THE ACHIEVABILITY PROOF

Let us first start with a motivating example for the case of $P \geq \frac{K-M}{2}$.

**Example 1.** *Suppose we have $N = 2$ servers, $K = 4$ files, $M = 1$ file as side information, and the user requests $P = 2$ new files. First, we review the achievable scheme proposed in [7] which is designed for the $M = 0$ case, i.e., no side information. In their scheme, each file contents is permuted independently and uniformly at random, resulting in permuted files A, B, C, and D. Moreover, it is assumed that these permutations are not known by the servers. Next, each of the files is partitioned into $N^2 = 4$ equal-sized chunks, i.e., $A = (a_1, \ldots, a_4)$, $B = (b_1, \ldots, b_4)$, $C = (c_1, \ldots, c_4)$, and $D = (d_1, \ldots, d_4)$.*

*Their scheme consists of two phases. Suppose the user requires to privately retrieve files A and B. In the first phase, each server sends a chunk from each file, i.e., Server 1 sends $a_1, b_1, c_1, d_1$ and Server 2 sends $a_2, b_2, c_2, d_2$. In the second phase, the scheme uses a Reed-Solomon generator matrix in $\mathbb{F}_q$ as follows*

$$G_{2\times3} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}.$$

The user generates two new matrices $G_1$ and $G_2$ by applying two independent random permutations on the columns of $G$. Then, the user requests from the first and second servers

$$G_1 \begin{bmatrix} a_3 & b_3 & c_2 & d_2 \end{bmatrix}^\mathsf{T} \qquad and \qquad G_2 \begin{bmatrix} a_4 & b_4 & c_1 & d_1 \end{bmatrix}^\mathsf{T},$$

respectively. Since user has received $c_2$ and $d_2$ from the second server in the first phase, it can decode $a_3$ and $b_3$ from the above two linear equations sent by the first server in the second phase. Similarly, it can decode $a_4$ and $b_4$. Thus, the user has retrieved files $A$ and $B$. This scheme results in the load of $D^{\mathsf{PSI}}(N = 2, K = 4, P = 2, M = 0) = 3/2$ per decoded file.

Now, suppose the user has access to a private side information of size $M = 1$. We use the fact that the user can construct some of the above transmissions from its side information directly. In the above, Server 1 sends 6 coded chunks where one of the file chunks in the first phase is already available as side information, which can be used to reduce the load of the first server. Since the side information should remain private to the server, we employ an MDS generator matrix $G_{5\times 6}$. Then Server 1 sends 5 linear equations of the original 6 coded chunks with coefficients obtained from the rows of $G_{5\times 6}$. By removing the term already available as side information, the user can form a linear system of 5 equations to decode the remaining chunks. The same procedure is followed by the second server. This results in the load of $D^{\mathsf{PSI}}(N = 2, K = 4, P = 2, M = 1) = 5/4$ per decoded file. Notice that as stated in Theorem 1, this scheme achieves the optimal load.

$\square$

Although the above example was stated for the case $P \geq \frac{K-M}{2}$, since the main idea of introducing the role of side information into the achievable scheme for the other case of $P \leq \frac{K-M}{2}$ is the same, we skip presenting an example for this case. In the following, we explain the general achievable scheme.

The proposed scheme is along the same line introduced in [16]. Let us first review the concept of linear PIR schemes. Suppose the user chooses $\pi_i$, $i \in [1 : K]$, independently and uniformly at random from the set of all permutations of $[1 : L]$, hidden from the servers. We define the scrambled version of file $W_i$ as $U_i = (u_i(1), \ldots, u_i(L)) = (w_i(\pi_i(1)), \ldots, w_i(\pi_i(L)))$. Then, we group these permuted symbols into chunks of $c$ symbols, i.e., $U_i = (v_i(1), \ldots, v_i(\frac{L}{c}))$ where $v_i(j) = (u_i((j-1)c + 1), \ldots, u_i(jc))$ for $j \in [1 : L/c]$, in which for the sake of presentation clarity we have assumed that $c$ divides $L$.

A $k$-sum of type $(j_1, \ldots, j_k)$ with coefficients vector $(\delta_1, \ldots, \delta_k)$ of chunks is defined as $\delta_1 v_{j_1}(i_1) + \cdots + \delta_k v_{j_k}(i_k)$ where $j_1, \ldots, j_k$ are distinct elements of $[1 : K]$, $i_1, \ldots, i_k \in [1 : L/c]$, $\delta_i \in \mathbb{F}$, and all the operations are performed element-wise in $\mathbb{F}$.

In a general linear PIR scheme, each server transmits $K$ blocks where the block $k \in [1 : K]$ consists of all possible types of $k$-sums. This symmetric structure is a consequence of the privacy requirement [3]. Moreover, each type of $k$-sums appears in $\alpha_{N,K}(k)$ distinct instances, by involving different chunks from the corresponding files in the $k$-sum. Also, each distinct instance mentioned above appears $\beta_{N,K}(k)$ times with possibly different coefficient vectors. Thus in total, each block consists of $\binom{K}{k}\alpha_{N,K}(k)\beta_{N,K}(k)$ chunks which results in the total load of

$$p(N, K) \triangleq c_{N,K} \sum_{k=1}^{K} \binom{K}{k} \alpha_{N,K}(k)\beta_{N,K}(k) \qquad (3)$$

symbols, imposed to each server. Notice that in a general achievable scheme, the chunk size $c$ depends on $N$ and $K$, hence we denote it by $c_{N,K}$. A linear PIR achievable scheme is defined to be *valid* if it satisfies both the decodability and privacy constraints, defined in Section II.

Now, for the general PIR problem with private side information, we describe an achievable scheme by employing MDS codes on top of the PIR scheme without private side information, similar to the special case proposed in [16]. In order to do this, assume $p(N, K)$ symbols are transmitted from each server in the PIR scheme without side information. Then, if the user is equipped with side information, it is clear that a subset of these symbols can be constructed directly from the side information, and thus should not be transmitted. Let us denote the number of such symbols by $q(N, K, M)$, which can be calculated as follows

$$q(N, K, M) \triangleq c_{N,K} \sum_{k=1}^{M} \binom{M}{k} \alpha_{N,K}(k)\beta_{N,K}(k). \qquad (4)$$

Now, consider an MDS code[3] $[2p - q, p]$ over the finite field $\mathbb{F}$ where each server encodes its $p$ symbols with such an MDS code. Then, instead of sending the original $p$ symbols, each server only transmits $p - q$ symbols corresponding to the non-systematic part of such a code. Accordingly, the total number of symbols, per decoded file, required to be transmitted by all servers, in the presence of private side information, is

$$D^{\mathsf{PSI}}(N, K, P, M) = \frac{N}{PL}\Big[p(N, K) - q(N, K, M)\Big].$$

It can be easily verified that the new scheme is valid if the original scheme is valid. That is because, first, the user can recover the remaining $p - q$ elements from the coded symbols and the side information due to the MDS code properties. Second, the new scheme sends the same set of queries as the original scheme which satisfy the privacy constraint. Hence, the total load of new scheme is $p - q$ symbols per server.

The above construction is also used in [16] on top of the PIR problem studied in [3]. The interesting finding of [16] is that the load of the PIR problem with private side information of size $M$ is equal to the load of a PIR problem without private side information where the library size is reduced by $M$, i.e., $K - M$.

---

[3]In the following, we may remove the dependency of functions $p$ and $q$ on $N$, $K$, and $M$ wherever it is clear from the context.

In this section, we use the general construction introduced above to develop an achievable scheme for the multi-message PIR scheme with private side information, based on the original scheme proposed in [7]. Thus, the main question to be answered is whether in this case providing the user with private side information of size $M$ will reduce the effective library size from $K$ to $K - M$.

The answer to the above question is a *Yes* if the following constraint is satisfied

$$p(N, K) - q(N, K, M) = p(N, K - M), \qquad (5)$$

where $p$ and $q$, defined in (3) and (4), are determined by the achievable scheme through the coefficients $\alpha_{N,K}(k)$, $\beta_{N,K}(k)$ and $c_{N,K}$. Thus, it just remains to check if the corresponding coefficients used in the achievable scheme proposed for the multi-message problem without private side information in [7] satisfy (5) or not. In the rest of this section we answer this question in the affirmative.

### A. Analysis of the Achievable Scheme for $P \geq \frac{K-M}{2}$

For this regime, considering the scheme introduced in [7] and without going into the details, one can verify the followings

$$\alpha_{N,K}(k) = \begin{cases} 1 & : \quad k = 1, \\ 0 & : \quad k = 2, \ldots, K-1, \\ N-1 & : \quad k = K, \end{cases}$$

$$\beta_{N,K}(k) = \begin{cases} 1 & : \quad k = 1, \\ 0 & : \quad k = 2, \ldots, K-1, \\ P & : \quad k = K, \end{cases}$$

and $c_{N,K} = \frac{L}{N^2}$. Then, we have

$$p(N, K) - q(N, K, M) = c_{N,K} \sum_{k=1}^{K} \binom{K}{k} \alpha_{N,K}(k)\beta_{N,K}(k)$$

$$- c_{N,K} \sum_{k=1}^{M} \binom{M}{k} \alpha_{N,K}(k)\beta_{N,K}(k)$$

$$= \frac{L}{N^2}\Big(K - M + P(N-1)\Big).$$

On the other hand we have

$$p(N, K - M)$$

$$= c_{N,K-M} \sum_{k=1}^{K-M} \binom{K-M}{k} \alpha_{N,K-M}(k)\beta_{N,K-M}(k)$$

$$= \frac{L}{N^2}\Big(K - M + P(N-1)\Big)$$

which confirms (5), and concludes the proof.

### B. Analysis of the Achievable Scheme for $P \leq \frac{K-M}{2}$

By inspecting the scheme introduced in [7] for the case of $P \leq \frac{K-M}{2}$, we can verify that

$$\alpha_{N,K}(k) = \sum_{i=1}^{P} \gamma_i^{N,K} r_i^{K-P-k} = r^{K-P-k} \sum_{i=1}^{P} \gamma_i^{N,K},$$

where according to [7] we have $r_i = r$ and $r$ satisfies $(1 + \frac{1}{r})^K = N^{K/P}$. Here, $(\gamma_1^{N,K}, \ldots, \gamma_P^{N,K})$ is the solution of the linear equation [7, Eq. (64)]. Notice that we do not need the exact values of $\gamma_i^{N,K}$'s since they will cancel out later from our computations. Moreover, we have $\beta_{N,K}(k) = 1$ in this regime.

The constant $c_{N,K}$ of the scheme in [7] can be derived as follows. Each $k$-sum transmitted by each server can be one of the two following types. The first type includes those sums containing at least some chunks from the requested files and maybe some chunks from the not-requested files. The second type only consists of sums containing not-requested file chunks, which are being transmitted due to the privacy requirements. For the decodability constraint we require that the total number of useful equations (equations containing at least one chunk from the $P$ requested files) transmitted by all the servers should be equal to the total number of symbols required to recover the $P$ requested files (i.e., $PL$ symbols). This will result in

$$c_{N,K} = \frac{1}{N} \frac{PL}{\sum_{k=1}^{K} \binom{K}{k} \alpha_{N,K}(k) - \sum_{k=1}^{K-P} \binom{K-P}{k} \alpha_{N,K}(k)}.$$

Hence, to verify (5) we can proceed as follows

$$p(N, K) - q(N, K, M)$$

$$= \frac{PL}{N} \frac{\sum_{k=1}^{K} \binom{K}{k} \alpha_{N,K}(k) - \sum_{k=1}^{M} \binom{M}{k} \alpha_{N,K}(k)}{\sum_{k=1}^{K} \binom{K}{k} \alpha_{N,K}(k) - \sum_{k=1}^{K-P} \binom{K-P}{k} \alpha_{N,K}(k)}$$

$$= \frac{PL}{N} \frac{\sum_{k=1}^{K} \binom{K}{k} r^{K-P-k} - \sum_{k=1}^{M} \binom{M}{k} r^{K-P-k}}{\sum_{k=1}^{K} \binom{K}{k} r^{K-P-k} - \sum_{k=1}^{K-P} \binom{K-P}{k} r^{K-P-k}}$$

$$= \frac{PL}{N} \frac{\sum_{k=1}^{K} \binom{K}{k} r^{-k} - \sum_{k=1}^{M} \binom{M}{k} r^{-k}}{\sum_{k=1}^{K} \binom{K}{k} r^{-k} - \sum_{k=1}^{K-P} \binom{K-P}{k} r^{-k}}$$

$$= \frac{PL}{N} \frac{(1 + r^{-1})^K - (1 + r^{-1})^M}{(1 + r^{-1})^K - (1 + r^{-1})^{K-P}}$$

$$= \frac{PL}{N} \frac{N^{K/P} - N^{M/P}}{N^{K/P} - N^{(K-P)/P}}$$

$$= \frac{PL}{N} \frac{1 - \left(\frac{1}{N}\right)^{\frac{K-M}{P}}}{1 - \frac{1}{N}}$$

$$= \frac{PL}{N} \cdot D^{\mathsf{PSI}}(N, K - M, P, 0).$$

which is equal to $p(N, K - M)$ according to [7]. This concludes the proof.

## V. THE CONVERSE PROOF

### A. The Converse Argument for $P \geq \frac{K-M}{2}$

To prove the converse for this case, let us proceed as follows. Without loss of generality we assume $\mathcal{S} = \{1, \ldots, M\}$, $\mathcal{P} = \{M+1, \ldots, M+P+1\}$. Then, we can state the following lemma.

**Lemma 1.** *Under the assumptions of Theorem 1, we have*

$$(K - M)L \leq \sum_{n=1}^{N} H(A_n^{\mathcal{S},\mathcal{P}} | Q, W_{\mathcal{S}}) + (K - M - P)L$$

$$- H(A_1|Q, W_{\mathcal{S}}, W_{\mathcal{P}}). \tag{6}$$

*Proof.* The proof is provided in [17]. $\qquad\square$

Now, we can lower bound $D^{\mathsf{PSI}}$ as follows

$$
\begin{aligned}
D^{\mathsf{PSI}} &= \frac{1}{PL} \sum_{n=1}^{N} H(A_n^{\mathcal{S},\mathcal{P}}) \\
&\geq \frac{1}{PL} \sum_{n=1}^{N} H(A_n^{\mathcal{S},\mathcal{P}}|Q, W_{\mathcal{S}}) \\
&\overset{(a)}{\geq} \frac{1}{PL} \Big( PL + H(A_1|Q, W_{\mathcal{S}}, W\mathcal{P}) \Big) \\
&\overset{(b)}{\geq} 1 + \frac{K - M - P}{NP}
\end{aligned}
$$

where (a) follows from Lemma 1 and (b) holds due the following lemma, Lemma 2. This concludes the proof.

**Lemma 2.** *If $P \geq \frac{K-M}{2}$, then we have*

$$H(A_1|Q, W_{\mathcal{S}}, W_{\mathcal{P}}) \geq \frac{(K - M - P)L}{N}.$$

*Proof.* The proof is provided in [17]. $\qquad\square$

*B. The Converse Argument for $P \leq \frac{K-M}{2}$*

To prove the converse for this regime, we can state the following lemma.

**Lemma 3.** *Under the assumption of Theorem 1, we have*

$$
\begin{aligned}
(K - M)L &\leq NH(A_1|Q, W_{\mathcal{S}}) \\
&\quad + (N - 1)[NH(A_1|Q, W_{\mathcal{S}}) - PL] \\
&\quad + (K - M - 2P)L \\
&\quad - H(A_1|W_{[M+1:M+2P]}, Q, W_{\mathcal{S}}).
\end{aligned}
$$

*Proof.* The proof is provided in [17]. $\qquad\square$

The result stated in Lemma 3 can be equivalently written as

$$
\begin{aligned}
NH(A_1|Q, W_{\mathcal{S}}) &\geq \left(1 + \frac{1}{N}\right) PL \\
&\quad + \frac{1}{N} H(A_1|W_{[M+1:M+2P]}, Q, W_{\mathcal{S}}). \tag{7}
\end{aligned}
$$

Equation (7) provides a lower bound on $H(A_1|Q, W_{\mathcal{S}})$. Applying the same chain of inequalities used to prove Lemma 3, one can derive a similar inequality for $H(A_1|W_{[M+1:M+2P]}, Q, W_{\mathcal{S}})$. Hence, by applying (7) inductively over the library size, one can derive a lower bound for $H(A_1|Q, W_{\mathcal{S}})$, similar to [7, Equation (125)]. The important difference of above calculations with those appeared in [7] is that here we start the induction from the message $W_{M+1}$ while the starting point of [7] is the message $W_1$.

Similar to the path followed from [7, Equation (124)] to [7, Equation (133)], we arrive at

$$
D^{\mathsf{PSI}} \geq \left( \frac{1 - \left(\frac{1}{N}\right)^{\left\lfloor \frac{K-M}{P} \right\rfloor}}{1 - \frac{1}{N}} + \frac{\left(\frac{K-M}{P} - \left\lfloor \frac{K-M}{P} \right\rfloor\right)}{N^{\left\lfloor \frac{K-M}{P} \right\rfloor}} \right)
$$

which completes the proof for the case of $P \leq \frac{K-M}{2}$ and $\frac{K-M}{P} \in \mathbb{N}$.

## VI. Conclusions

In this paper, we have characterized the capacity of a multi-message PIR problem where the user has access to a private side information, for certain regimes. Our result shows that the role of private side information is equivalent to reducing the effective library size by the size of side information. This result, along with a similar conclusion for a single-message PIR setup in [16], motivate the conjecture that for any PIR scheme, adding a private side information will be equivalent to reducing the effective library size.

## References

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

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

[3] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4075–4088, July 2017.

[4] R. Tajeddine, O. W. Gnilke, and S. E. Rouayheb, "Private information retrieval from mds coded data in distributed storage systems," *IEEE Transactions on Information Theory*, pp. 1–1, 2018.

[5] K. Banawan and S. Ulukus, "The capacity of private information retrieval from coded databases," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1945–1956, March 2018.

[6] H. Sun and S. A. Jafar, "The capacity of robust private information retrieval with colluding databases," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2361–2370, April 2018.

[7] K. A. Banawan and S. Ulukus, "Multi-message private information retrieval: Capacity results and near-optimal schemes," *CoRR*, vol. abs/1702.01739, 2017. [Online]. Available: http://arxiv.org/abs/1702.01739

[8] Q. Wang and M. Skoglund, "Secure symmetric private information retrieval from colluding databases with adversaries," *CoRR*, vol. abs/1707.02152, 2017. [Online]. Available: http://arxiv.org/abs/1707.02152

[9] K. A. Banawan and S. Ulukus, "Asymmetry hurts: Private information retrieval under asymmetric traffic constraints," *CoRR*, vol. abs/1801.03079, 2018. [Online]. Available: http://arxiv.org/abs/1801.03079

[10] H. Sun and S. A. Jafar, "The capacity of private computation," *CoRR*, vol. abs/1710.11098, 2017. [Online]. Available: http://arxiv.org/abs/1710.11098

[11] M. Mirmohseni and M. A. Maddah-Ali, "Private function retrieval," *CoRR*, vol. abs/1711.04677, 2017. [Online]. Available: http://arxiv.org/abs/1711.04677

[12] R. Tandon, "The capacity of cache aided private information retrieval," *CoRR*, vol. abs/1706.07035, 2017. [Online]. Available: http://arxiv.org/abs/1706.07035

[13] Y. Wei, K. A. Banawan, and S. Ulukus, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *CoRR*, vol. abs/1709.01056, 2017. [Online]. Available: http://arxiv.org/abs/1709.01056

[14] ——, "Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits," *CoRR*, vol. abs/1712.07021, 2017. [Online]. Available: http://arxiv.org/abs/1712.07021

[15] S. Kadhe, B. Garcia, A. Heidarzadeh, S. Y. E. Rouayheb, and A. Sprintson, "Private information retrieval with side information," *CoRR*, vol. abs/1709.00112, 2017. [Online]. Available: http://arxiv.org/abs/1709.00112

[16] Z. Chen, Z. Wang, and S. Jafar, "The capacity of private information retrieval with private side information," *CoRR*, vol. abs/1709.03022, 2017. [Online]. Available: http://arxiv.org/abs/1709.03022

[17] S. Pooya Shariatpanahi, M. Jafari Siavoshani, and M. A. Maddah-Ali, "Multi-Message Private Information Retrieval with Private Side Information," *CoRR*, vol. abs/1805.11892, May 2018. [Online]. Available: http://arxiv.org/abs/1805.11892