Converse of Private Information Retrieval

April 16, 2020

Chapter 1

Introduction of PIR

We begin the chapter with the background of private information retrieval, basic algebra codes and retrieval schemes, then we introduce the basic procedures of a capacity of PIR problem, and we list some possible applications with PIR, such as locally decodable codes, cache coding and topological interference alignment.

1.1 Background of Private Information Retrieval

Marked by paradigm-shifting developments such as big data, cloud computing, and internet of things, the modern information age presents researchers with an unconventional set of challenges. The rapidly evolving research landscape continues to blur traditional boundaries between computer science, communication and information theory, coding and signal processing. For example, the index coding problem which was introduced by computer scientists in 1998, is now a very active research topic in information theory because of its fundamental connections to a broad range of questions that includes topological interference management, network coding, distributed storage capacity, hat guessing, and non-Shannon information inequalities. Evidently, the crossover of problems across fields creates exciting opportunities for fundamental progress through a consolidation of complementary perspectives. The pursuit of such crossovers brings us to the private information retrieval (PIR) problem.

Introduced in 1995 by Chor et al., the private information retrieval (PIR) problem seeks the most efficient way for a user to retrieve a desired message from a set of distributed databases, each of which stores all the messages, without revealing any information about which message is being retrieved to any individual database. The user can hide his interests trivially by requesting all the information, but that could be very inefficient (expensive). The goal of the PIR problem is to find the most efficient solution.

Besides its direct applications, PIR is of broad interest because it shares intimate connections to many other prominent problems. PIR attracted our attention initially in because of its curious similarities to Blind Interference

Alignment. PIR protocols are the essential ingredients of oblivious transfer, instance hiding, multiparty computation, secret sharing schemes and locally decodable codes. Through the connection between locally decodable and locally recoverable codes, PIR also connects to distributed data storage repair, index coding and the entire umbrella of network coding in general. As such PIR holds tremendous promise as a point of convergence of complementary perspectives. The characterization of the information theoretic capacity of PIR that we undertake in this work, is a step in this direction.

1.2 Problem Statement

The term *"Retrieval"* implies the process which one user sends queries $\mathcal{Q}$ based on the desired symbol index k (out of total K messages) to get W_k from N server replies $\mathcal{A}$, where each server contains full or part of data from database $W_{[K]}$. It can be formulated as:

Definition 1 *(Retrieval)A standard (N,K,k)-retrieval is an combination of an algorithm set $\{Q, A, L\}$ and a random vector U, constrained by:*

$$Q^k_{\mathcal{N}} = \{Q^k_n | n \in \mathcal{N}\}, \quad A^k_{\mathcal{N}} = \{A^k_n | n \in \mathcal{N}\}, \quad k \in [K]; \tag{1.1}$$

$$\mathcal{Q} = Q^k_{[N]} | k \in [K], \quad \mathcal{A} = A^k_{[N]} | k \in [K]; \tag{1.2}$$

$$Q^k_{[N]} = Q(U, k); \tag{1.3}$$

$$A^k_n = A(Q^k_n, W_{[K]}), \quad k \in [K]; \tag{1.4}$$

$$W_k = L(A^k_{[N]}, U, k), \quad \frac{dL}{dU} = 0; \tag{1.5}$$

We define a Private Information Retrieval (abbreviated as PIR) scheme as follows:

Definition 2 *(PIR Scheme)We call a retrieval set $\mathcal{P}$ a (N,K)-PIR scheme if*

$$\mathcal{P} = \{(N, K, k) | \forall k_1, k_2 \in [K], \forall n \in [N], (Q^{k_1}_n, A^{k_1}_n, W_{[K]}) \sim (Q^{k_2}_n, A^{k_2}_n, W_{[K]})\} \tag{1.6}$$

for any (N,K,k)-retrieval form $\mathcal{P}(N, K)$, we call the retrieval a k-realization of $\mathcal{P}$.

Lemma 1 *A (N,K)-PIR scheme is equivalent to the variable ($\mathcal{A}$,$\mathcal{Q}$,$W_{[K]}$), as long as the (Q,A,U,L) satisfies:*

$$(QueryConstraint, QC) : \forall k \in [K], \quad I(W_{[K]}; Q^{[k]}_{[N]}) = 0; \tag{1.7}$$

$$(AnswerConstraint, AC) : \forall k \in [K], \quad \forall n \in [N], H(A^{[k]}_n | Q^{[k]}_n, W_{[K]}) = 0; \tag{1.8}$$

$$(CorrectnessConstraint, CC) : H(W_k | A^{[k]}_{[N]}, Q^{[k]}_{[N]}) = 0; \tag{1.9}$$

$$(PrivacyConstraint, PC) : I(k; A^k_n, Q^k_n, W_{[K]}) = 0; \tag{1.10}$$

By applying information constraints, we transform the problem to the relationship of above 3 main variables.

Definition 3 *We call a (N,K)-PIR R-achieved, if the scheme satisfies:*

$$R = \frac{H(W_k)}{\sum_{n=1}^{N} H(A_n^k|Q_n^k)} \tag{1.11}$$

we note

$$C = sup\{R|\exists(N,K) - PIR \quad Scheme \quad is \quad R - achieved.\} \tag{1.12}$$

as the (N,K)-PIR Capacity.

Chapter 2

PIR Capacity

For classical PIR which all servers contain the same replicated database, we need to add data constraints(DC) to , i.e.,

Definition 4 *A (N,K)-PIR scheme is Constrained by:*

$$(DC): H(W_{[K]}) = H(W_1) + \cdots + H(W_K), H(W_1) = \cdots = H(W_K) = L. \tag{2.1}$$

$$(QC): \forall k \in [K], \quad I(W_{[K]}; Q_{[N]}^{[k]}) = 0; \tag{2.2}$$

$$(AC): \forall k \in [K], \quad \forall n \in [N], H(A_n^{[k]} | Q_n^{[k]}, W_{[K]}) = 0; \tag{2.3}$$

$$(CC): H(W_k | A_{[N]}^{[k]}, Q_{[N]}^{[k]}) = 0; \tag{2.4}$$

$$(PC): I(k; A_n^k, Q_n^k, W_{[K]}) = 0; \tag{2.5}$$

2.1 Achievability

left Blank.

2.2 Converse

Lemma 2 *(QC+AC):for $\forall \mathcal{N} \in \mathcal{B}([N])$, $\forall \mathcal{K} \in \mathcal{B}([K])$:*

$$H(A_{\mathcal{N}}^{[k]} | \mathcal{Q}, W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}) = H(A_{\mathcal{N}}^{[k]} | W_{\mathcal{K}}, Q_{\mathcal{N}}^{[k]}); \tag{2.6}$$

Proof:

$$
\begin{aligned}
&I(A_{\mathcal{N}}^{[k]};\mathcal{Q}|W_{\mathcal{K}},Q_{\mathcal{N}}^{[k]}) && (2.7)\\
&\quad\le I(A_{\mathcal{N}}^{[k]},W_{[1:K]};\mathcal{Q}|W_{\mathcal{K}},Q_{\mathcal{N}}^{[k]}) && (2.8)\\
&\quad= I(W_{[1:K]};\mathcal{Q}|W_{\mathcal{K}},Q_{\mathcal{N}}^{[k]}) + \ I(A_{\mathcal{N}}^{[k]};\mathcal{Q}|W_{[1:K]},W_{\mathcal{K}},Q_{\mathcal{N}}^{[k]}) && (2.9)\\
&\quad\overset{(AC)}{=} I(W_{[1:K]};\mathcal{Q}|W_{\mathcal{K}},Q_{\mathcal{N}}^{[k]}) && (2.10)\\
&\quad= H(W_{[1:K]}|W_{\mathcal{K}},Q_{\mathcal{N}}^{[k]}) - H(W_{[1:K]}|W_{\mathcal{K}},\mathcal{Q}) && (2.11)\\
&\quad\overset{(QC)}{\le}\ H(W_{[1:K]\setminus\mathcal{K}}) - H(W_{[1:K]\setminus\mathcal{K}}|\mathcal{Q}) && (2.12)\\
&\quad= I(W_{[1:K]\setminus\mathcal{K}};\mathcal{Q}) && (2.13)\\
&\quad\overset{(QC)}{=} 0, && (2.14)
\end{aligned}
$$