Private Information Retrieval

Soheil Nematihaji

Department of Computer Science, University of Virginia

This document presents a survey on PIR[1]

Introduction

Consider the following scenario. Alice wants to obtain information from a database but does not want the database to learn which information she wanted. One solution is for Alice to ask for the entire database. Can she obtain what she wants with less communication? We can extend this problem to something more general. Consider the same scenario that Alice wants to obtain her information from K different database which have the same information. And we can assume that there are k non-communicating copies of the database.Also We assume that the databases have unlimited computational power or we can assume that they are computationally bounded by $Poly(n)$.

Definitions and theorems on PIR

We need some definition to starting our survey.

Definition 1. We model a database as an $n - bit$ string $x = x_1, x_2, \cdots, x_n$ together with a computational agent[2] that can do computations based on both x and queries made to it.

[1] Private Information Retrieval

[2] Consider agent as a Turing Machine

Definition 2. A $1 - round\ k - DB$ Information Retrieval Scheme With $x \in \{0,1\}^n$ and k databases has the following form:

1. Alice wants to know x_i. There are k copies of the database which all have $x = x_1, x_2, \cdots, x_n$. The DB's do not communicate with each other.

2. Alice flips coins and, based on the coin flips and i, computes (query) strings $q_1, q_2, \cdots, q_n$. Alice sends q_j to database DB_j

3. For all $j\ 1 \leqslant j \leqslant k, DB_j$sends back a (answer) string $ANS_j\ (q_j)$.

4. Using the value of i, the coin flips, and the $ANS_j(q_j)$, Alice computes x_i

The complexity of the above PIR scheme is $\sum\limits_{j=1}^{k} |qj| + |ANS_j(q_j)|$.

Now we can define the privacy. We can consider two type of privacy depend on the DB complexity.so for each computationally unbounded or computationally bounded DB we can define privacy:

Definition 3. A $S - round\ k - DB$Private Information Retrieval Scheme with$x \in \{0,1\}^n$ and k Databases is an information retrieval scheme such that, after the query is made and answered, the database does not have any information about what i is.actually they should not can distinguish between transcript of i and j The database is assumed to be computationally unbounded. Hence we need to ensure that the database does not have enough information to figure out anything about i . For these PIR schemes we will need multiple copies of the database.

Definition 4. A $S - round\ k - DB$ Computationally Private Information Retrieval Scheme with $x \in \{0,1\}^n$ and k databases is an information retrieval scheme such that, assuming some limitations on what the database can compute, after the query is made and answered, the database does not have any information about what i is. Hence we need to ensure that computing anything about i is beyond the computational limits of the database.

in the rest of the survey Our focus is on $1 - round$.

notation 1. If σ is a string and $i \leqslant |\sigma|$ then $\sigma \oplus i$ is the string with the ith bit flipped.

Theorem 1. there is a $4 - DB$, $O(\sqrt{n})$-bit PIR scheme.

Proof. Each index of the database is represented as an ordered pair (i_1, i_2) , where i_1 and i_2 are written in base $\lceil\sqrt{n}\rceil$. The databases are labeled DB_{00} , DB_{01} , DB_{10} and DB_{11}.

1. Alice wants to know bit $x_{(i_1,i_2)}$.

2. Alice generates σ , $\tau \in \{0 , 1\}^{\sqrt{n}}$.

3. Alice then generates two additional $\sqrt{n}$ bits strings from the first two strings: $\sigma' = \sigma \oplus i_1$ and$\tau' = \tau \oplus i_2$.

4. Alice sends two strings to each database. DB_{00}receives σ, τ . DB_{01} receives σandτ' . DB_{10}receives σ'andτ . DB_{11} receives σ'andτ'.

5.D_{00} sends $\oplus_{\sigma(j_1)=1,\tau(j_2)=1}x_{j1,j2}$. D_{01} sends $\oplus_{\sigma(j_1)=1,\tau'(j_2)=1}x_{j1,j2}$.$D_{10}$ sends $\oplus_{\sigma'(j_1)=1,\tau(j_2)=1}x_{j1,j2}$. D_{11} sends $\oplus_{\sigma'(j_1)=1,\tau'(j_2)=1}x_{j1,j2}$.

6. Alice XORs the four bits. Since $x_{i1,i2}$ is the only bit that appeared an odd number of times, the result is $x_{i1,i2}$.

Note that the number of bits sent is $8\sqrt{n} + 4$.

if we want to explain why this protocol is secure for Alice we should calculate what is the advantage of each DB to guess i correctly.consider that we have to transcript for i and j. if the database doesnt communicate with each other then what is the probability that they can distinguish between two transcript.now we can prove what we want and actually we can define what exactly the privacy is.first we define privacy mathematically then we prove our claim.so we need to define our scheme again with perfect privacy.

Definition 5. (Private Information Retrieval one-round schemes): consider$k \geq 2$ a k - server Private Information Retrieval (PIR) scheme for database length n consists of:

k query functions , $Q_1, \cdots, Q_k : [n] \times \{0,1\}^{l_r nd} \to \{0,1\}^q$.

k answer functions , $A_1, \cdots, A_k : \{0,1\}^n \times \{0,1\}^{l_q} \to \{0,1\}^{l_a}$.

Privacy: For every$i, j \in [n], s \in [k]$, and $q \in \{0,1\}^{l_q}$,

$$Pr(Q_s(i,r) = q) = Pr(Q_s(j,r) = q)$$

now back to our problem.in theorem 2 we want to show that this protocol has privacy for Alice.for prove the privacy we should show that the distribution of two transcript for tow different index is equal, for doing this consider two case :
case1: $DB_{0,0}$ and index i
case1: $DB_{0,0}$ and index j the distribution of σ and τ do not depend on i and j.so the distributions of them are equal.we have the same thing for $DB_{i,j}$ because the σ and τ chose at random.
Theorem 2. For all $k \in N$ there is a k-DB, $o((k \log k)n^{\frac{1}{\log k}})$ -bit PIR scheme.
the proof of this theorem is like theorem 1. we need to represent the DB in the base $n^{\frac{1}{\log k}}$. and the rest of the proof is the same as proof of theorem 1.
we can extend this idea and prove better bound for PIR.
Theorem 3. For all $k \in N$ there is a k-DB, $o((k \log k)n^{\frac{1}{\log k + \log \log k}})$ -bit PIR scheme.
Theorem 4. Assume there are k vectors in $\{0,1\}^d$ that cover $\{0,1\}^d$. Then there is a k-DB, $O(n^{\frac{1}{d}})$ - bit PIR scheme.
The theorem 4 implies theorem 3.
Theorem 5. For all k there is a k-DB $O(2^{k^2}(n^{\frac{1}{2k-1}}))$ -bit scheme.
Theorem 6. For all k there is a k-DB $O(k^3(n^{\frac{1}{2k-1}}))$ -bit scheme.
Theorem 7. One-way Functions Imply O (n^ϵ) 2-DB PIRs

Single-Database PIR

Single-database PIR has a close connection to the notion of Oblivious Transfer (OT).Informally, 1-out-of-n OT is a protocol for two players: A sender who initially has n secrets $x_1, \cdots, x_n$and a receiver who initially holds an index $1 \leqslant i \leqslant n$. At the end of the protocol the receiver knows x_i but has no information about the other secrets, while the sender has no information about the index i . Note that OT is different from PIR in that there is no communication complexity requirement but, on the other hand, secrecy is required for both players, while for PIR it is required only for the user.and there is a connection between PIR and collision-resistant. Any one-round Single-Database PIR protocol is also a collision-resistant hash function.and if there is a 1-DB ($\frac{n}{2}$) -bit PIR scheme then there is a weak bit- commitment scheme.

Theorem 8. Assume that the quadratic residue problem is 'hard' for m the product of two primes and $|m| \geqslant n^{\delta}$. Then there exists a 1-DB, $O(n^{\frac{1}{2}+\delta})$ -bit PIR scheme.

proof.we represent the DB as a $\sqrt{n}\times\sqrt{n}$ Array. Alice wants bit $x_{i,j}$. Alice generates two primes p_1, p_2 of the same length such that $m = p_1p_2$ has length n. Alice generates $\sqrt{n}$ elements of Z_m^* which we call $r_1, \cdots, r_{\sqrt{n}}$. Alice makes sure that all of them are quadratic residues except r_i . Make sure that r_i has Jacobi symbol 1 (i.e., it is a non-square modulo both p_1 and p_2). Alice sends $m, r_1, ..., r\sqrt{n}4$ to the database. Note that this takes$O(n\sqrt{n}$). The database computes the following matrix. $c_{a,b} = z_b^2$ if $x_{a,b} = 1$,$c_{a,b} = z_b$ if $x_{a,b} = 0$. The database computes the products of the rows. The database sends over$r_1, \cdots, r_{\sqrt{n}}$. 8. Alice sees if r_j is a QR. If it is then $x_{i,j} = 1$, otherwise $x_{i,j} = 0$.

The following summarizes what is known about assumptions for sublinear 1-DB PIR:

TDP ⇒1-DB $(n - o(n))$ -bit PIR

($n - o(n)$) -bit PIR ⇒ OT

OT ⇒ One-Way function

One-Way function ⇒ 2-DB $n - o(n)$

HES[3] ⇒1-DB(n^{ϵ})-bit PIR

[3] Homomorphic Encryption Scheme