Private Information Retrieval: An Introduction

Mostly Yuval Ishai- CS Dept, Technion

Some Slides- Eyal Kushilevitz- CS Dept, Technion

Edited by William Gasarch- CS Dept, Univ of MD.

Private Information Retrieval (PIR)

[CGKS95]

- Goal: allow a user to access a database while hiding what she is after.
- Motivation: patent databases, stock tips, web searches, etc.

Modeling

- Database: n-bit string x

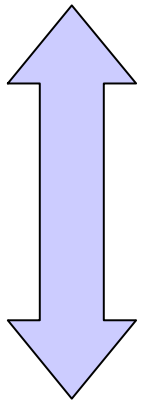
- User: wishes to
  - retrieve x_i and
  - keep i private

Modeling

$x \in \{0,1\}^n$

Server

???



x_i

User

$i \in \{1,\ldots,n\}$

Some “solutions”

1. User downloads entire database.
   Drawback: n communication bits.
   Main research goal: minimize *communication complexity*.

2. User masks i with additional random indices.
   Drawback: gives a lot of information about i.

Two Approaches

Computational PIR [KO97,CMS99,...]

Computational privacy, based on cryptographic assumptions. E.g. assume database cannot compute quadratic residues.

Information-Theoretic PIR [CGKS95, AMB97,]

Replicate database among k servers

Bounds for Computational PIR

	servers	comm.	assumption
[CG97]	2	$O(n^{\varepsilon})$	one-way function
[KO97]	1	$O(n^{\varepsilon})$	QRA / homomorphic encryption
[CMS99]	1	polylog(n)	Φ-hiding
[KO00]	1	n-$o(n)$	trapdoor permutation
[L05]	1	$O(\log^2 n)$	public key
[GR05]	1	$O(\log n)$	Φ-hiding

Bounds for I.T. PIR

Upper bounds:

- $O(\log n / \log\log n)$ servers, $\mathrm{polylog}(n)$ [BF90,BFKR91,CGKS95]
- 2 servers, $O(n^{1/3})$; k servers, $O(n^{1/k})$ [CGKS95]
- k servers, $O(n^{1/(2k-1)})$ [Amb97,Ito99, IK99, BI01,WY05]
- k servers, $O(n^{c\log\log k /(k\log k)})$ [BIKR02].

Lower bounds:

- $\log n + 1$ (no privacy)
- 2 servers, $\sim 5\log n$; k servers, $c_k \log n$ [Man98,WdW04]
- $\Omega(n^{1/3})$ [RY06]

Why *Information-Theoretic* PIR?

Cons:

- Requires multiple servers
- Privacy against limited collusions
- Worse asymptotic complexity (with const. k):

  $O(n^c)$ vs. polylog(n)

Pros:

- Neat question
- Unconditional privacy
- Better "real-life" efficiency
- Allows very short queries or very short answers (+apps [DIO98,BIM99]

Open Questions: Assumptions

- Sufficient assumptions for 1-server PIR
  - OT → nontrivial PIR ?
    - Known: PIR → OT
  - Trapdoor permutation → *good* PIR?
    - Known: TDP → *n*-*o*(*n*) comm.
  - Your favorite assumption → *great* PIR?
    - Known: Phi-hiding, CRA

Open Questions: I.T. PIR

- Better upper bounds
  - Known: $O(n^{c\log\log k/(k\log k)})$
- Better lower bounds
  - Known: $c\log n$
  - Simplest cases:
    - k=2, $\Omega(n^{1/3})$
    - k=3, single answer bit per server

]

]

Open Question

- Is PIR practical?
- That is Topic of this Panel!