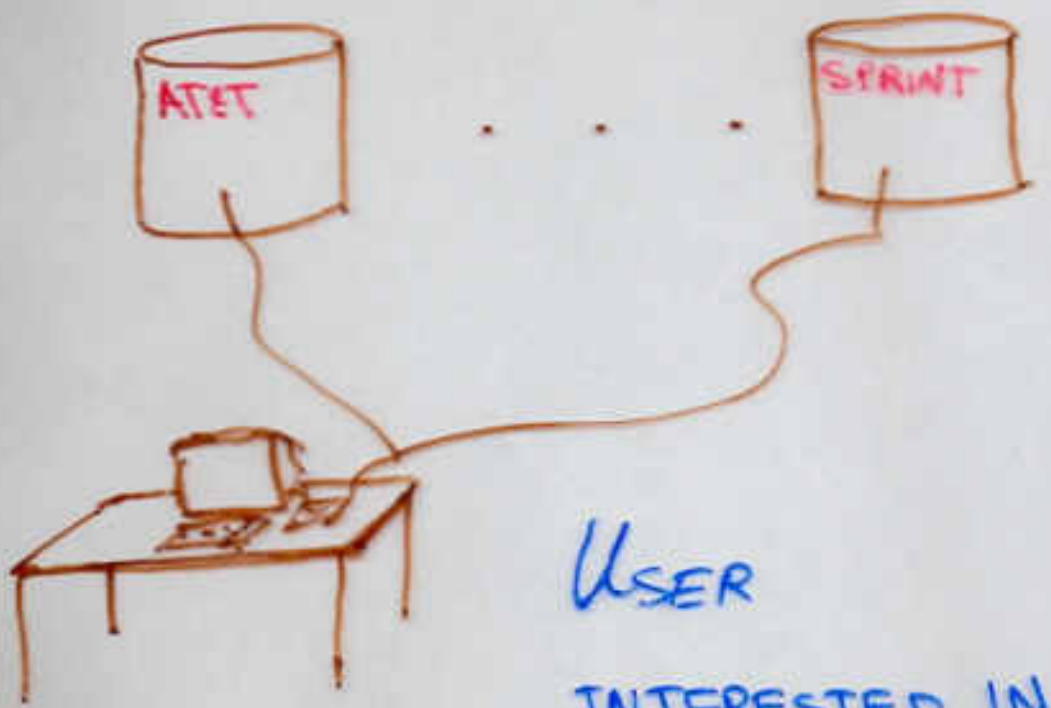
PRIVATE INFORMATION RETRIEVAL

Benny CHOR
Oded GOLDREICH
Eyal KUSHILEVITZ
Madhu SUDAN

THE PROBLEM/MODEL

- k (REPLICATED) DATABASES ($k=2$)
  EACH HOLDING THE SAME n ITEMS



USER

INTERESTED IN ITEM $i \in [n]$

- PRIVACY: NO SINGLE DATABASE SHOULD KNOW i

- COST: COMMUNICATION (BETW' USER & DATABASES).

SOLUTIONS — PIR SCHEMES

- FOR $K=2$ DATABASES: COST $= 12 \cdot \sqrt[3]{n}$
- $\forall\ K>2$ DATABASES: COST $= O(\sqrt[k]{n})$
- FOR $K = \frac{1}{3} \cdot \log_2 n$, COST $= \frac{1}{3} \log_2^2 n \cdot \log_2 \log_2 n$

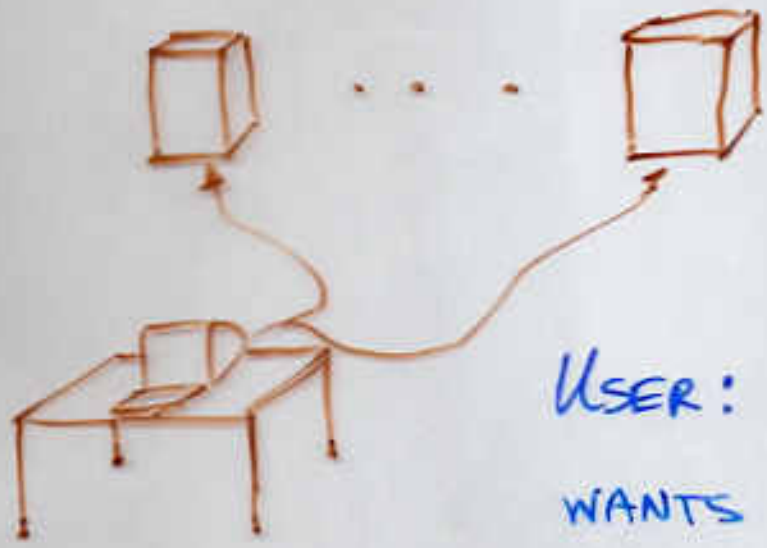
THE TRIVIAL SOLUTION ALLOW $K=1$

BUT REQUIRES COST $= n$.

AFK ≙ ABADI, FEIGENBAUM & KILIAN
BF ≙ BEAVER & FEIGENBAUM
BFKR ≙ BEAVER, FEIGENBAUM, KILIAN & ROGAWAY

A RELATED MODEL – "INSTANCE HIDING"

- k "POWERFUL" COMPUTERS



USER: INSTANCE $i \in \{0,1\}^\ell$
WANTS $f(i)$, WHERE f
IS HARD TO COMPUTE (FOR U).

- PRIVACY: AS BEFORE
- COST: USER OPERATES IN POLY($|i|$)-TIME.

RELATION TO PIR

$[n] \longleftrightarrow \{0,1\}^\ell \equiv [2^\ell]$

DATABASE $\longleftrightarrow$ $f(1) \cdot f(2) \cdots f(2^\ell)$

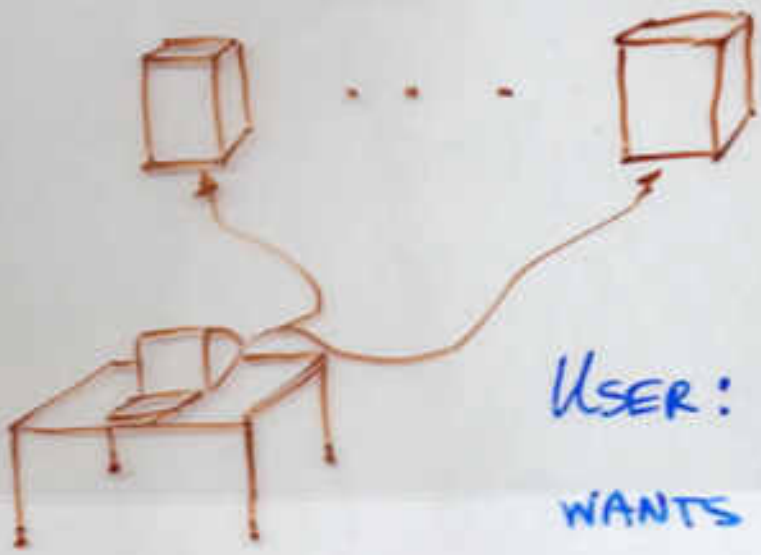
→ SUMMARY: PIR = IH scaled down
+ change of focus/complexity

– IN PIR = Pol
& FOCUS ON C
– IN I.H. = Pol

AFK ≙ ABADI, FEIGENBAUM & KILIAN
BF ≙ BEAVER & FEIGENBAUM
BFKR ≙ BEAVER, FEIGENBAUM, KILIAN & ROGA

A RELATED MODEL – "INSTANCE HIDING"

- k "POWERFUL" COMPUTERS



- [AFK], $k=1$, NEG.
- [BF] $k=\ell=\log_2 n$
- [BFKR] $\forall k>2$

USER: INSTANCE $i \in \{0,1\}^\ell$
WANTS $f(i)$, WHERE f
IS HARD TO COMPUTE (FOR U).

- PRIVACY: AS BEFORE
- COST: USER OPERATES IN POLY($|i|$)-TIME.

RELATION TO PIR

$[n] \longleftrightarrow \{0,1\}^\ell \equiv [2^\ell]$

DATABASE $\longleftrightarrow$ $f(1) \cdot f(2) \cdots f(2^\ell)$

→ SUMMARY: PIR = IH scaled down + change of focus/complexity

– IN PIR = P
& FOCUS ON
– IN I.H. = P

A SIMPLE PIR FOR $k=2$

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

DESIRED ITEM $i \in [n]$

USER SELECTS UNIFORMLY $S \subseteq [n]$

- SENDS S TO DATABASE$_1$
- SENDS $S \oplus \{i\}$ TO DATABASE$_2$

DATABASE, UPON RECEIVING $R \subseteq [n]$,

RETURNS $\bigoplus_{j \in R} x_j$

USER XORs THE 2 BITS.

- PRIVACY: OK.

- COST: LINEAR IN n

NO BETTER THAN "TRIVIAL".

So why did I waste your time?

... because this idea can be generalized to something useful

A CLOSER LOOK AT $\sqrt[3]{n}$ SOLUTION

COMMUNICATION BETWEEN USER & DATABASE

USER —— 3 SUBSETS = 3·m BITS ——→ Database

←—— 1 (XOR) BIT ——

- SUPPOSE DB_{000} RECEIVES (R_1, R_2, R_3).
- IT KNOWS THAT DB_{100} HAS RECEIVED ONE OF THE FOLLOWING $(R_1 \oplus \{j\}, R_2, R_3)$, $j = 1, \ldots, m$.
- IT CAN "SIMULATE" DB_{100} BY REPLYING WITH ALL m POSSIBILITIES. (m BITS!)

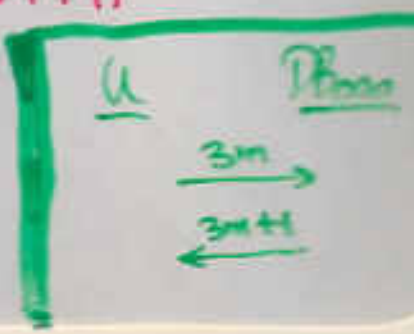
$\Rightarrow$ DB_{000} SIMULATES DB_{100}, DB_{010}, DB_{001}

& DB_{111} SIMULATES DB_{011}, DB_{101}, DB_{110}

$\Rightarrow$ 2 DATABASE SCHEME WITH

PRIVACY : ok.

COST : $\alpha(\sqrt[3]{n})$.



IN GENERAL $(d \geq 3)$

- A PIR FOR 2^d DATABASES WITH COMMUNICATION

  USER $\xrightarrow{\text{d SUBSETS OF } [\sqrt[d]{n}]}$ DATABASE

  $= d \cdot \sqrt[d]{n}$ BITS

  $\xleftarrow{\text{1 (XOR) BIT}}$

- A COVERING CODE (OF RADIUS 1) FOR $\{0,1\}^d$ USING K CODEWORDS. $\left(\frac{2^d}{d+1} \leq k \leq 2^d\right)$ ← "VOLUME BOUND"

$\Rightarrow$ A PIR FOR K DATABASES WITH COMMUNICATION COST $\approx 2^d \cdot d \cdot \sqrt[d]{n}$.

d	2^d	K	VOLUME BOUND.	
3	8	2	2	PERFECT
4	16	4	4	
5	32	7	6	
6	64	12	10	
7	128	16	16	PERFECT

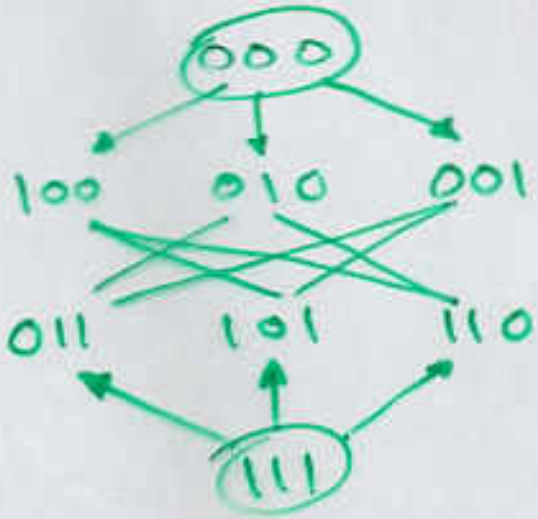
COVERING CODES (OF RAD' 1)

$C \subseteq \{0,1\}^d$ IS A COVERING CODE FOR $\{0,1\}^d$ IF

$\forall x \in \{0,1\}^d \quad \exists y \in C$ S.T. $\text{DIST}(x,y) \leq 1$.

↑ HAMMING DISTANCE.

E.G. $\{000, 111\}$ IS A COVERING CODE FOR $\{0,1\}^3$.



VOLUME BOUND → EACH CODEWORD COVERS $d+1$ STRINGS

$\Rightarrow$ #CODEWORDS $\geq \frac{2^d}{d+1}$

↑ EQUALITY FOR PERFECT COVER

POLYNOMIAL INTERPOLATION PIRs

$x \in \{0,1\}^n \implies X : [n] \to \{0,1\}$

APPLICATION: [BF]

$r_1, \ldots, r_q \in_R F \qquad (q \in \{\ell, m\})$

$f_i(t) \triangleq \hat{X}(i_1 + t \cdot r_1, \ldots, i_q + t \cdot r_q)$

WHERE $(i_1, \ldots, i_q)$ IS THE ITEM SOUGHT.

OBTAIN FROM j^{TH} DATABASE THE VALUE $f_i(j)$

INTERPOLATE TO OBTAIN $f(0) = X(i_1, \ldots, i_q)$.

POLYNOMIAL INTERPOLATION PIRs

$x \in \{0,1\}^n \implies \chi : [n] \to \{0,1\}$

$\chi : \binom{[m]}{d} \to \{0,1\}, \quad m \approx \sqrt[d]{n}$

$\hat{\chi} : F^m \to F$ DEGREE d MULTI-LIN' EXT'

≈ [BFKR]

I.E., $\hat{\chi}(z_1, \ldots, z_m) \triangleq \sum_{S \in \binom{[m]}{d}} \left(\prod_{j \in S} z_j\right) \cdot \chi(S)$.

APPLICATION: [BF]

$r_1, \ldots, r_q \in_R F \qquad (q \in \{\ell, m\})$

$f(t) \triangleq \hat{\chi}(i_1 + t \cdot r_1, \ldots, i_q + t \cdot r_q)$

WHERE $(i_1, \ldots, i_q)$ IS THE ITEM SOUGHT.

OBTAIN FROM j^{TH} DATABASE THE VALUE $f(j)$

INTERPOLATE TO OBTAIN $f(0) = \chi(i_1, \ldots, i_q)$.

OPEN PROBLEM

BEST PIR SCHEMES

#DB	COST	CONJ.
2	$\alpha(\sqrt[3]{n})$	$\Omega(\sqrt[3]{n})$
$k>2$	$\alpha(\sqrt[k]{n})$	$\Omega(\sqrt[2k-1]{n})$

$\Longleftrightarrow$ NOT TIGHT

"BEST" LOWER BOUND

FOR $K=2$, IF USER IS ONLY ALLOWED A SINGLE BOOLEAN QUERY THEN $|\text{QUERY}| = \Omega(n)$.

ADDENDUM (JULY 2008)

The conjectured lowerbound for multiple-server PIRs(i.e., the case of $k > 2$) was disproved a couple of years afterwards by Ambainis in his paper ``An Upper Bound On The Communication Complexity of Private Information Retrieval" (24th ICALP, LNCS 1256, pages 401--407, 1997).Ambainis established an upper bound of $n^{1/(2k-1)}$, which became my revised conjecture for a tight result.

This conjecture was disproved by Beimel, Ishai, Kushilevitz, and Raymond in their paper ``Breaking the $O(n^{1/(2k-1)})$ barrier for information-theoretic private information retrieval"(43rd FOCS, pages 261--270, 2002).

My last attempt at a conjecture was that for every k there exists a constant $c = c(k)$ such that a k-server PIR requires communication complexity n^c. Furthermore, I expected $c(k)$ to equal the reciprocal of some small polynomial.

This last conjecture seems to be disproven by Yekhanin in his paper``Towards 3-Query Locally Decodable Codes of Subexponential Length" (39th STOC, pages 266--274, 2007).

I dare not make further conjectures....