Efficient Private Information Retrieval protocols based on transversal designs

Julien Lavauzelle

Team GRACE
LIX & INRIA Saclay, Université Paris-Saclay

Workshop Code-Based Cryptography 2017, Tenerife, Spain
02/06/2017

1. Definitions
2. Transversal designs for efficient PIR protocols
3. Constructions

1. Definitions

2. Transversal designs for efficient PIR protocols

3. Constructions

   First construction: affine transversal designs

   Second construction: with orthogonal arrays

Issue

Given a file F,

can we retrieve F_i
without leaking any information on i?

Examples:

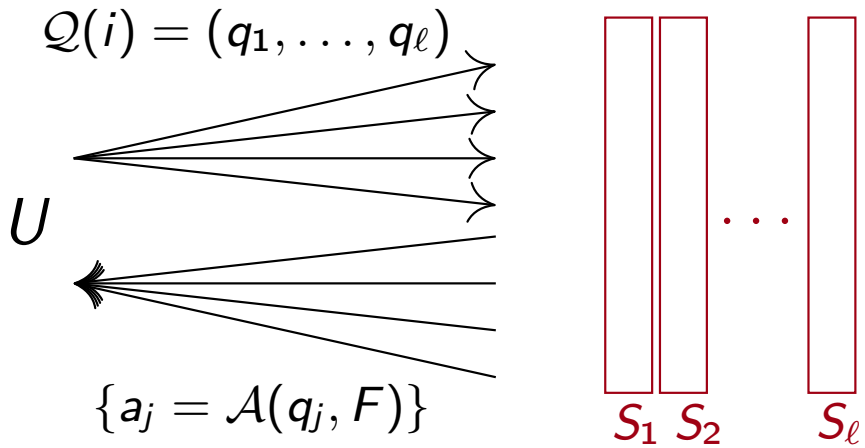
- confidential medical data,
- stock exchange prices...

Private Information Retrieval protocols

Let F be a file stored on a DSS with ℓ servers $S_1, \ldots, S_\ell$.

Private Information Retrieval (PIR) protocol: a user U wants to recover F_i privately.

1. U generates a query vector $\mathbf{q} = \mathcal{Q}(i)$ and sends q_j to S_j
2. Each server S_j computes $a_j = \mathcal{A}(q_j, F)$ and sends it back to U
3. U recovers $F_i = \mathcal{R}(\mathbf{q}, \mathbf{a}, i)$



IT-security: we want $\mathbb{P}(q_j|i) = \mathbb{P}(q_j)$, $\forall j = 1, \ldots, \ell$.

Parameters and previous works

Design goals:

- Low communication complexity (exchanged bits).
- Low computation complexity for $\mathcal{A}$ (server) and $\mathcal{R}$ (user).
- Low storage overhead (for the servers).

Parameters and previous works

Design goals:

- Low communication complexity (exchanged bits).
- Low computation complexity for $\mathcal{A}$ (server) and $\mathcal{R}$ (user).
- Low storage overhead (for the servers).

Existing solutions:

- Download the whole file F...

Parameters and previous works

Design goals:

- Low communication complexity (exchanged bits).
- Low computation complexity for $\mathcal{A}$ (server) and $\mathcal{R}$ (user).
- Low storage overhead (for the servers).

Existing solutions:

- Download the whole file F... inefficient, but it's the best solution with only one server [Chor Goldreich Kushilevitz Sudan '95].

Parameters and previous works

Design goals:

- Low communication complexity (exchanged bits).
- Low computation complexity for $\mathcal{A}$ (server) and $\mathcal{R}$ (user).
- Low storage overhead (for the servers).

Existing solutions:

- Download the whole file F... inefficient, but it's the best solution with only one server [Chor Goldreich Kushilevitz Sudan '95].
- Use *smooth locally decodable codes* with locality ℓ:
  - ℓ servers, each storing a copy of F
  - use the ℓ-query local decoding algorithm to recover F_i
  - *smoothness* ensures security

Parameters and previous works

Design goals:

- Low communication complexity (exchanged bits).
- Low computation complexity for $\mathcal{A}$ (server) and $\mathcal{R}$ (user).
- Low storage overhead (for the servers).

Existing solutions:

- Download the whole file F... inefficient, but it's the best solution with only one server [Chor Goldreich Kushilevitz Sudan '95].
- Use *smooth locally decodable codes* with locality ℓ:
  - ℓ servers, each storing a copy of F (heavy storage overhead)
  - use the ℓ-query local decoding algorithm to recover F_i (complexity?)
  - *smoothness* ensures security

1. Definitions

2. Transversal designs for efficient PIR protocols

3. Constructions
   - First construction: affine transversal designs
   - Second construction: with orthogonal arrays

Motivations

Storage: split an encoded version of the file over the servers (instead of replicating)

Security: the code must have a "smooth" set of parity-check equations for recovering any symbol F_i

An example

Let $\mathbb{F}_q^m = \{P_1, \ldots, P_{q^m}\}$. A q-ary Reed-Muller code is:

$$\mathrm{RM}_q(m, r) = \left\{ \big(f(P_1), \ldots, f(P_{q^m})\big),\ f \in \mathbb{F}_q[X_1, \ldots, X_m], \deg f \leq r \right\}.$$

For $r \leq q - 2$, every $c \in \mathrm{RM}_q(m, r)$ satisfies:

$$\sum_{P \in L} c_P = 0, \quad \forall \text{ line } L \subset \mathbb{F}_q^m$$

An example (cont'd)

Let $\mathcal{G} = \{G_1, \ldots, G_q\}$ be a partition of $\mathbb{F}_q^m$ into q hyperplanes.

1) Encode F into c with $\mathrm{RM}_q(m, r)$. Give $c_{|G_j}$ to server S_j.

2) To recover $F_i = c_i$ **for some** $i \in \mathbb{F}_q^m$**:**

- Pick a line L through i
- Ask server S_j for c_{P_j} where $\{P_j\} = L \cap G_j$, except if $P_j = i$.
- Reconstruct

$$c_i = -\sum_{i \neq P_j \in L} c_{P_j}$$

Security: there is a line between i and any other point of $\mathbb{F}_q^m$.

An example (cont'd [2])

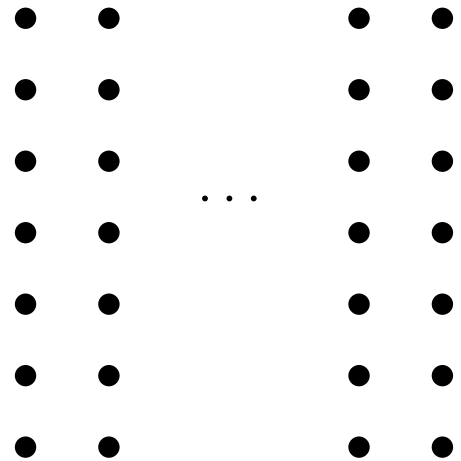
But $\mathrm{RM}_q(m, r)$ with $r < q$ has rate $\leq \frac{1}{m!}$

Generalization: build a similar code through its parity-check equations according to an appropriate incidence structure.

Transversal designs

A **transversal design** $\mathcal{T} = \mathrm{TD}(\ell, s)$ is a 3-tuple $(X, \mathcal{B}, \mathcal{G})$ of sets:

- X is the set of *points*, $|X| = n = s\ell$,

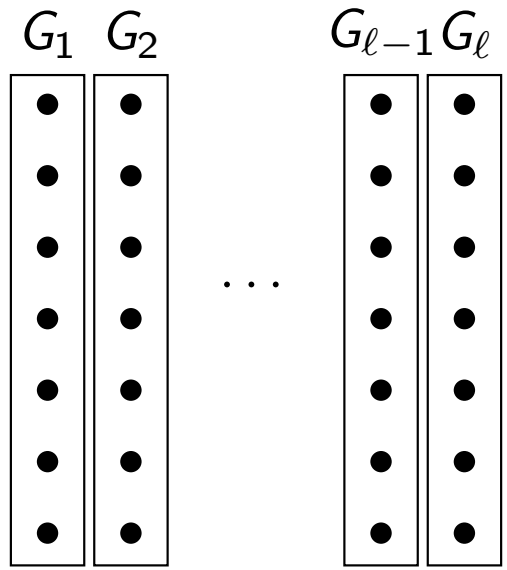


Transversal designs

A **transversal design** $\mathcal{T} = \mathrm{TD}(\ell, s)$ is a 3-tuple $(X, \mathcal{B}, \mathcal{G})$ of sets:

- X is the set of *points*, $|X| = n = s\ell$,
- the *groups* $\mathcal{G} = \{G_j\}_{1 \le j \le \ell}$ satisfy

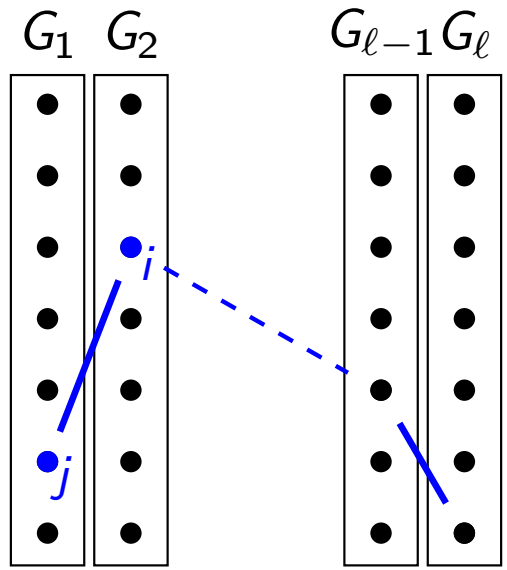
$$X = \coprod_{i=1}^{\ell} G_j \text{ and } |G_j| = s\,,$$



Transversal designs

A **transversal design** $\mathcal{T} = \mathrm{TD}(\ell, s)$ is a 3-tuple $(X, \mathcal{B}, \mathcal{G})$ of sets:

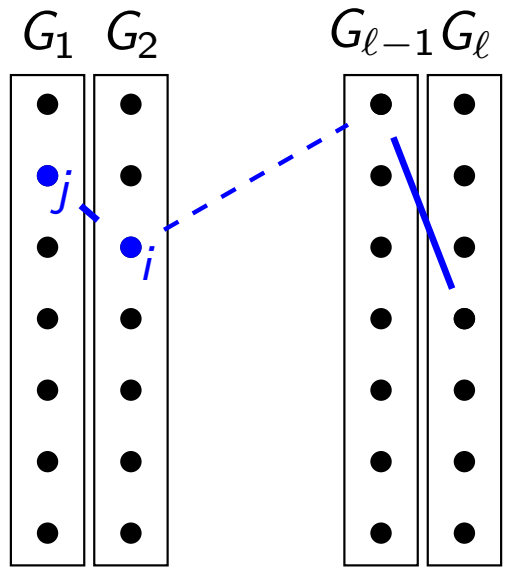
- X is the set of *points*, $|X| = n = s\ell$,
- the *groups* $\mathcal{G} = \{G_j\}_{1 \le j \le \ell}$ satisfy
$$X = \coprod_{i=1}^{\ell} G_j \text{ and } |G_j| = s\,,$$
- the *blocks* $B \in \mathcal{B}$ satisfy:
  - $B \subset X$ and $|B| = \ell$;
  - $\{i, j\} \subset X$ lie in the same group, or $\exists! B \in \mathcal{B}$ such that $\{i, j\} \subset B$



Transversal designs

A **transversal design** $\mathcal{T} = \mathrm{TD}(\ell, s)$ is a 3-tuple $(X, \mathcal{B}, \mathcal{G})$ of sets:

- X is the set of *points*, $|X| = n = s\ell$,
- the *groups* $\mathcal{G} = \{G_j\}_{1 \le j \le \ell}$ satisfy
$$X = \coprod_{i=1}^{\ell} G_j \text{ and } |G_j| = s\,,$$
- the *blocks* $B \in \mathcal{B}$ satisfy:
  - $B \subset X$ and $|B| = \ell$;
  - $\{i, j\} \subset X$ lie in the same group, or $\exists! B \in \mathcal{B}$ such that $\{i, j\} \subset B$



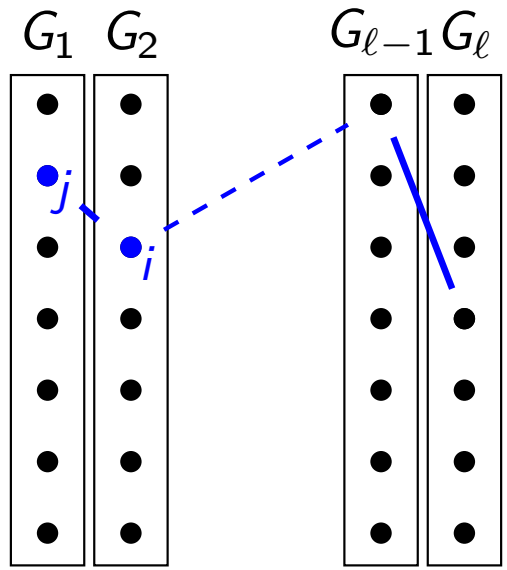
Transversal designs

A **transversal design** $\mathcal{T} = \mathrm{TD}(\ell, s)$ is a 3-tuple $(X, \mathcal{B}, \mathcal{G})$ of sets:

- X is the set of *points*, $|X| = n = s\ell$,
- the *groups* $\mathcal{G} = \{G_j\}_{1 \le j \le \ell}$ satisfy

$$X = \coprod_{i=1}^{\ell} G_j \text{ and } |G_j| = s\,,$$

- the *blocks* $B \in \mathcal{B}$ satisfy:
  - $B \subset X$ and $|B| = \ell$;
  - $\{i, j\} \subset X$ lie in the same group, or $\exists! B \in \mathcal{B}$ such that $\{i, j\} \subset B$



Its **incidence matrix** M has size $|\mathcal{B}| \times |X|$ and is defined by:

$$M_{i,j} = \begin{cases} 1 & \text{if } x_j \in B_i \\ 0 & \text{otherwise} \end{cases}$$

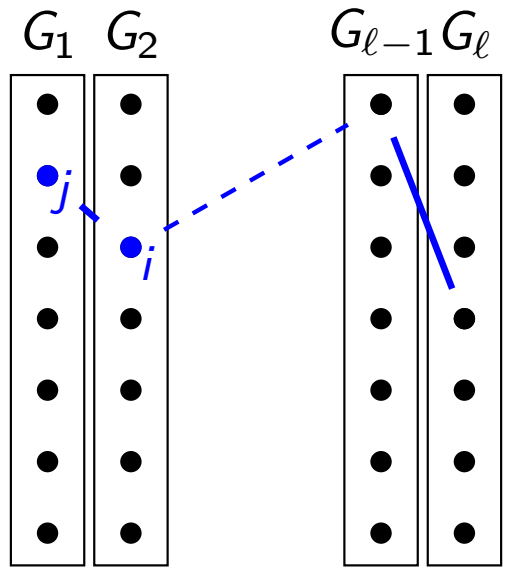
Transversal designs

A **transversal design** $\mathcal{T} = \mathrm{TD}(\ell, s)$ is a 3-tuple $(X, \mathcal{B}, \mathcal{G})$ of sets:

- X is the set of *points*, $|X| = n = s\ell$,
- the *groups* $\mathcal{G} = \{G_j\}_{1 \leq j \leq \ell}$ satisfy

$$X = \coprod_{i=1}^{\ell} G_j \text{ and } |G_j| = s\,,$$

- the *blocks* $B \in \mathcal{B}$ satisfy:
  - $B \subset X$ and $|B| = \ell$;
  - $\{i, j\} \subset X$ lie in the same group, or $\exists! B \in \mathcal{B}$ such that $\{i, j\} \subset B$



Its **incidence matrix** M has size $|\mathcal{B}| \times |X|$ and is defined by:

$$M_{i,j} = \begin{cases} 1 & \text{if } x_j \in B_i \\ 0 & \text{otherwise} \end{cases}$$

The **code** $\mathcal{C}$ **based on** $\mathcal{T}$ **over** $\mathbb{F}_q$ is the $\mathbb{F}_q$-linear code having M as parity-check matrix.

Our PIR protocol construction

Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a code based on a $\mathrm{TD}(\ell, s)$.

Initialisation. User U encodes $c = \mathcal{C}(F)$, and gives $c_{|G_j}$ to server S_j for $j = 1, \ldots, \ell$.

Our PIR protocol construction

Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a code based on a $\mathrm{TD}(\ell, s)$.

Initialisation. User U encodes $c = \mathcal{C}(F)$, and gives $c_{|G_j}$ to server S_j for $j = 1, \ldots, \ell$.

To recover $F_i = c_i$:

1. User U randomly picks a block $B \in \mathcal{B}$ containing i. Then it defines:

$$q_j = \mathcal{Q}(i)_j = \begin{cases} B \cap G_j & \text{if } i \notin G_j \\ \text{a random point in } G_j & \text{otherwise} \end{cases}$$

2. each server S_j sends back $a_j = \mathcal{A}(q_j, c_{|G_j}) = c_{q_j}$
3. U recovers

$$c_i = -\sum_{i \notin G_j} c_{q_j}$$

Properties

Theorem.– If the servers do not collude, then our PIR protocol is information-theoretically secure.

Properties

Theorem.– If the servers do not collude, then our PIR protocol is information-theoretically secure.
Proof:

- the only server which holds F_i received a random query;
- for each other server S_j, there is a constant (=1) number of blocks passing through i and each $q_j \in G_j \Rightarrow$ no information leaks on i.

Properties

Theorem.– If the servers do not collude, then our PIR protocol is information-theoretically secure.
Proof:

- the only server which holds F_i received a random query;
- for each other server S_j, there is a constant (=1) number of blocks passing through i and each $q_j \in G_j \Rightarrow$ no information leaks on i.

Properties. For a $k \log q$ bits file, with $k = \dim_{\mathbb{F}_q} \mathcal{C} \leq n = s\ell$.

- communication complexity: $\ell(\log s + \log q)$ bits
- computational complexity:
  - $\mathcal{O}(1)$ for $\mathcal{A}$ (instead of $\Omega(k \log q)$)
  - $\mathcal{O}(\ell)$ $\mathbb{F}_q$-operations for $\mathcal{R}$
- storage overhead: $(n-k)\log q$ bits (instead of $(\ell-1)k \log q$)

Properties

Theorem.– If the servers do not collude, then our PIR protocol is information-theoretically secure.
Proof:

- the only server which holds F_i received a random query;
- for each other server S_j, there is a constant (=1) number of blocks passing through i and each $q_j \in G_j \Rightarrow$ no information leaks on i.

Properties. For a $k \log q$ bits file, with $k = \dim_{\mathbb{F}_q} \mathcal{C} \leq n = s\ell$.

- communication complexity: $\ell(\log s + \log q)$ bits
- computational complexity:
  - $\mathcal{O}(1)$ for $\mathcal{A}$ (instead of $\Omega(k \log q)$)
  - $\mathcal{O}(\ell)$ $\mathbb{F}_q$-operations for $\mathcal{R}$
- storage overhead: $(n-k)\log q$ bits (instead of $(\ell-1)k \log q$)

Main issue: best k depending on ℓ, n?

1. Definitions

2. Transversal designs for efficient PIR protocols

3. Constructions

First construction: affine transversal designs

Second construction: with orthogonal arrays

1. Definitions

2. Transversal designs for efficient PIR protocols

3. Constructions
   - First construction: affine transversal designs
   - Second construction: with orthogonal arrays

A classical TD: points/lines/hyperplanes

Let $\mathcal{T}_A$ be the **classical affine TD**:

- $X = \mathbb{F}_q^m$, $m \geq 2$,
- $\mathcal{G}$ a set of q disjoint hyperplanes partitionning X,
- $\mathcal{B} = \{\text{affine lines } L \text{ secant to each group of } \mathcal{G}\}$.

A classical TD: points/lines/hyperplanes

Let $\mathcal{T}_A$ be the **classical affine TD**:

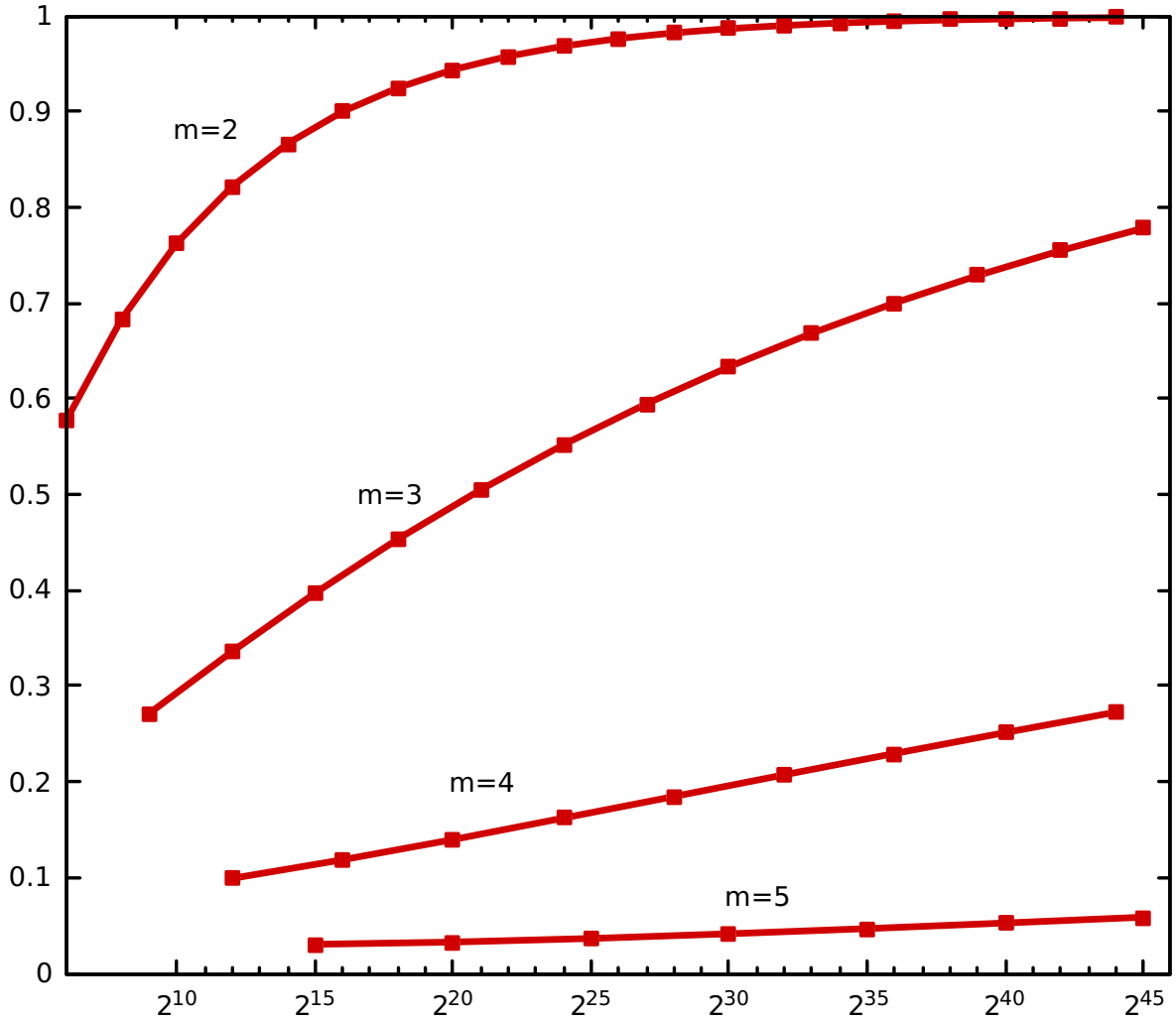
- $X = \mathbb{F}_q^m$, $m \geq 2$,
- $\mathcal{G}$ a set of q disjoint hyperplanes partitionning X,
- $\mathcal{B} = \{\text{affine lines } L \text{ secant to each group of } \mathcal{G}\}$.

The associated $\mathbb{F}_q$-linear code has

- length $n = q^m$
- block size $\ell = q$
- dimension?
  - its parity-check matrix has q^m columns and q^{2m-2} rows...
  - ... but it contains $\mathrm{RM}(m, q-2)$ which has rate $\simeq 1/m!$,
  - and sometimes it is even larger:

Rate of classical TDs

rate $R = k/n$



length $n = 2^{em}$

Particular case: $m = 2$

For $m = 2$, $q = p^e$, using Hamada's formula [Ham68] we obtain:

$$n = p^{2e}, \quad k \geq p^{2e} - \binom{p+1}{2}^e, \quad \ell = \sqrt{n},$$

that is

$$\begin{cases} R = k/n & = 1 - \Theta(n^{c_p}) \\ \ell & = \Theta(n^{1/2}) \end{cases}$$

where $c_p = \frac{1}{2}(\log_p(\frac{p+1}{2}) - 1) < 0$.

We have $c_p \nearrow$, with $c_2 = -0.208$ and $c_\infty = 0$.

Questions:

- is this construction optimal?
- bounds on ℓ and R?

1. Definitions

2. Transversal designs for efficient PIR protocols

3. Constructions
   - First construction: affine transversal designs
   - Second construction: with orthogonal arrays

TDs and orthogonal arrays

An *orthogonal array* $\mathrm{OA}(t, \ell, s)$ *of strength* t may be seen as a code over S, with:

- $|S| = s$,
- length ℓ,
- cardinality $N = s^t$,
- and dual distance $d^\perp = t + 1$

TDs and orthogonal arrays

An *orthogonal array* $\mathrm{OA}(t, \ell, s)$ *of strength* t may be seen as a code over S, with:

- $|S| = s$,
- length ℓ,
- cardinality $N = s^t$,
- and dual distance $d^\perp = t + 1$

$$OA(2,3,2) = \begin{bmatrix} a & b & b \\ b & b & a \\ b & a & b \\ a & a & a \end{bmatrix}$$

TDs and orthogonal arrays

An *orthogonal array* $\mathrm{OA}(t, \ell, s)$ *of strength* t may be seen as a code over S, with:

- $|S| = s$,
- length ℓ,
- cardinality $N = s^t$,
- and dual distance $d^\perp = t + 1$

$$OA(2,3,2) = \begin{bmatrix} a & b & b \\ b & b & a \\ b & a & b \\ a & a & a \end{bmatrix}$$

Construction OA $\rightarrow$ TD :

- $X = S \times [1, \ell]$
- $\mathcal{G} = \{S \times \{i\}, i \in [1, \ell]\}$

$$\begin{array}{ccc} (a,1) & (a,2) & (a,3) \\ (b,1) & (b,2) & (b,3) \end{array}$$

TDs and orthogonal arrays

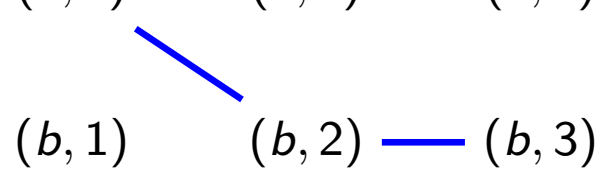
An *orthogonal array* $\mathrm{OA}(t, \ell, s)$ *of strength* t may be seen as a code over S, with:

- $|S| = s$,
- length ℓ,
- cardinality $N = s^t$,
- and dual distance $d^\perp = t + 1$

$$OA(2,3,2) = \begin{bmatrix} a & b & b \\ b & b & a \\ b & a & b \\ a & a & a \end{bmatrix}$$

Construction OA $\rightarrow$ TD :

- $X = S \times [1, \ell]$
- $\mathcal{G} = \{S \times \{i\}, i \in [1, \ell]\}$
- $\mathcal{B} = \{\{(c_i, i), 1 \leq i \leq \ell\}, c \in \mathrm{OA}\}$



TDs and orthogonal arrays

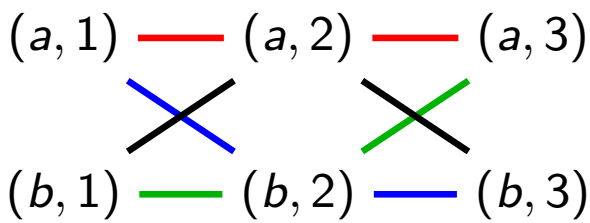
An *orthogonal array* $\mathrm{OA}(t, \ell, s)$ *of strength* t may be seen as a code over S, with:

- $|S| = s$,
- length ℓ,
- cardinality $N = s^t$,
- and dual distance $d^\perp = t + 1$

$$OA(2,3,2) = \begin{bmatrix} a & b & b \\ b & b & a \\ b & a & b \\ a & a & a \end{bmatrix}$$

Construction OA $\rightarrow$ TD :

- $X = S \times [1, \ell]$
- $\mathcal{G} = \{S \times \{i\}, i \in [1, \ell]\}$
- $\mathcal{B} = \{\{(c_i, i), 1 \leq i \leq \ell\}, c \in \mathrm{OA}\}$



TDs and orthogonal arrays

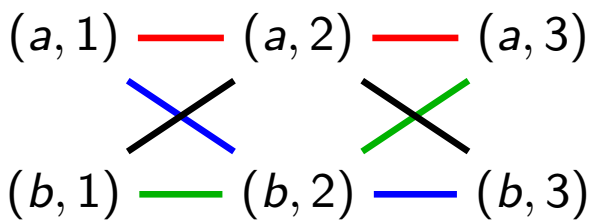
An *orthogonal array* $\mathrm{OA}(t, \ell, s)$ *of strength* t may be seen as a code over S, with:

- $|S| = s$,
- length ℓ,
- cardinality $N = s^t$,
- and dual distance $d^\perp = t + 1$

$$OA(2,3,2) = \begin{bmatrix} a & b & b \\ b & b & a \\ b & a & b \\ a & a & a \end{bmatrix}$$

Construction OA $\to$ TD :

- $X = S \times [1, \ell]$
- $\mathcal{G} = \{S \times \{i\}, i \in [1, \ell]\}$
- $\mathcal{B} = \{\{(c_i, i), 1 \le i \le \ell\}, c \in \mathrm{OA}\}$



Prop. If $t = 2$, then we obtain a $\mathrm{TD}(\ell, s)$ from an $\mathrm{OA}(t, \ell, s)$.

Resistance to collusions

What about $\mathrm{OA}(t, \ell, s)$ with $t > 2$?

For each t-tuple of points lying in t different groups, there is a block which contains them all.

$\Rightarrow$ Our PIR protocol resists $t - 1$ collusive servers.

Resistance to collusions

What about $\mathrm{OA}(t, \ell, s)$ with $t > 2$?

For each t-tuple of points lying in t different groups, there is a block which contains them all.

$\Rightarrow$ Our PIR protocol resists $t - 1$ collusive servers.

But in practice, the PIR storage overhead increases with t (see later).

The "coded-queries code" construction

Definition.– We call $\mathcal{C}_0$-*coded-queries code* (denoted $\mathrm{Code}_q(\mathcal{C}_0)$) the $\mathbb{F}_q$-linear code $\mathcal{C}$ coming from the successive constructions:

$$\mathcal{C}_0 = \mathrm{OA}(t, \ell, s) \quad \mapsto \quad \text{generalized } \mathrm{TD}(\ell, s; t) \quad \mapsto \quad \mathcal{C} = \mathrm{Code}_q(\mathcal{C}_0)$$

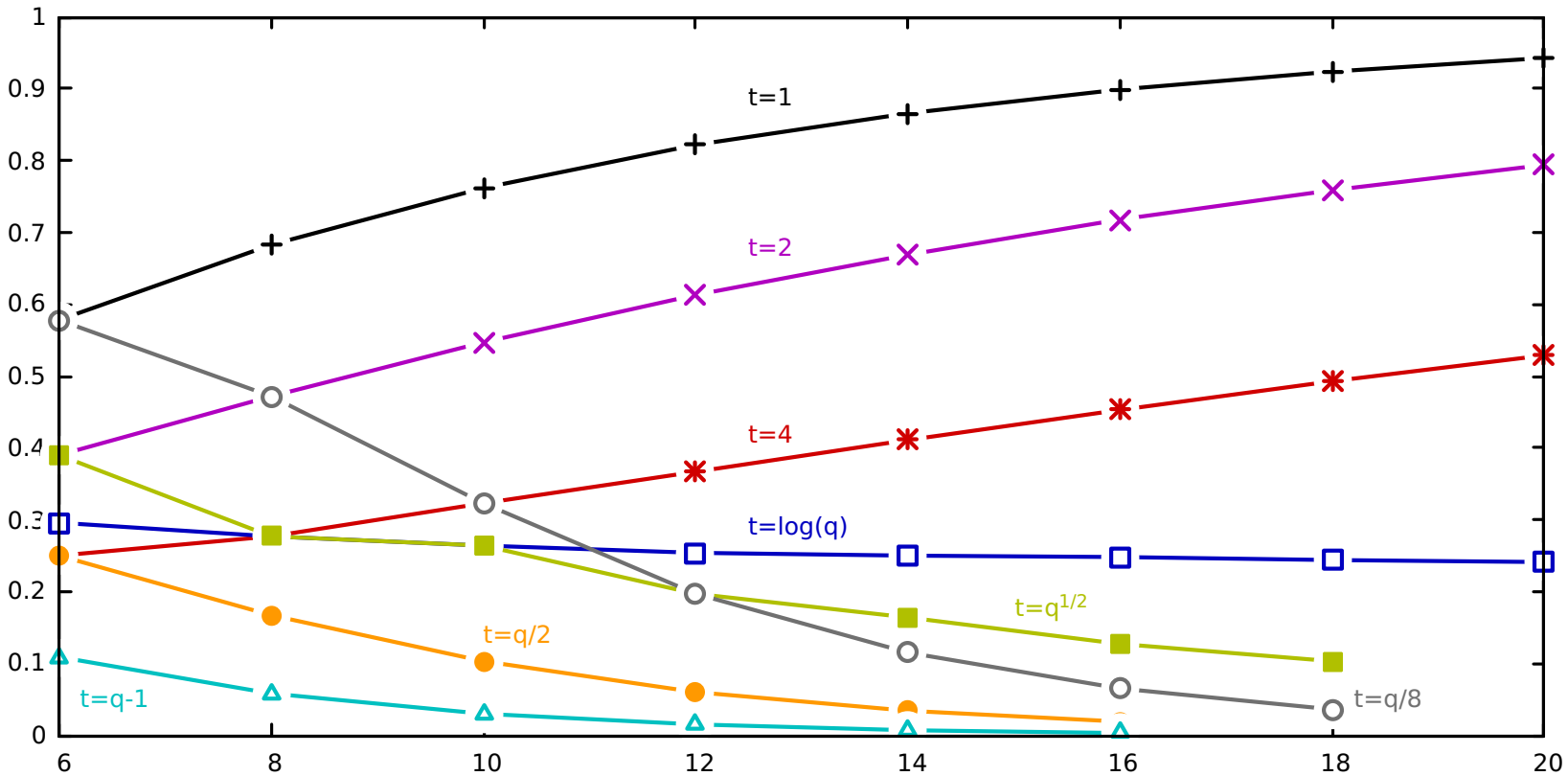
We derive PIR parameters from those of $\mathcal{C}_0$:

- $d^\perp - 2$ is the number of collusive servers the protocol resists
- the larger $\mathcal{C}_0$, the larger PIR storage overhead

$$\Rightarrow \text{ let's use MDS codes}$$

Reed-Solomon-coded-queries codes

Example: for $\mathrm{OA}(t+1, \ell = q, s = q) = \mathrm{RS}(\mathbb{F}_q, t+1)$:



Conclusion

Summary: (server-)efficient PIR protocols can be built upon codes from transversal designs

Current issues:

- transversal designs with low-rank parity-check matrices?
- bounds, optimal constructions?
- (divisible projective codes $\mathcal{C}_0$ over large alphabets?)

Questions?

Remarks on coded-queries codes

Proposition.– For all codes $\mathcal{C}_0$ of length ℓ over $\mathbb{F}_s$, $\mathrm{Code}_q(\mathcal{C}_0)$ is an $[n, k]_q$ code with:

- $n = s\ell$,
- $\ell - 1 \leq k \leq n - \sqrt{n}$.

Proposition.– Let H be the parity-check matrix of $\mathrm{Code}_q(\mathcal{C}_0)$. Then,

$$HH^T = \ell J - D(\mathcal{C}_0)\,,$$

where J is the all-1 matrix and

$$D(\mathcal{C}_0)_{c,c'} = d(c, c'), \quad \forall c, c' \in \mathcal{C}_0$$

Divisible codes for efficient PIR protocols

A p-divisible code is a code whose codewords' weights are divisible by p.

Corollary.– If $\mathcal{C}_0$ is p-divisible for $p = \mathrm{char}(\mathbb{F}_q)$, then:

$$k = \dim \mathrm{Code}_q(\mathcal{C}_0) \geq \frac{n-1}{2}\,.$$

Furthermore, if $p \,|\, \ell$, then:

$$HH^T = 0 \quad \Rightarrow \quad \mathcal{C}^\perp \subseteq \mathcal{C}$$

Theorem.– If there exists a p-divisible code $\mathcal{C}_0$ of length ℓ and dual distance $t+2$, then there exists a PIR protocol resisting to t colluding servers, with rate $\gtrsim 1/2$.

Question.– Do there exist projective ($d^\perp \geq 3$) p-divisible codes of length ℓ over $\mathbb{F}_q$, with $q \gg \ell$?