Private information retrieval schemes based on codes

Julien Lavauzelle

IRMAR, Université de Rennes 1

Séminaire Mathématiques Discrètes, Codes et Cryptographie

20/02/2020

Cryptographic primitives based on codes

Best-known **cryptographic primitive** based on **coding theory**:

McEliece public-key encryption scheme (+ many variants)

Cryptographic primitives based on codes

Best-known **cryptographic primitive** based on **coding theory**:

McEliece public-key encryption scheme (+ many variants)

But... many other ones:

- signature (CFS, Wave), identification schemes (Stern),
- secret sharing schemes (Shamir),

Cryptographic primitives based on codes

Best-known **cryptographic primitive** based on **coding theory**:

McEliece public-key encryption scheme (+ many variants)

But... many other ones:

- signature (CFS, Wave), identification schemes (Stern),
- secret sharing schemes (Shamir),
- proofs of retrievability,

Cryptographic primitives based on codes

Best-known **cryptographic primitive** based on **coding theory**:

McEliece public-key encryption scheme (+ many variants)

But... many other ones:

- signature (CFS, Wave), identification schemes (Stern),
- secret sharing schemes (Shamir),
- proofs of retrievability,
- **private information retrieval**.

Outline

1. Private information retrieval

2. PIR schemes with low computation
   - Transversal designs and codes
   - A PIR scheme with transversal designs
   - Towards collusion resistance
   - PIR schemes with lifted codes

3. Other constructions of PIR schemes

4. Conclusion

Outline

1. Private information retrieval

2. PIR schemes with low computation
   - Transversal designs and codes
   - A PIR scheme with transversal designs
   - Towards collusion resistance
   - PIR schemes with lifted codes

3. Other constructions of PIR schemes

4. Conclusion

Problem statement

Private information retrieval (PIR):

Given a **remote** database $F = (F_1, \ldots, F_M) \in \Sigma^M$
and an index $i \in [1, M]$,
can we **retrieve** the entry/file F_i,
without leaking any information on i?

Problem statement

Private information retrieval (PIR):

Given a **remote** database $F = (F_1, \ldots, F_M) \in \Sigma^M$
and an index $i \in [1, M]$,
can we **retrieve** the entry/file F_i,
without leaking any information on i?

Application: private search, medical data, etc.

Problem statement

Private information retrieval (PIR):

Given a **remote** database $F = (F_1, \ldots, F_M) \in \Sigma^M$
and an index $i \in [1, M]$,
can we **retrieve** the entry/file F_i,
without leaking any information on i?

Application: private search, medical data, etc.

Trivial solution: full download.

Definition of PIR

Introduced in:

Private Information Retrieval. Chor, Goldreich, Kushilevitz, Sudan. FOCS. **1995**.

A database $F = (F_1, \ldots, F_m)$ stored (in some way) on n servers $S_1, \ldots, S_n$.
A user U wants to recover F_i privately.

A **Private Information Retrieval protocol** is a set of algorithms $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$:

Definition of PIR

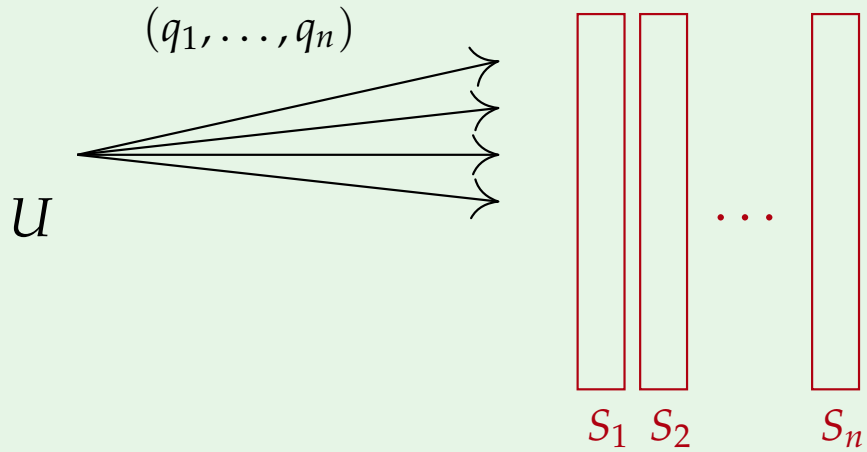
Introduced in:

Private Information Retrieval. Chor, Goldreich, Kushilevitz, Sudan. FOCS. **1995**.

A database $F = (F_1, \ldots, F_m)$ stored (in some way) on n servers $S_1, \ldots, S_n$.
A user U wants to recover F_i privately.

A **Private Information Retrieval protocol** is a set of algorithms $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$:

1. U generates a query vector $\boldsymbol{q} = (q_1, \ldots, q_n) \leftarrow \mathcal{Q}(i)$ and sends q_j to server S_j



Definition of PIR

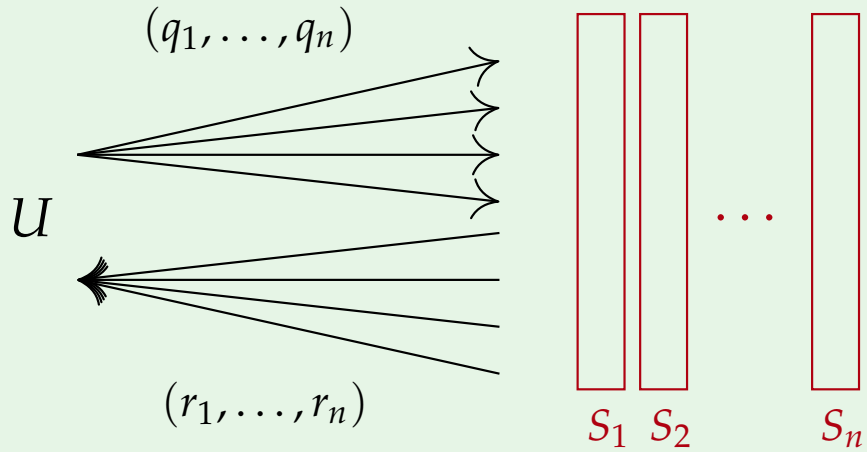
Introduced in:

Private Information Retrieval. Chor, Goldreich, Kushilevitz, Sudan. FOCS. **1995**.

A database $F = (F_1, \ldots, F_m)$ stored (in some way) on n servers $S_1, \ldots, S_n$.
A user U wants to recover F_i privately.

A **Private Information Retrieval protocol** is a set of algorithms $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$:

1. U generates a query vector $\boldsymbol{q} = (q_1, \ldots, q_n) \leftarrow \mathcal{Q}(i)$ and sends q_j to server S_j
2. Each server S_j computes $r_j = \mathcal{A}(q_j, F_{|S_j})$ and sends it back to U



Definition of PIR

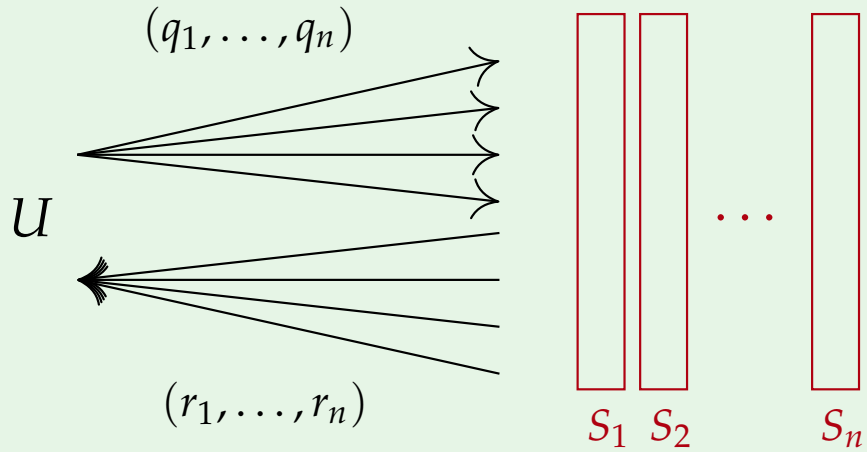
Introduced in:

Private Information Retrieval. Chor, Goldreich, Kushilevitz, Sudan. FOCS. **1995**.

A database $F = (F_1, \ldots, F_m)$ stored (in some way) on n servers $S_1, \ldots, S_n$.
A user U wants to recover F_i privately.

A **Private Information Retrieval protocol** is a set of algorithms $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$:

1. U generates a query vector $\boldsymbol{q} = (q_1, \ldots, q_n) \leftarrow \mathcal{Q}(i)$ and sends q_j to server S_j
2. Each server S_j computes $r_j = \mathcal{A}(q_j, F_{|S_j})$ and sends it back to U
3. U recovers $F_i = \mathcal{R}(\boldsymbol{q}, \boldsymbol{r}, i)$



Privacy

A **collusion of servers**: set of servers $\{S_j : j \in T\}$, where $T \subset [1, n]$, which exchange information about queries, data, etc.

$$t := \max\{|T|, T \subseteq [1, n] \text{ is a collusion}\} \geq 1$$

Privacy

A **collusion of servers**: set of servers $\{S_j : j \in T\}$, where $T \subset [1, n]$, which exchange information about queries, data, etc.

$$t := \max\{|T|, T \subseteq [1, n] \text{ is a collusion}\} \geq 1$$

- **Information-theoretic privacy:**

$$\mathrm{I}(i; \boldsymbol{q}_{|T}) = 0, \quad \forall T \subseteq [1, n], |T| \leq t.$$

- **Computational privacy:** by varying the index i, distributions of queries $\boldsymbol{q}_{|T} = \mathcal{Q}(i)_{|T}$ are computationally indistinguishable.

Privacy

A **collusion of servers**: set of servers $\{S_j : j \in T\}$, where $T \subset [1, n]$, which exchange information about queries, data, etc.

$$t := \max\{|T|, T \subseteq [1, n] \text{ is a collusion}\} \geq 1$$

- **Information-theoretic privacy:**

$$\mathrm{I}(i; \boldsymbol{q}_{|T}) = 0, \quad \forall T \subseteq [1, n], |T| \leq t.$$

- **Computational privacy:** by varying the index i, distributions of queries $\boldsymbol{q}_{|T} = \mathcal{Q}(i)_{|T}$ are computationally indistinguishable.

Theorem [CGKS95, CG97]. If $t = n$ (in particular if $n = 1$), then:

- for IT-privacy, **no better solution than full download**,
- computational privacy is possible (but remains **expensive** as of now).

Main parameters of PIR schemes

We mostly focus on **IT-privacy**
(hence need $n \geq 2$ servers)

Main parameters of PIR schemes

We mostly focus on **IT-privacy**
(hence need $n \geq 2$ servers)

Parameters to be taken into account:

- **communication** complexity (upload and download)
- **computation** complexity (client and servers)
- server storage overhead
- maximum size of collusions (t)

Main parameters of PIR schemes

We mostly focus on **IT-privacy**
(hence need $n \geq 2$ servers)

Parameters to be taken into account:

- **communication** complexity (upload and download)
- **computation** complexity (client and servers)
- server storage overhead
- maximum size of collusions (t)

Several possible **settings**:

- bounded vs. unbounded number of entries,
- replicated database vs. coded database,
- dynamic database vs. static database,
- unresponsive or byzantine servers, etc.

Outline

1. Private information retrieval

2. PIR schemes with low computation
   - Transversal designs and codes
   - A PIR scheme with transversal designs
   - Towards collusion resistance
   - PIR schemes with lifted codes

3. Other constructions of PIR schemes

4. Conclusion

Outline

1. Private information retrieval

2. PIR schemes with low computation
   - Transversal designs and codes
   - A PIR scheme with transversal designs
   - Towards collusion resistance
   - PIR schemes with lifted codes

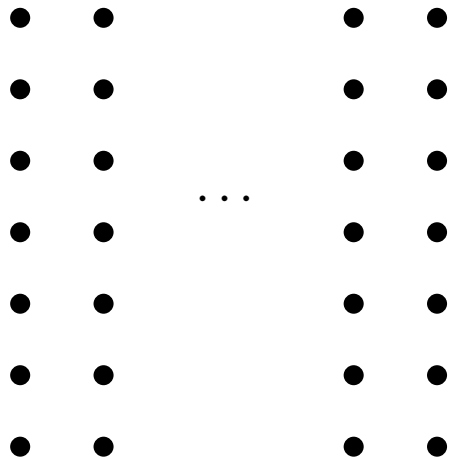
3. Other constructions of PIR schemes

4. Conclusion

Transversal designs

A **transversal design** $\mathrm{TD}(n,s) = (X, \mathcal{B}, \mathcal{G})$ is given by:

- X a set of *points*, $|X| = N = ns$,

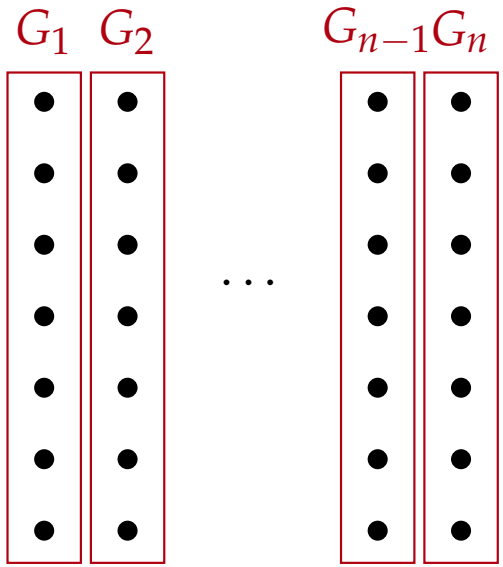


Transversal designs

A **transversal design** $\mathrm{TD}(n,s) = (X, \mathcal{B}, \mathcal{G})$ is given by:

- X a set of *points*, $|X| = N = ns$,
- *groups* $\mathcal{G} = \{G_j\}_{1 \leq j \leq n}$ satisfying

$$X = \coprod_{j=1}^{n} G_j \text{ and } |G_j| = s\,,$$



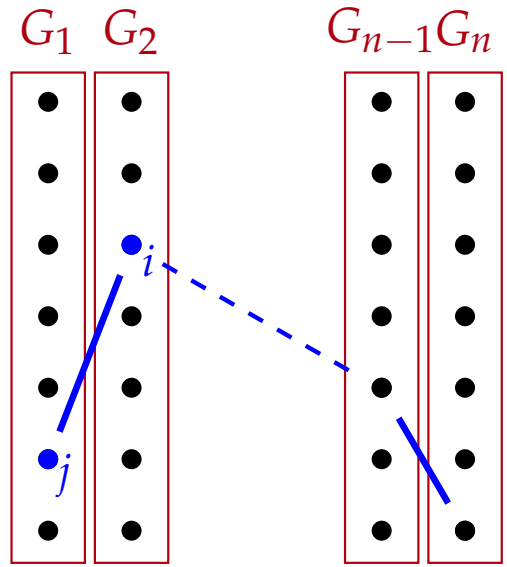
Transversal designs

A **transversal design** $\mathrm{TD}(n,s) = (X, \mathcal{B}, \mathcal{G})$ is given by:

- X a set of *points*, $|X| = N = ns$,
- *groups* $\mathcal{G} = \{G_j\}_{1 \le j \le n}$ satisfying

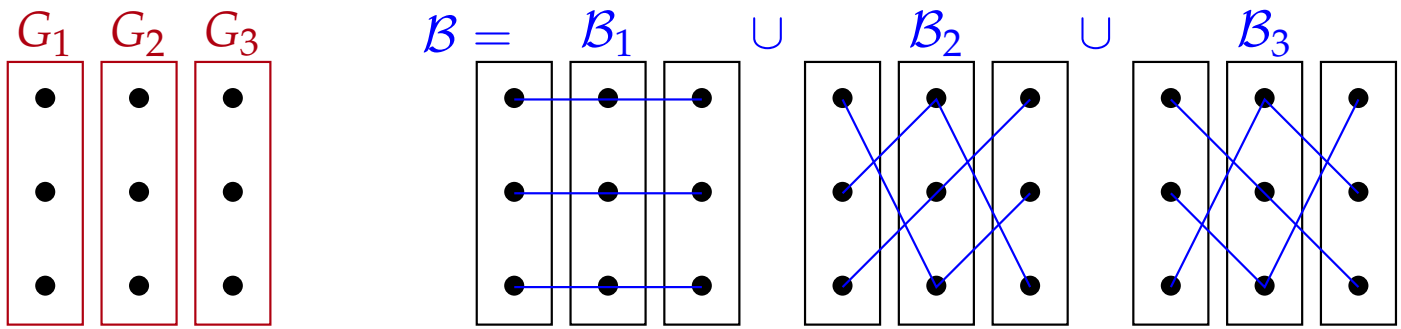
$$X = \coprod_{j=1}^{n} G_j \text{ and } |G_j| = s\,,$$

- *blocks* $B \in \mathcal{B}$ satisfying
  - $B \subset X$ and $|B| = n$;
  - for all $\{i,j\} \subset X$, $\{i,j\}$ lie:
    either in a single group $G \in \mathcal{G}$,
    or in a unique block $B \in \mathcal{B}$



Example: a TD(3,3)

- $ns = 9$ points
- $s = 3$ groups G_1, G_2, G_3 of size 3
- $ns = 9$ blocks of $n = 3$ points, partionned into 3 parallel classes $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$



Codes from designs

Let $\mathcal{T}$ be a transversal design $\mathrm{TD}(n,s) = (X, \mathcal{B}, \mathcal{G})$.

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

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

Codes from designs

Let $\mathcal{T}$ be a transversal design $\mathrm{TD}(n,s) = (X, \mathcal{B}, \mathcal{G})$.

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

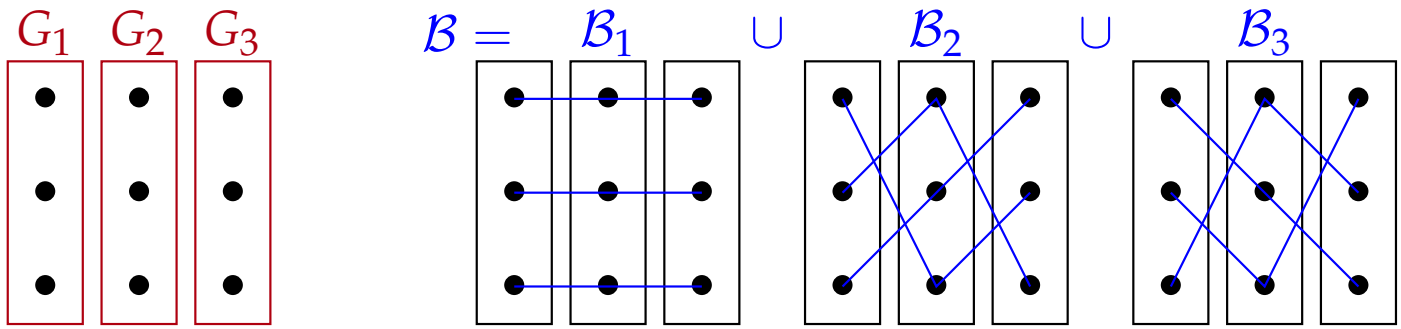
$$M_{i,j} = \begin{cases} 1 & \text{if } x_j \in B_i \\ 0 & \text{otherwise.} \end{cases}$$

Definition. The **code $\mathcal{C}$ based on $\mathcal{T}$ over $\mathbb{F}_q$** is the $\mathbb{F}_q$-linear code admitting $\boldsymbol{M}$ as a parity-check matrix (*i.e.* $\mathcal{C}^\perp$ is generated by $\boldsymbol{M}$).

- $\mathrm{length}(\mathcal{C}) = |X| = ns$,
- $\dim(\mathcal{C}) = \dim(\ker \boldsymbol{M})$,
- every $B \in \mathcal{B}$ gives an $\boldsymbol{h} \in \mathcal{C}^\perp$ such that $\mathrm{wt}(\boldsymbol{h}_{|G_j}) = 1$, $\forall j = 1, \ldots, n$.

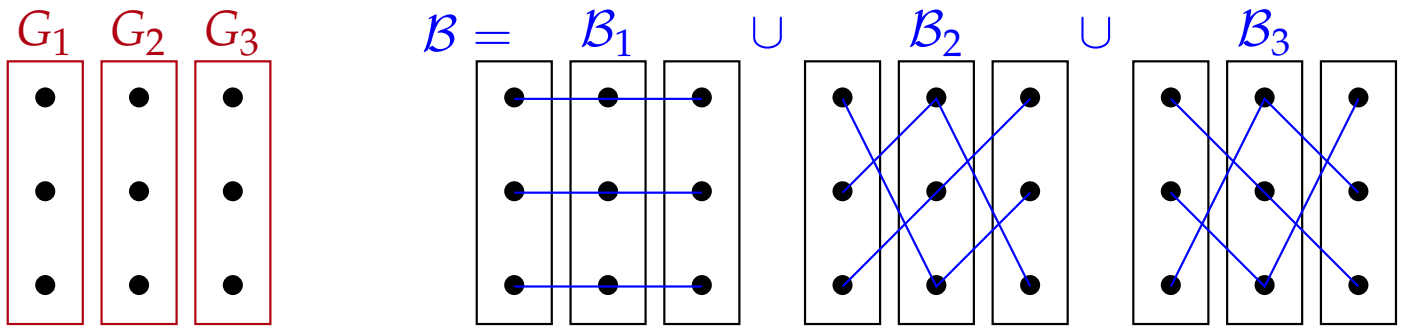
Example

The transversal design $\mathrm{TD}(3,3)$ represented by:



Example

The transversal design TD$(3,3)$ represented by:

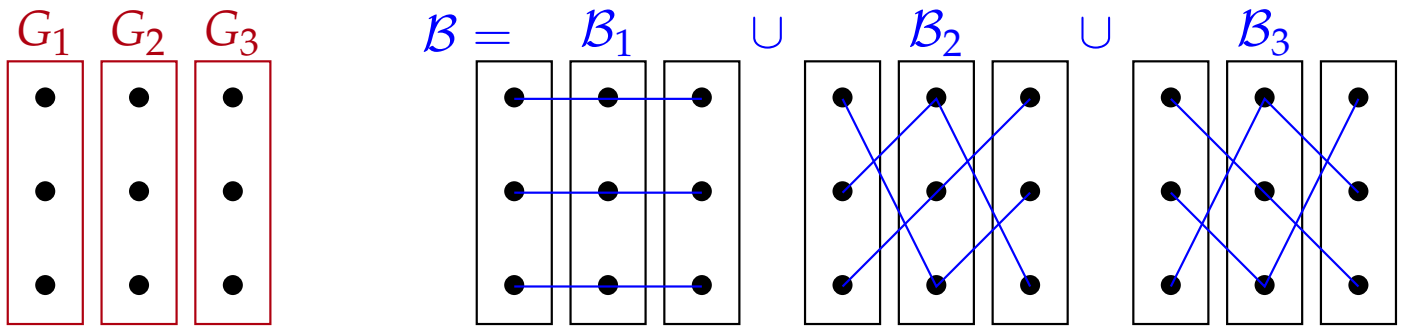


gives an incidence matrix

$$\boldsymbol{M} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Example

The transversal design TD$(3,3)$ represented by:



gives an incidence matrix

$$\boldsymbol{M} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Its rank over $\mathbb{F}_3$ is 6 $\quad\Longrightarrow\quad$ the associated code $\mathcal{C}$ is a $[9,3]_3$ code.

Outline

1. Private information retrieval

2. PIR schemes with low computation
   - Transversal designs and codes
   - A PIR scheme with transversal designs
   - Towards collusion resistance
   - PIR schemes with lifted codes

3. Other constructions of PIR schemes

4. Conclusion

The PIR scheme

Private Information Retrieval from Transversal Designs. **L.**. IEEE-TIT. **2019**.

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

The PIR scheme

Private Information Retrieval from Transversal Designs. **L.**. IEEE-TIT. **2019**.

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

- **Initialisation.** User U encodes $F \mapsto \boldsymbol{c} \in \mathcal{C}$, and gives $\boldsymbol{c}_{|G_j}$ to server S_j.

The PIR scheme

Private Information Retrieval from Transversal Designs. **L.**. IEEE-TIT. **2019**.

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

- **Initialisation.** User U encodes $F \mapsto \boldsymbol{c} \in \mathcal{C}$, and gives $\boldsymbol{c}_{|G_j}$ to server S_j.
- **To recover** $F_i = c_i$, with $i \in X$:
  1. User U randomly picks a block $B \in \mathcal{B}$ containing i.
     Then U defines:

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

  2. Each server S_j sends back c_{q_j}
  3. U recovers

$$c_i = -\sum_{j:\, i \notin G_j} c_{q_j} = -\sum_{b \in B \setminus \{i\}} c_b$$

Privacy and parameters

Theorem. This PIR protocol is information-theoretically private.

Proof:

- the only server which holds F_i received a random query;
- for each other server S_j, query q_j gives no information on the block B which has been picked $\Rightarrow$ no information leaks on i.

Privacy and parameters

Theorem. This PIR protocol is information-theoretically private.

Proof:

- the only server which holds F_i received a random query;
- for each other server S_j, query q_j gives no information on the block B which has been picked $\Rightarrow$ no information leaks on i.

Features.

- communication complexity: $n \log s$ uploaded bits, $n \log q$ downloaded bits
- computational complexity:
  - **only 1 read for each server** (optimal)
  - $\leq n$ additions over $\mathbb{F}_q$ for the user
- storage overhead: $(ns - k) \log q$ bits, where $k = \dim(\mathcal{C})$

Privacy and parameters

Theorem. This PIR protocol is information-theoretically private.

Proof:

- the only server which holds F_i received a random query;
- for each other server S_j, query q_j gives no information on the block B which has been picked $\Rightarrow$ no information leaks on i.

Features.

- communication complexity: $n \log s$ uploaded bits, $n \log q$ downloaded bits
- computational complexity:
  - **only 1 read for each server** (optimal)
  - $\leq n$ additions over $\mathbb{F}_q$ for the user
- storage overhead: $(ns - k) \log q$ bits, where $k = \dim(\mathcal{C})$

Question: transversal designs with good $\dim(\mathcal{C})$ depending on (n, s)?

Instances with geometric designs

$\mathcal{T}_A^{q,m}$, the **classical affine transversal design**:

- $X = \mathbb{F}_q^m$ for $m \geq 2$,
- $\mathcal{G}$ a partition of X into q hyperplanes $G_1, \ldots, G_q$,
- $\mathcal{B} = \{\text{affine lines } L \text{ secant to each } G_j\}$.

The code has:

- length $ns = q^m$,
- "locality" $n = q$.

Instances with geometric designs

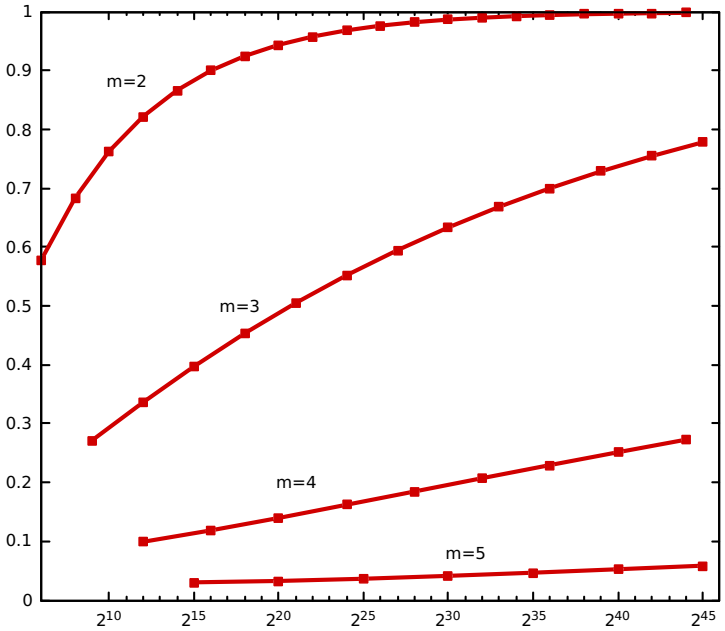
$\mathcal{T}_A^{q,m}$, the **classical affine transversal design**:

- $X = \mathbb{F}_q^m$ for $m \geq 2$,
- $\mathcal{G}$ a partition of X into q hyperplanes $G_1, \ldots, G_q$,
- $\mathcal{B} = \{\text{affine lines } L \text{ secant to each } G_j\}$.

The code has:

- length $ns = q^m$,
- "locality" $n = q$.

rate k/N



length $N = ns = 2^{em}$

Instances with geometric designs

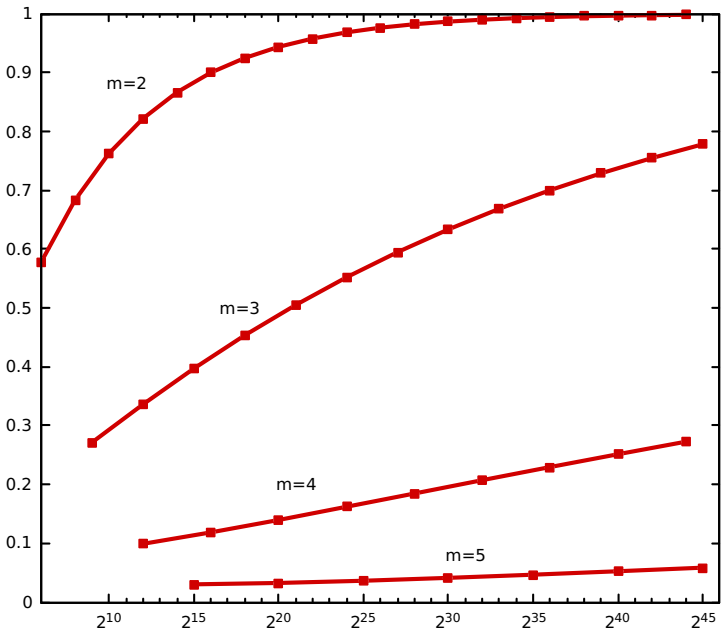
$\mathcal{T}_A^{q,m}$, the **classical affine transversal design**:

- $X = \mathbb{F}_q^m$ for $m \geq 2$,
- $\mathcal{G}$ a partition of X into q hyperplanes $G_1, \ldots, G_q$,
- $\mathcal{B} = \{\text{affine lines } L \text{ secant to each } G_j\}$.

The code has:

- length $ns = q^m$,
- "locality" $n = q$.

rate k/N



length $N = ns = 2^{em}$

Question: better instances?

Outline

1. Private information retrieval

2. PIR schemes with low computation
   - Transversal designs and codes
   - A PIR scheme with transversal designs
   - Towards collusion resistance
   - PIR schemes with lifted codes

3. Other constructions of PIR schemes

4. Conclusion

Orthogonal arrays

An **orthogonal array** $\mathrm{OA}(t, n, s)$ of strength t is a list A of words
- of length n,
- over a finite set S, $|S| = s$,
- such that, for every $I \subset [1, n]$ of size t, $A_{|I} = S^t$.

Equivalently, an $\mathrm{OA}(t, n, s)$ is a code $A \subset S^n$ with dual distance $t + 1$.

$$S = \{a, b\}$$

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

Orthogonal arrays

An **orthogonal array** $\mathrm{OA}(t,n,s)$ of strength t is a list A of words
- of length n,
- over a finite set S, $|S| = s$,
- such that, for every $I \subset [1,n]$ of size t, $A_{|I} = S^t$.

Equivalently, an $\mathrm{OA}(t,n,s)$ is a code $A \subset S^n$ with dual distance $t+1$.

$$S = \{a,b\}$$

Construction OA $\rightarrow$ TD :
- $X = S \times [1,n]$
- $\mathcal{G} = \{S \times \{i\}, 1 \leq i \leq n\}$

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

$$\begin{matrix} (a,1) \\ (b,1) \end{matrix} \qquad \begin{matrix} (a,2) \\ (b,2) \end{matrix} \qquad \begin{matrix} (a,3) \\ (b,3) \end{matrix}$$

Orthogonal arrays

An **orthogonal array** $\mathrm{OA}(t,n,s)$ of strength t is a list A of words
- of length n,
- over a finite set S, $|S| = s$,
- such that, for every $I \subset [1,n]$ of size t, $A_{|I} = S^t$.

Equivalently, an $\mathrm{OA}(t,n,s)$ is a code $A \subset S^n$ with dual distance $t+1$.

$$S = \{a,b\}$$

Construction OA $\rightarrow$ TD :
- $X = S \times [1,n]$
- $\mathcal{G} = \{S \times \{i\}, 1 \leq i \leq n\}$
- $\mathcal{B} = \{\{(c_i,i), 1 \leq i \leq n\}, \boldsymbol{c} \in \mathrm{OA}\}$

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

$(a,1)$	$(a,2)$	$(a,3)$
$(b,1)$	$(b,2)$	$(b,3)$

Orthogonal arrays

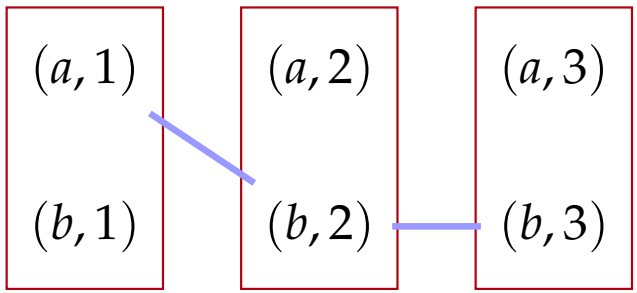
An **orthogonal array** $\mathrm{OA}(t,n,s)$ of strength t is a list A of words
- of length n,
- over a finite set S, $|S| = s$,
- such that, for every $I \subset [1,n]$ of size t, $A_{|I} = S^t$.

Equivalently, an $\mathrm{OA}(t,n,s)$ is a code $A \subset S^n$ with dual distance $t+1$.

$S = \{a,b\}$

Construction OA $\to$ TD :
- $X = S \times [1,n]$
- $\mathcal{G} = \{S \times \{i\}, 1 \le i \le n\}$
- $\mathcal{B} = \{\{(c_i,i), 1 \le i \le n\}, \boldsymbol{c} \in \mathrm{OA}\}$

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



Orthogonal arrays

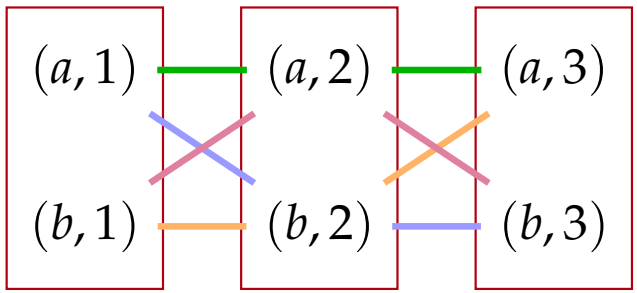
An **orthogonal array** $\mathrm{OA}(t,n,s)$ of strength t is a list A of words
- of length n,
- over a finite set S, $|S| = s$,
- such that, for every $I \subset [1,n]$ of size t, $A_{|I} = S^t$.

Equivalently, an $\mathrm{OA}(t,n,s)$ is a code $A \subset S^n$ with dual distance $t+1$.

$S = \{a,b\}$

Construction OA $\to$ TD :
- $X = S \times [1,n]$
- $\mathcal{G} = \{S \times \{i\}, 1 \le i \le n\}$
- $\mathcal{B} = \{\{(c_i,i), 1 \le i \le n\}, \boldsymbol{c} \in \mathrm{OA}\}$

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



Resisting collusions

Proposition. For $t = 2$, an $\mathrm{OA}(t, n, s)$ gives a $\mathrm{TD}(n, s)$.

Resisting collusions

Proposition. For $t = 2$, an $\mathrm{OA}(t, n, s)$ gives a $\mathrm{TD}(n, s)$.

Experimentally, for $t = 2$ and small n and s, codes based on classical affine TDs have the largest dimension.

Resisting collusions

Proposition. For $t = 2$, an $\mathrm{OA}(t, n, s)$ gives a $\mathrm{TD}(n, s)$.

Experimentally, for $t = 2$ and small n and s, codes based on classical affine TDs have the largest dimension.

For $t \geq 3$, we get TDs such that:

for every t-set T of points lying in t different groups,
there exists a unique block $B \in \mathcal{B}$ such that $T \subset B$.

Resisting collusions

Proposition. For $t = 2$, an $\mathrm{OA}(t, n, s)$ gives a $\mathrm{TD}(n, s)$.

Experimentally, for $t = 2$ and small n and s, codes based on classical affine TDs have the largest dimension.

For $t \geq 3$, we get TDs such that:

for every t-set T of points lying in t different groups,
there exists a unique block $B \in \mathcal{B}$ such that $T \subset B$.

$\Rightarrow$ The PIR protocol resists $t - 1$ colluding servers.

Resisting collusions

Proposition. For $t = 2$, an $\mathrm{OA}(t, n, s)$ gives a $\mathrm{TD}(n, s)$.

Experimentally, for $t = 2$ and small n and s, codes based on classical affine TDs have the largest dimension.

For $t \geq 3$, we get TDs such that:

for every t-set T of points lying in t different groups,
there exists a unique block $B \in \mathcal{B}$ such that $T \subset B$.

$\Rightarrow$ The PIR protocol resists $t - 1$ colluding servers.

- OAs with $t > 2$ exist (*e.g.* from Reed-Solomon codes)
- But associated TDs lead to codes with poor rates (except for $t \ll n$)

Private Information Retrieval from Transversal Designs. **L.**. IEEE-TIT. **2019**.

Outline

1. Private information retrieval

2. PIR schemes with low computation
   - Transversal designs and codes
   - A PIR scheme with transversal designs
   - Towards collusion resistance
   - PIR schemes with lifted codes

3. Other constructions of PIR schemes

4. Conclusion

Lifted codes

Definition. The (full-length) **Reed-Solomon code** of dimension k over $\mathbb{F}_q$ is:

$$\mathrm{RS}_q(k) := \{\mathrm{ev}_{\mathbb{A}^1}(f) := (f(x_1), \ldots, f(x_q)) \mid \deg(f) \leq k-1\}\,.$$

Reed-Muller codes = generalization to m-variate polynomials.

Lifted codes

Definition. The (full-length) **Reed-Solomon code** of dimension k over $\mathbb{F}_q$ is:

$$\mathrm{RS}_q(k) := \{\mathrm{ev}_{\mathbb{A}^1}(f) := (f(x_1), \ldots, f(x_q)) \mid \deg(f) \leq k-1\}\,.$$

Reed-Muller codes = generalization to m-variate polynomials.

Definition. The **m-th lifted Reed-Solomon code** of degree r over $\mathbb{F}_q$ is:

$$\mathrm{Lift}_q(m,r) := \{\mathrm{ev}_{\mathbb{A}^m}(f) \mid f \in \mathbb{F}_q[\mathbf{X}] \text{ and } \forall \text{ affine line } L \subset \mathbb{A}^m, \deg(f_{|L}) \leq r\}\,.$$

(where $f_{|L}$ is the lowest-degree univariate polynomial interpolating f over L)

Lifted codes

Definition. The (full-length) **Reed-Solomon code** of dimension k over $\mathbb{F}_q$ is:

$$\mathrm{RS}_q(k) := \{\mathrm{ev}_{\mathbb{A}^1}(f) := (f(x_1), \ldots, f(x_q)) \mid \deg(f) \leq k-1\}\,.$$

Reed-Muller codes = generalization to m-variate polynomials.

Definition. The **m-th lifted Reed-Solomon code** of degree r over $\mathbb{F}_q$ is:

$$\mathrm{Lift}_q(m,r) := \{\mathrm{ev}_{\mathbb{A}^m}(f) \mid f \in \mathbb{F}_q[\mathbf{X}] \text{ and } \forall \text{ affine line } L \subset \mathbb{A}^m, \deg(f_{|L}) \leq r\}\,.$$

(where $f_{|L}$ is the lowest-degree univariate polynomial interpolating f over L)

Lifted codes contain Reed-Muller codes, sometimes properly.

Example. For $q = 4, m = 2, r = 2$.

$$\mathrm{ev}(X^2Y^2) \in \mathrm{Lift}_4(2,2) \quad \text{but} \quad \mathrm{ev}(X^2Y^2) \notin \mathrm{RM}_4(2,2)$$

Weighted lifted codes

For convenience, here $m = 2$.

Definition. A t-curve is:

$$\mathcal{L} = \{(x, g(x)) \in \mathbb{A}^2 \mid g \in \mathbb{F}_q[X], \deg(g) \leq t\}$$

Weighted lifted codes

For convenience, here $m = 2$.

Definition. A t-curve is:

$$\mathcal{L} = \{(x, g(x)) \in \mathbb{A}^2 \mid g \in \mathbb{F}_q[X], \deg(g) \leq t\}$$

Definition. The **weighted lifted Reed-Solomon code** of degree r and weight t over $\mathbb{F}_q$ is:

$$\mathrm{WLift}_q(t, r) := \{\mathrm{ev}_{\mathbb{A}^2}(f) \mid f \in \mathbb{F}_q[X, Y] \text{ and } \forall\, t\text{-curve } \mathcal{L} \subset \mathbb{A}^2, \deg(f_{|\mathcal{L}}) \leq r\}$$

Weighted lifted codes

For convenience, here $m = 2$.

Definition. A t-curve is:

$$\mathcal{L} = \{(x, g(x)) \in \mathbb{A}^2 \mid g \in \mathbb{F}_q[X], \deg(g) \leq t\}$$

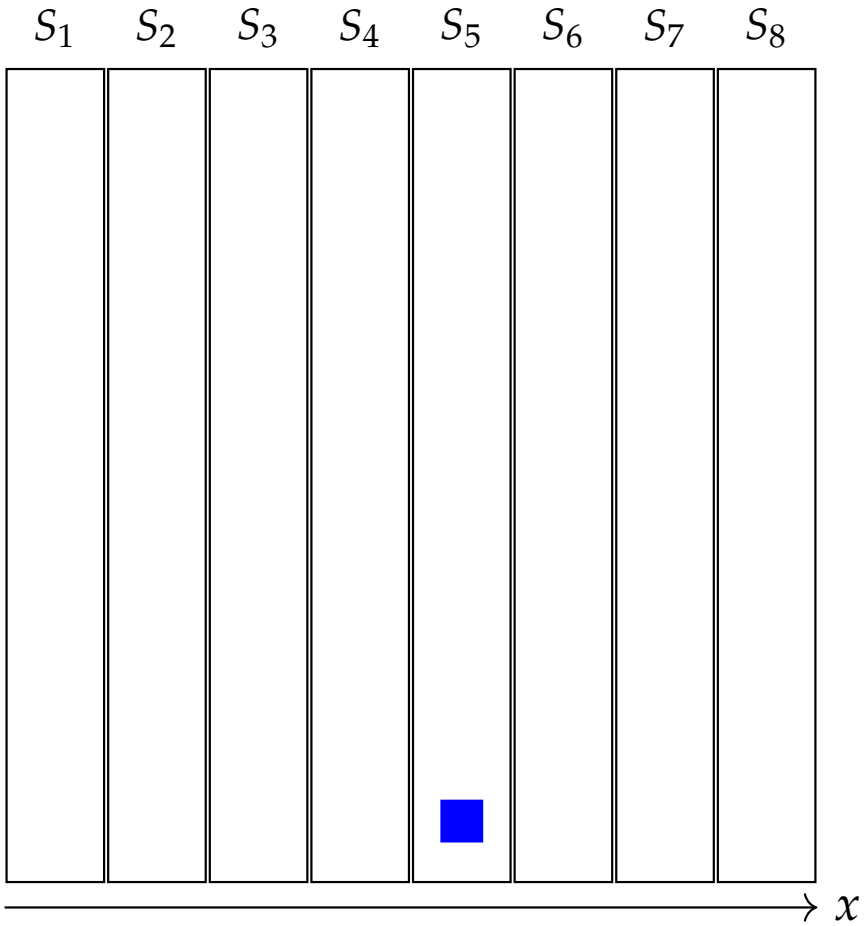
Definition. The **weighted lifted Reed-Solomon code** of degree r and weight t over $\mathbb{F}_q$ is:

$$\mathrm{WLift}_q(t, r) := \{\mathrm{ev}_{\mathbb{A}^2}(f) \mid f \in \mathbb{F}_q[X, Y] \text{ and } \forall\, t\text{-curve } \mathcal{L} \subset \mathbb{A}^2, \deg(f_{|\mathcal{L}}) \leq r\}$$

Consequence: for every $\boldsymbol{c} \in \mathrm{WLift}_q(t, r)$ and every t-curve $\mathcal{L}$, we have :

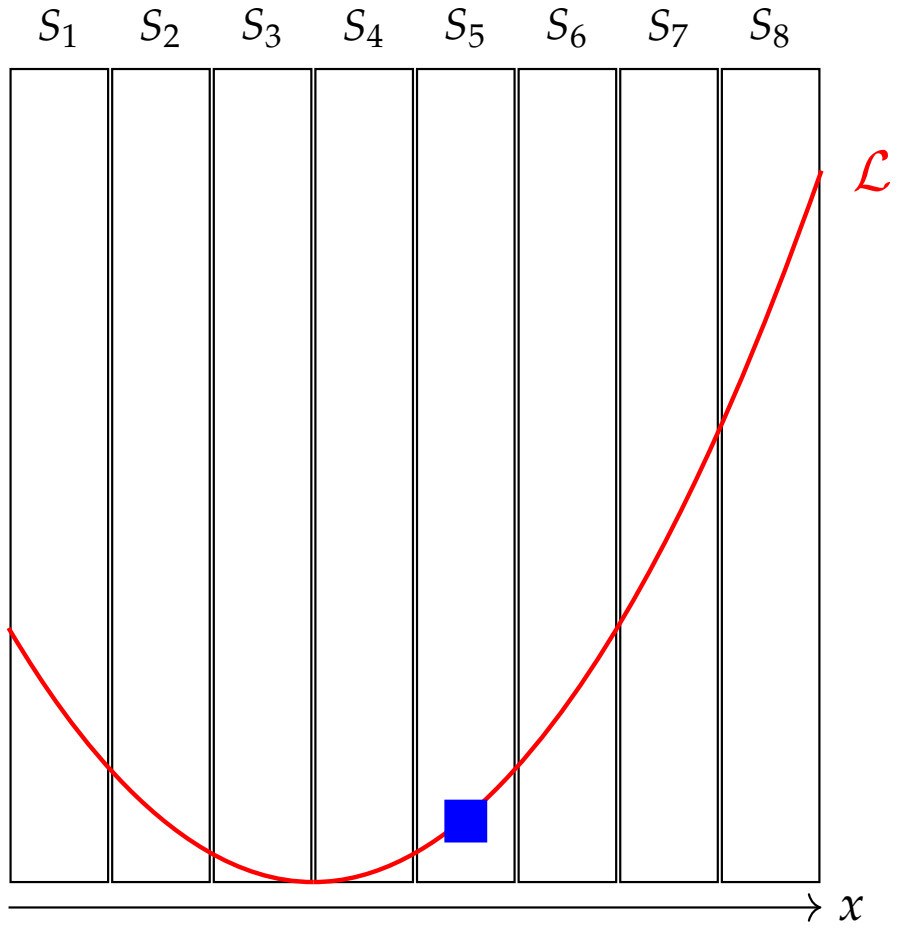
$$\boldsymbol{c}_{|\mathcal{L}} \in \mathrm{RS}_q(r)$$

A PIR scheme based on weighted lifted codes



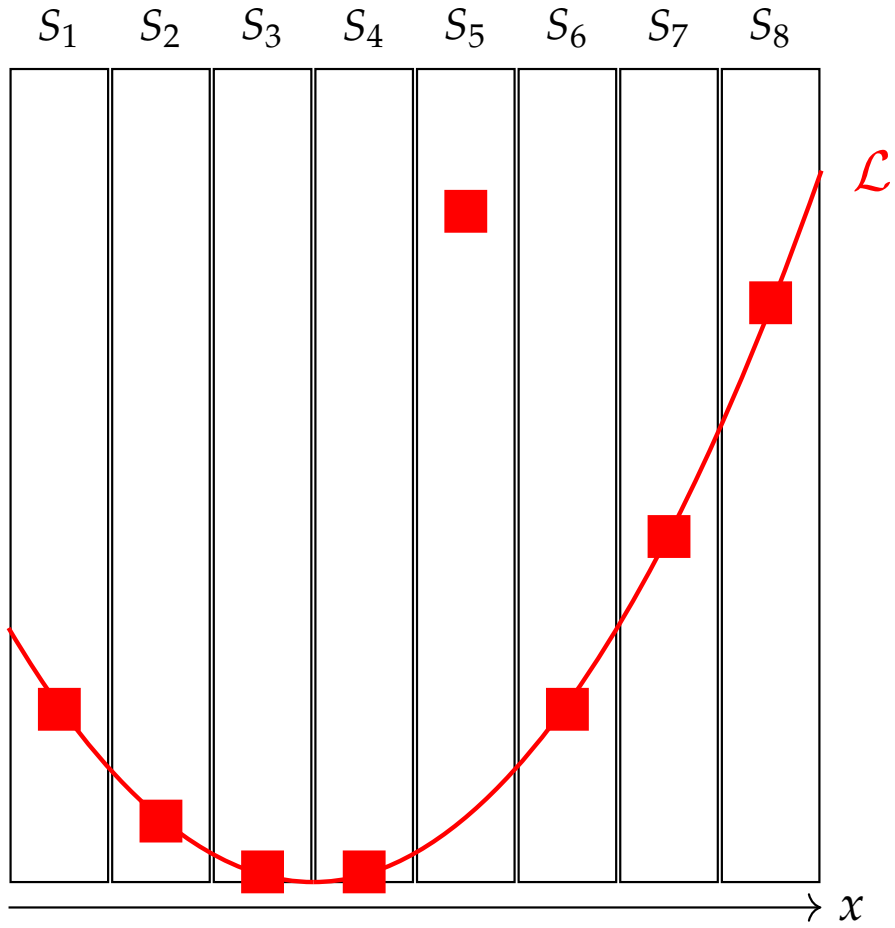
Database in encoded with $\mathrm{WLift}_q(t, r)$, then distributed across the servers

A PIR scheme based on weighted lifted codes



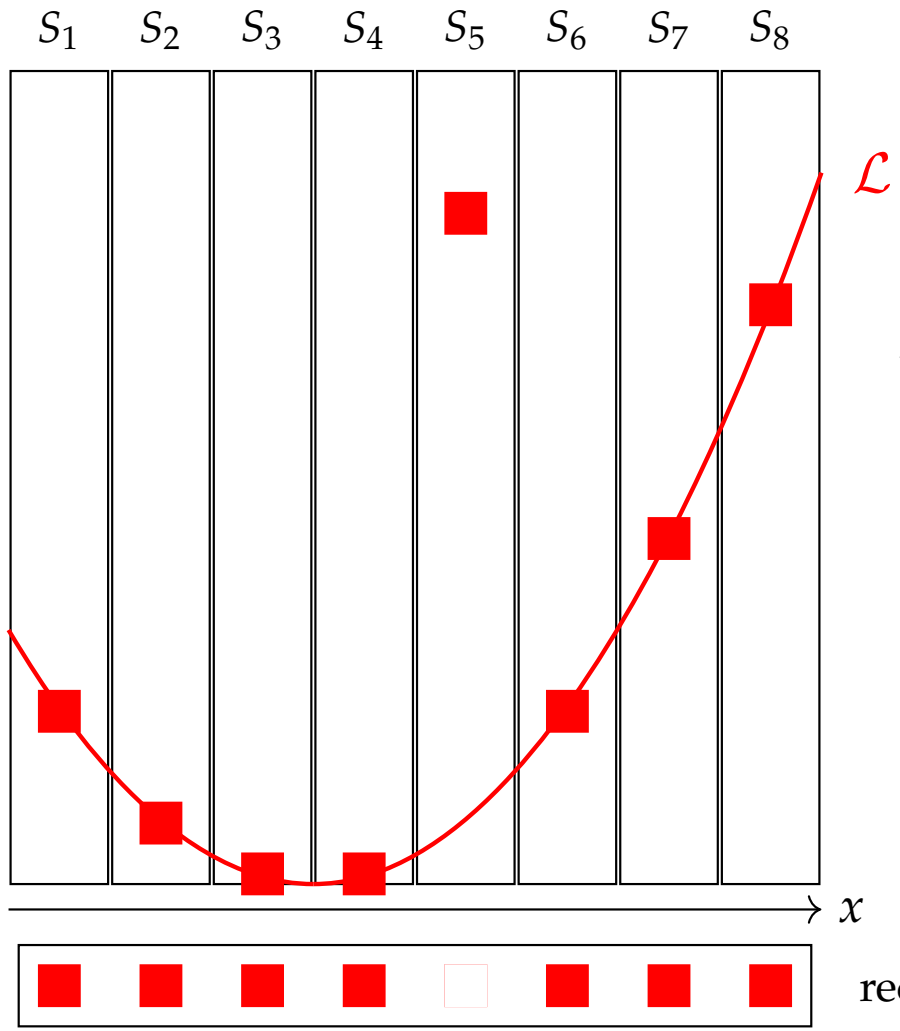
Database in encoded with $\mathrm{WLift}_q(t, r)$, then distributed across the servers

A PIR scheme based on weighted lifted codes



Database in encoded with $\mathrm{WLift}_q(t, r)$, then distributed across the servers

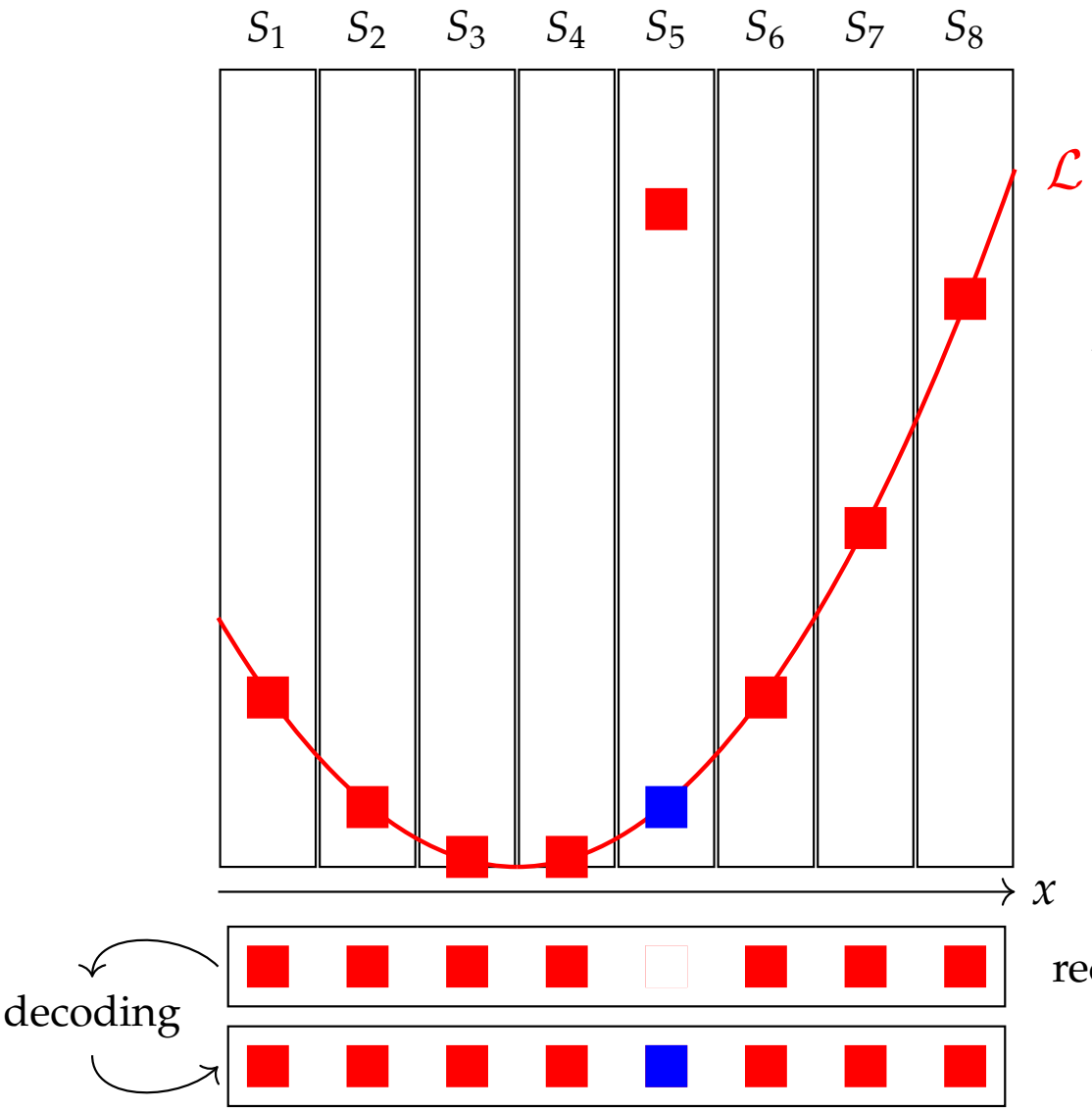
A PIR scheme based on weighted lifted codes



Database in encoded with $\mathrm{WLift}_q(t, r)$, then distributed across the servers

received answer

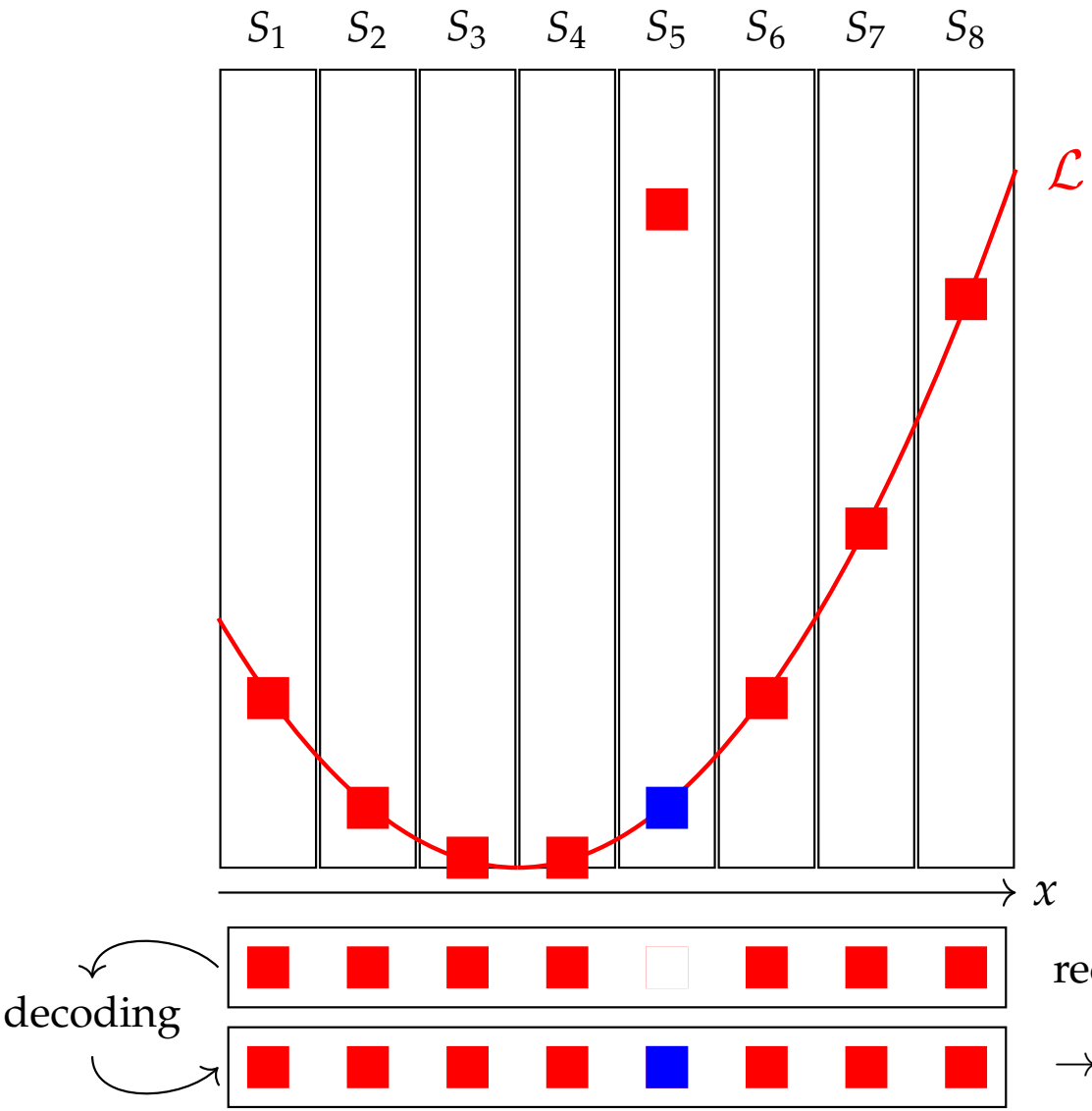
A PIR scheme based on weighted lifted codes



Database in encoded with $\mathrm{WLift}_q(t, r)$, then distributed across the servers

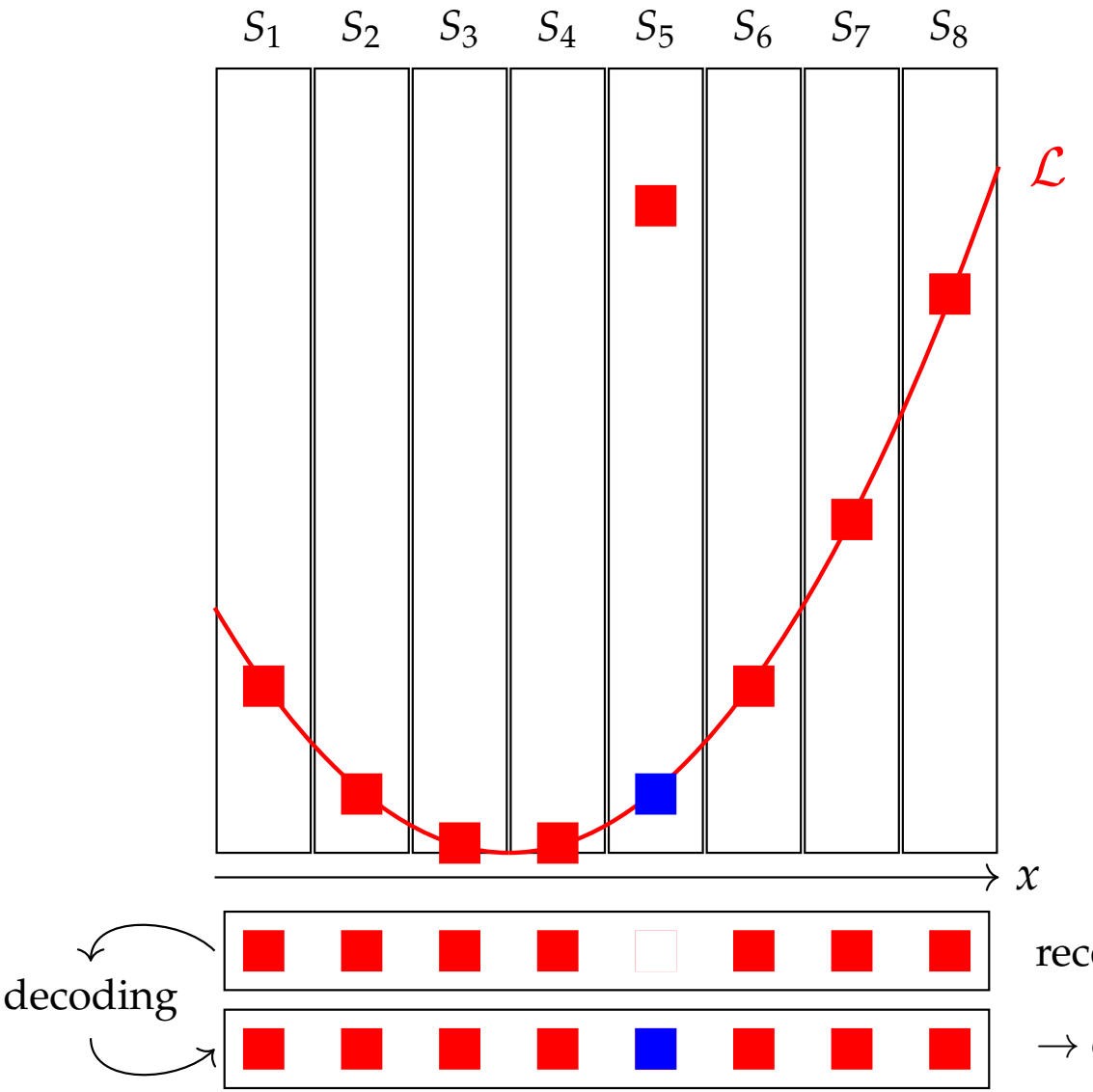
received answer

A PIR scheme based on weighted lifted codes



Database in encoded with $\mathrm{WLift}_q(t, r)$, then distributed across the servers

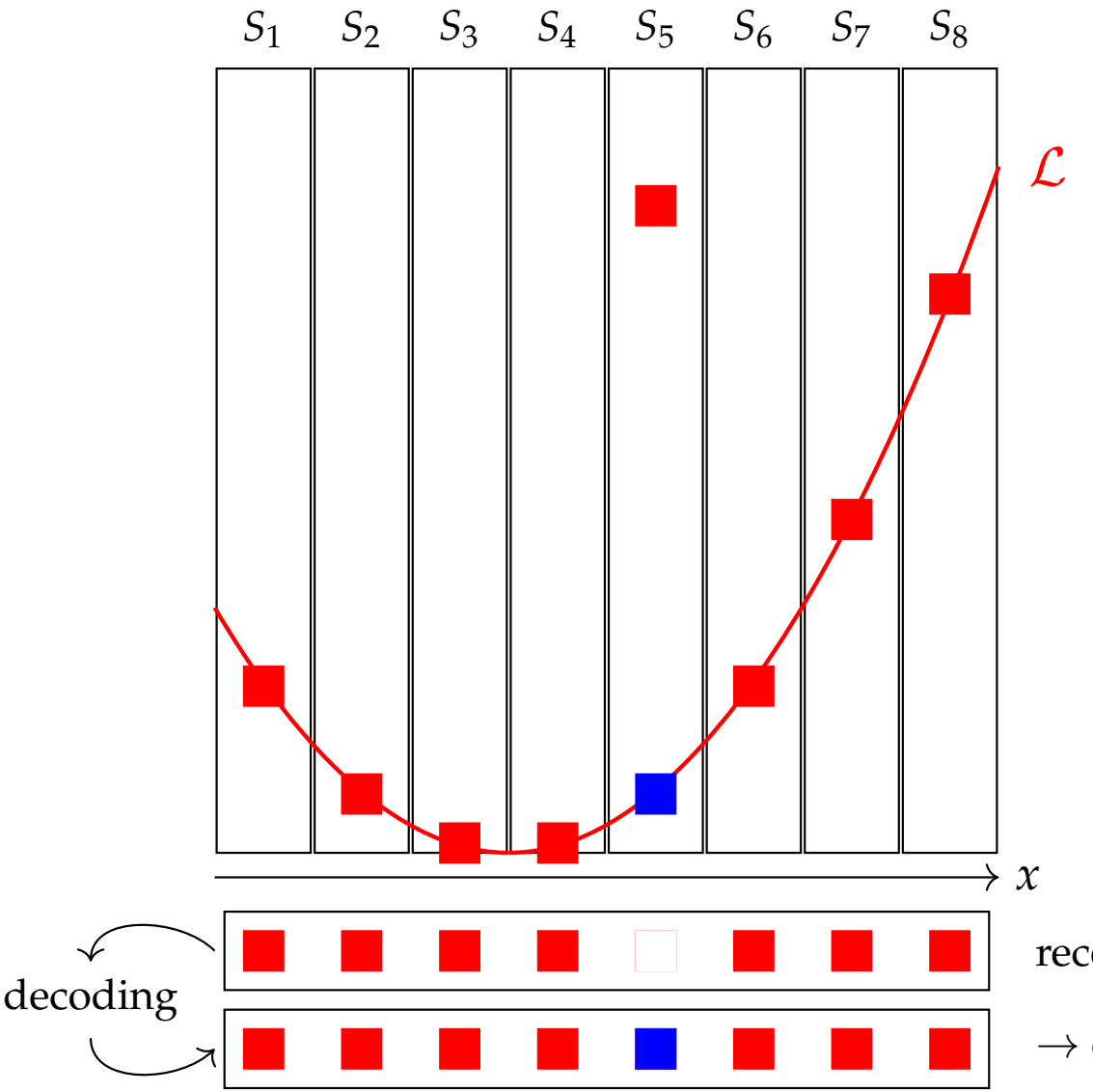
A PIR scheme based on weighted lifted codes



$\mathcal{L}$ is a t-curve

$\Longrightarrow$ no t-set of servers can find where is ■

A PIR scheme based on weighted lifted codes



$\mathcal{L}$ is a t-curve

$\Longrightarrow$ no t-set of servers can find where is ■

$\Longrightarrow$ the PIR scheme is t-private

Parameters

Weighted Lifted Codes: Local Correctabilities and Application to Robust Private Information Retrieval. **L.**, Nardi. arXiv:1904.08696. **2019**.

Theorem. Let p be a prime number, $t \geq 1$ and $\alpha \geq 2$ be fixed integers. Set $\mathcal{C}_e = \mathrm{WLift}_{p^e}(t, p^e - \alpha)$. Then, the information rate R_e of $\mathcal{C}_e$ grows to 1 when $e \to \infty$.

Corollary: PIR scheme with relative storage overhead $\to 0$,
for a constant number of adversaries

Parameters

Weighted Lifted Codes: Local Correctabilities and Application to Robust Private Information Retrieval. **L.**, Nardi. arXiv:1904.08696. **2019**.

Theorem. Let p be a prime number, $t \geq 1$ and $\alpha \geq 2$ be fixed integers. Set $\mathcal{C}_e = \mathrm{WLift}_{p^e}(t, p^e - \alpha)$. Then, the information rate R_e of $\mathcal{C}_e$ grows to 1 when $e \to \infty$.

Corollary: PIR scheme with relative storage overhead $\to 0$,
for a constant number of adversaries

Theorem. Let p be a prime number, $t \geq 1$ and $c \geq 1$ be fixed integers. Set $\gamma = 1 - p^{-c}$ and $\mathcal{C}'_e = \mathrm{WLift}_{p^e}(t, \gamma p^e)$. Then, the information rate R'_e of $\mathcal{C}'_e$ satisfies:

$$\lim_{e \to \infty} R'_e = K_{t,p,c} > 0$$

Corollary: PIR scheme with constant relative storage overhead,
for a constant number of adversaries and a constant fraction of errors

Outline

1. Private information retrieval

2. PIR schemes with low computation
   - Transversal designs and codes
   - A PIR scheme with transversal designs
   - Towards collusion resistance
   - PIR schemes with lifted codes

3. Other constructions of PIR schemes

4. Conclusion

A new PIR scheme on MBR codes

In previous schemes: **low computation**, but **moderate communication**.

Given a family of storage codes, **what is the lowest communication complexity** we can hope for a PIR protocol?

A new PIR scheme on MBR codes

In previous schemes: **low computation**, but **moderate communication**.

Given a family of storage codes, **what is the lowest communication complexity** we can hope for a PIR protocol?

Optimal constructions known for:

- repetition and parity-check codes [e.g. Sun–Jafar]
- any MDS code [e.g. Banawan–Ulukus]
- many other linear codes (some cyclic codes, etc.)

A new PIR scheme on MBR codes

In previous schemes: **low computation**, but **moderate communication**.

Given a family of storage codes, **what is the lowest communication complexity** we can hope for a PIR protocol?

Optimal constructions known for:

- repetition and parity-check codes [e.g. Sun–Jafar]
- any MDS code [e.g. Banawan–Ulukus]
- many other linear codes (some cyclic codes, etc.)

Regenerating codes: better storage systems for repairing node failures.

A new PIR scheme on MBR codes

In previous schemes: **low computation**, but **moderate communication**.

Given a family of storage codes, **what is the lowest communication complexity** we can hope for a PIR protocol?

Optimal constructions known for:

- repetition and parity-check codes [e.g. Sun–Jafar]
- any MDS code [e.g. Banawan–Ulukus]
- many other linear codes (some cyclic codes, etc.)

Regenerating codes: better storage systems for repairing node failures.

Result. A new PIR protocol featuring low communication for storage systems using MBR codes.

Private Information Retrieval Schemes with Product-Matrix MBR Codes. **L.**, Tajeddine, Freij-Hollanti, Hollanti. submitted. **2019**.

Main assumption of previous schemes:

$$n \geq 2 \text{ servers, not all colluding}$$

Single server PIR schemes

Main assumption of previous schemes:

$$n \geq 2 \text{ servers, not all colluding}$$

Computational PIR:

- only 1 server
- distinguishing index i is computationally hard for the server.

Single server PIR schemes

Main assumption of previous schemes:

$$n \geq 2 \text{ servers, not all colluding}$$

Computational PIR:

- only 1 server
- distinguishing index i is computationally hard for the server.

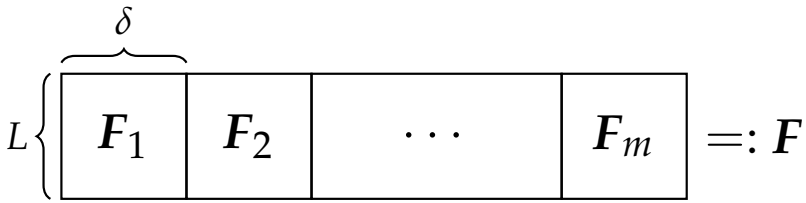
Very recently:

Computational Code-Based Single-Server Private Information Retrieval. Holzbaur, Hollanti, Wachter-Zeh. arXiv:2001.07049. **2020**.

Analysis of a code-based PIR scheme: model and query

System model.

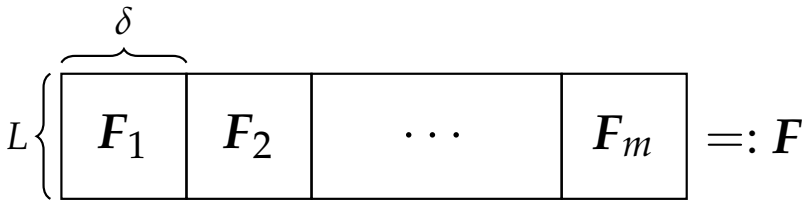
Entry $\boldsymbol{F}_j$ of the database $\boldsymbol{F}$ is an $(L \times \delta)$ matrix over $\mathbb{F}_q$.



Analysis of a code-based PIR scheme: model and query

System model.

Entry $\boldsymbol{F}_j$ of the database $\boldsymbol{F}$ is an $(L \times \delta)$ matrix over $\mathbb{F}_q$.



Query generation.

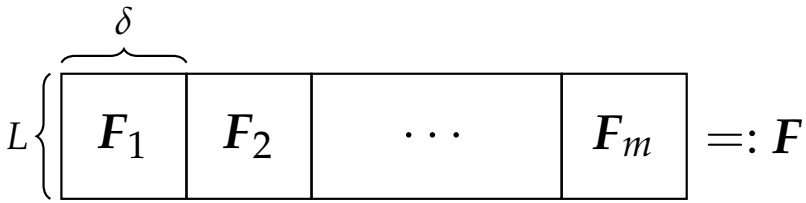
The user chooses at random:

- a code $\mathcal{C} \subseteq \mathbb{F}_{q^s}^n$ of dimension k,
- an information set $\mathcal{I} \subset [1, n]$ for $\mathcal{C}$,
- a basis $\{\gamma_1, \ldots, \gamma_s\}$ of $\mathbb{F}_{q^s}/\mathbb{F}_q$, and sets $V := \langle \gamma_1, \ldots, \gamma_v \rangle_{\mathbb{F}_q}$ and $W := \langle \gamma_{v+1}, \ldots, \gamma_s \rangle_{\mathbb{F}_q}$,

Analysis of a code-based PIR scheme: model and query

System model.

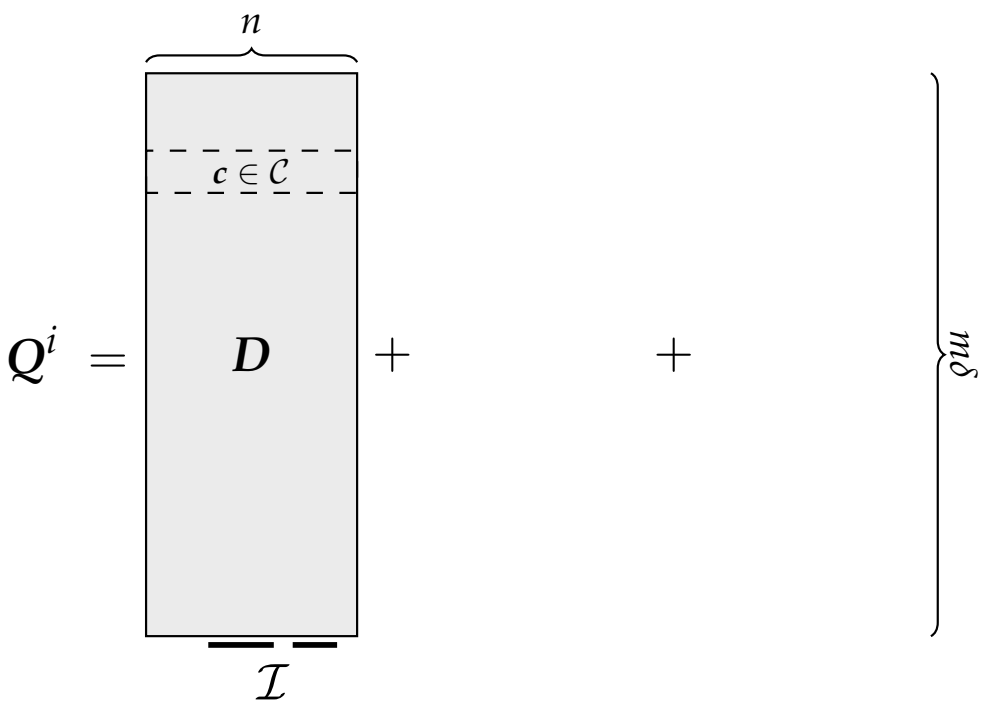
Entry $\boldsymbol{F}_j$ of the database $\boldsymbol{F}$ is an $(L \times \delta)$ matrix over $\mathbb{F}_q$.



Query generation.

The user chooses at random:

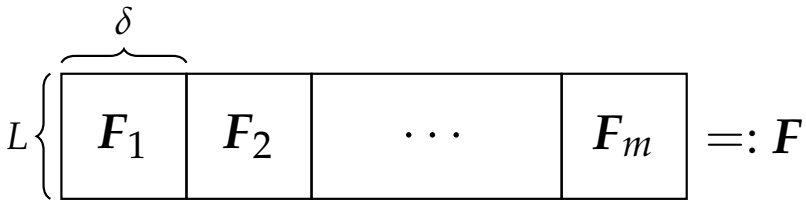
- a code $\mathcal{C} \subseteq \mathbb{F}_{q^s}^n$ of dimension k,
- an information set $\mathcal{I} \subset [1, n]$ for $\mathcal{C}$,
- a basis $\{\gamma_1, \ldots, \gamma_s\}$ of $\mathbb{F}_{q^s}/\mathbb{F}_q$, and sets $V := \langle \gamma_1, \ldots, \gamma_v \rangle_{\mathbb{F}_q}$ and $W := \langle \gamma_{v+1}, \ldots, \gamma_s \rangle_{\mathbb{F}_q}$,
- matrices $\boldsymbol{D} \in \mathbb{F}_{q^s}^{m\delta \times n}$, $\boldsymbol{E} \in V^{m\delta \times n}$ and $\boldsymbol{Z}^i \in W^{m\delta \times n}$ as follows:



Analysis of a code-based PIR scheme: model and query

System model.

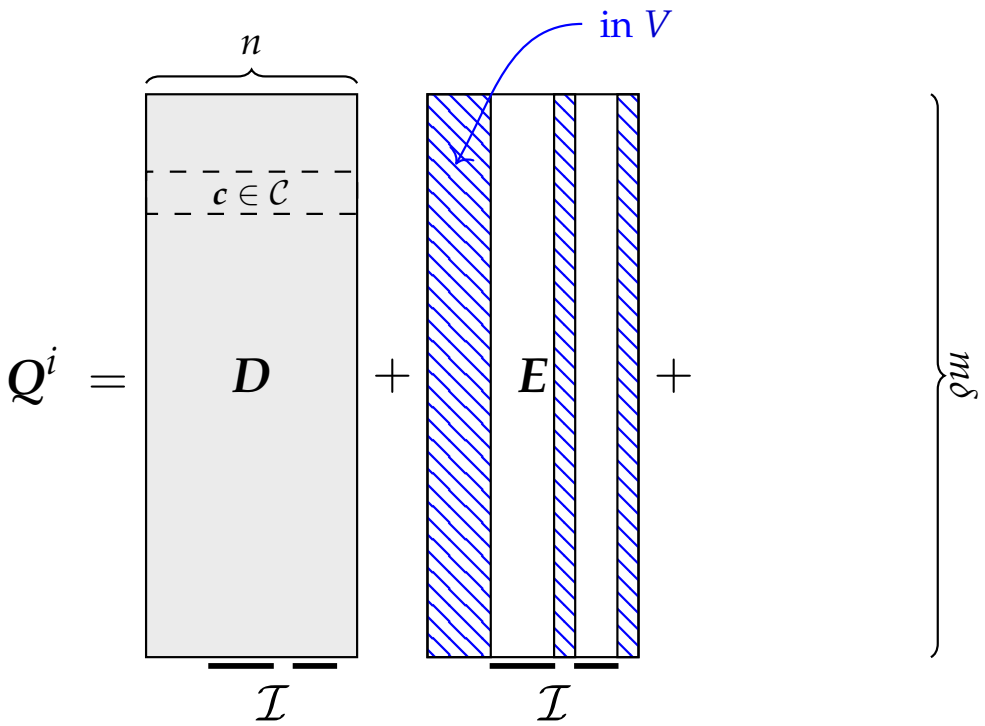
Entry $\boldsymbol{F}_j$ of the database $\boldsymbol{F}$ is an $(L \times \delta)$ matrix over $\mathbb{F}_q$.



Query generation.

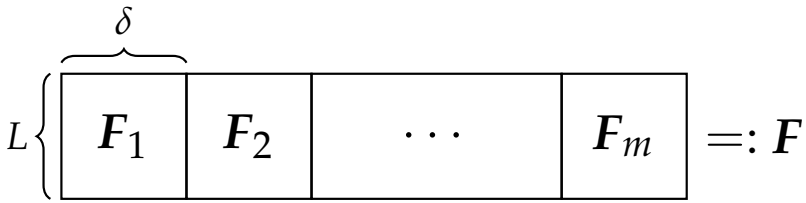
The user chooses at random:

- a code $\mathcal{C} \subseteq \mathbb{F}_{q^s}^n$ of dimension k,
- an information set $\mathcal{I} \subset [1, n]$ for $\mathcal{C}$,
- a basis $\{\gamma_1, \ldots, \gamma_s\}$ of $\mathbb{F}_{q^s}/\mathbb{F}_q$, and sets $V := \langle \gamma_1, \ldots, \gamma_v \rangle_{\mathbb{F}_q}$ and $W := \langle \gamma_{v+1}, \ldots, \gamma_s \rangle_{\mathbb{F}_q}$,
- matrices $\boldsymbol{D} \in \mathbb{F}_{q^s}^{m\delta \times n}$, $\boldsymbol{E} \in V^{m\delta \times n}$ and $\boldsymbol{Z}^i \in W^{m\delta \times n}$ as follows:



System model.

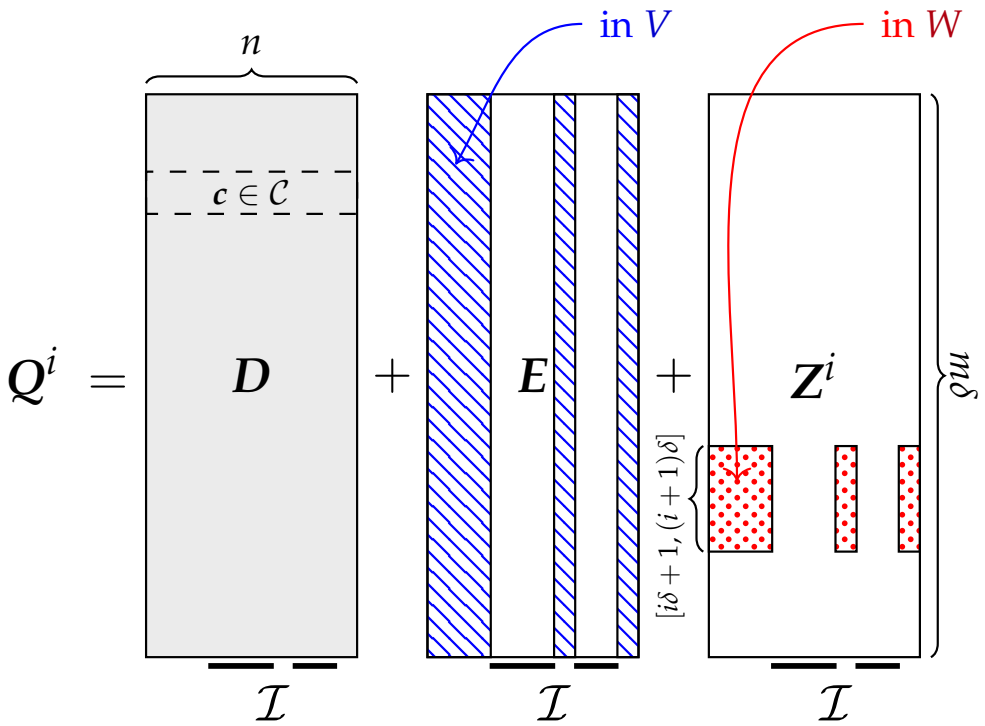
Entry $\boldsymbol{F}_j$ of the database $\boldsymbol{F}$ is an $(L \times \delta)$ matrix over $\mathbb{F}_q$.



Query generation.

The user chooses at random:

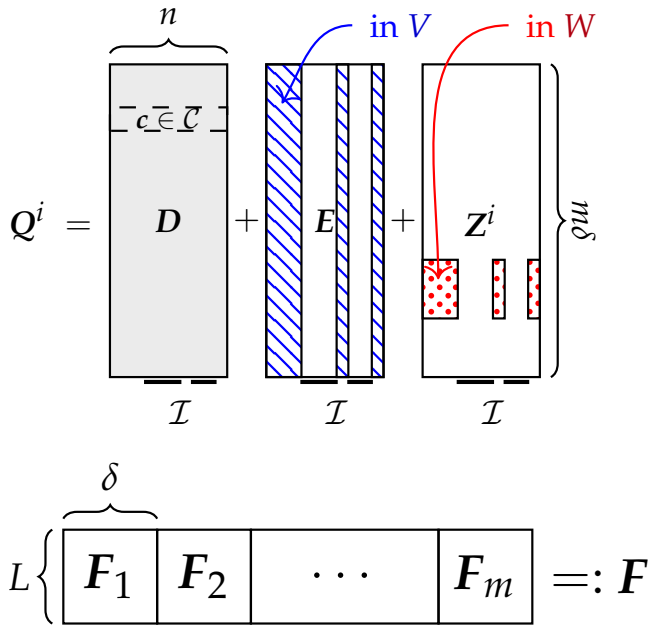
- a code $\mathcal{C} \subseteq \mathbb{F}_{q^s}^n$ of dimension k,
- an information set $\mathcal{I} \subset [1, n]$ for $\mathcal{C}$,
- a basis $\{\gamma_1, \ldots, \gamma_s\}$ of $\mathbb{F}_{q^s}/\mathbb{F}_q$, and sets $V := \langle \gamma_1, \ldots, \gamma_v \rangle_{\mathbb{F}_q}$ and $W := \langle \gamma_{v+1}, \ldots, \gamma_s \rangle_{\mathbb{F}_q}$,
- matrices $\boldsymbol{D} \in \mathbb{F}_{q^s}^{m\delta \times n}$, $\boldsymbol{E} \in V^{m\delta \times n}$ and $\boldsymbol{Z}^i \in W^{m\delta \times n}$ as follows:



Response.

The server computes

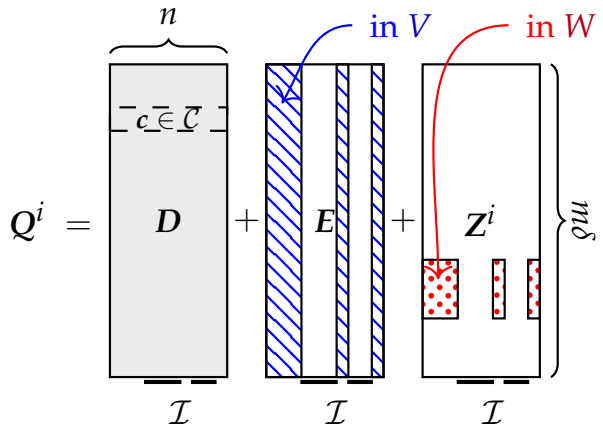
$$\boldsymbol{A}^i := \boldsymbol{F} \cdot \boldsymbol{Q}^i \in \mathbb{F}_{q^s}^{L \times n}$$



Response.

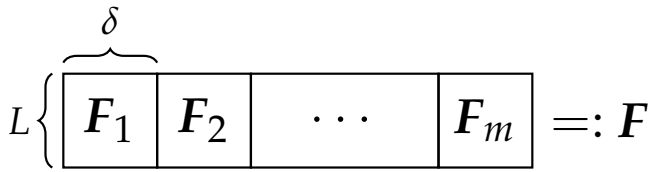
The server computes

$$A^i := F \cdot Q^i \in \mathbb{F}_{q^s}^{L \times n}$$



Decoding.

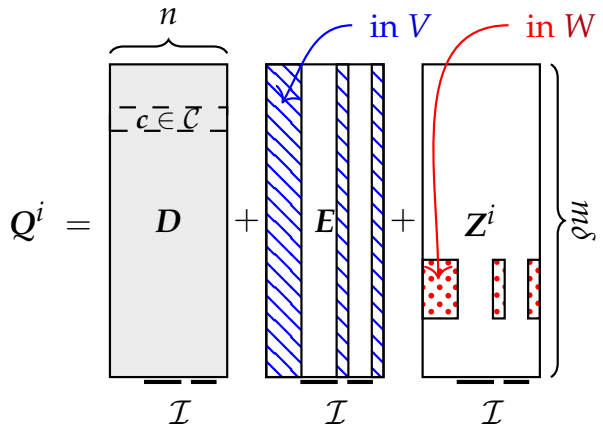
$$A^i = \sum_{r=1}^{m} F_r \cdot Q_r^i = \underbrace{\sum_{r=1}^{m} F_r \cdot D_r}_{\text{rows in } \mathcal{C}} + \underbrace{\sum_{r=1}^{m} F_r \cdot (E_r + Z_r^i)}_{\text{zero on } \mathcal{I}} .$$



Response.

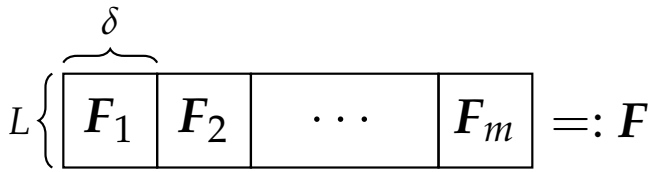
The server computes

$$A^i := F \cdot Q^i \in \mathbb{F}_{q^s}^{L \times n}$$



Decoding.

$$A^i = \sum_{r=1}^{m} F_r \cdot Q_r^i = \underbrace{\sum_{r=1}^{m} F_r \cdot D_r}_{\text{rows in } \mathcal{C}} + \underbrace{\sum_{r=1}^{m} F_r \cdot (E_r + Z_r^i)}_{\text{zero on } \mathcal{I}} .$$

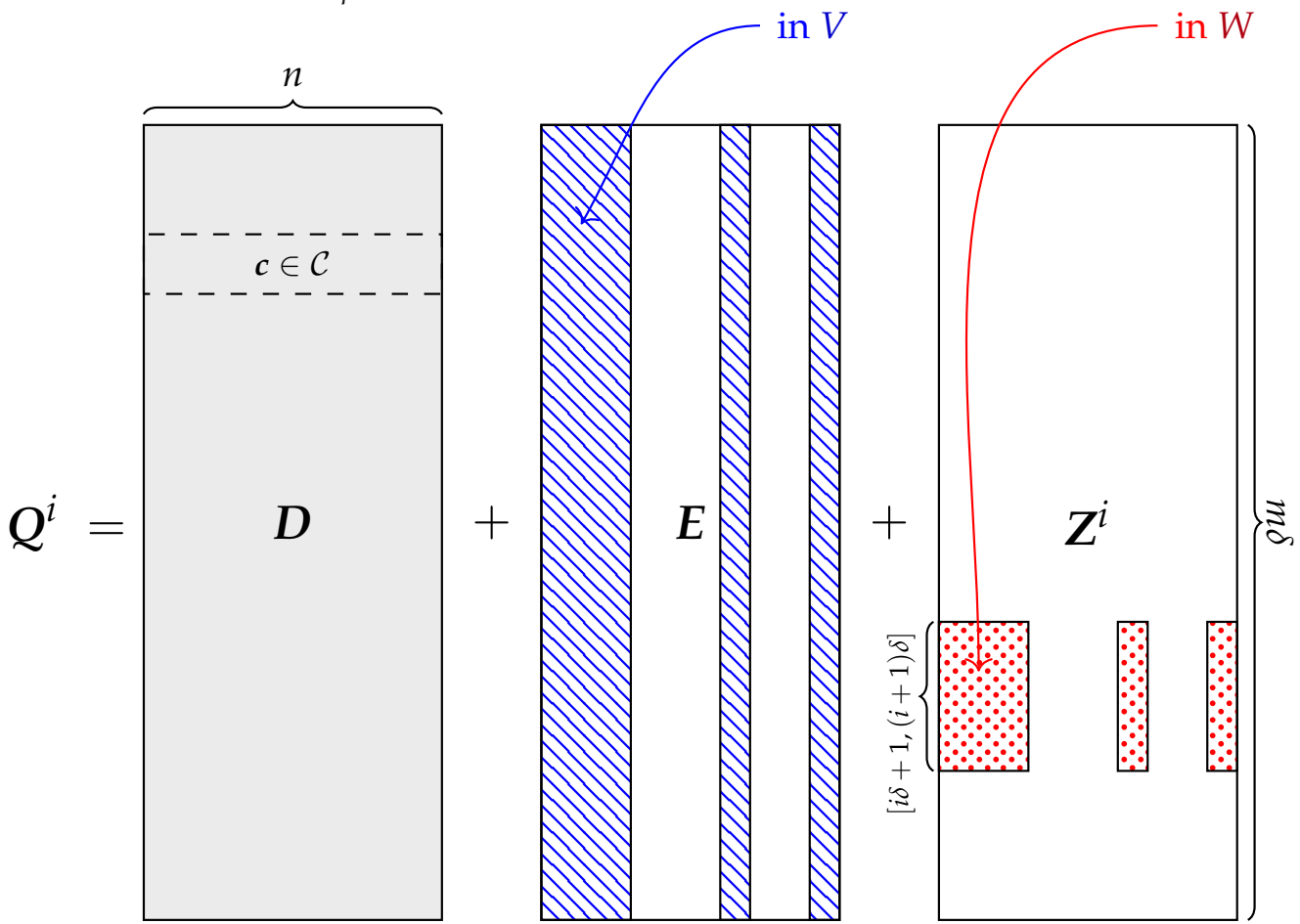


One gets:

$$\sum_{r=1}^{m} F_r \cdot (E_r + Z_r^i) = \underbrace{\left(\sum_{r=1}^{m} F_r \cdot E_r \right)}_{\text{rows in } V^n} + \underbrace{F_i \cdot Z_i^i}_{\text{rows in } W^n} .$$

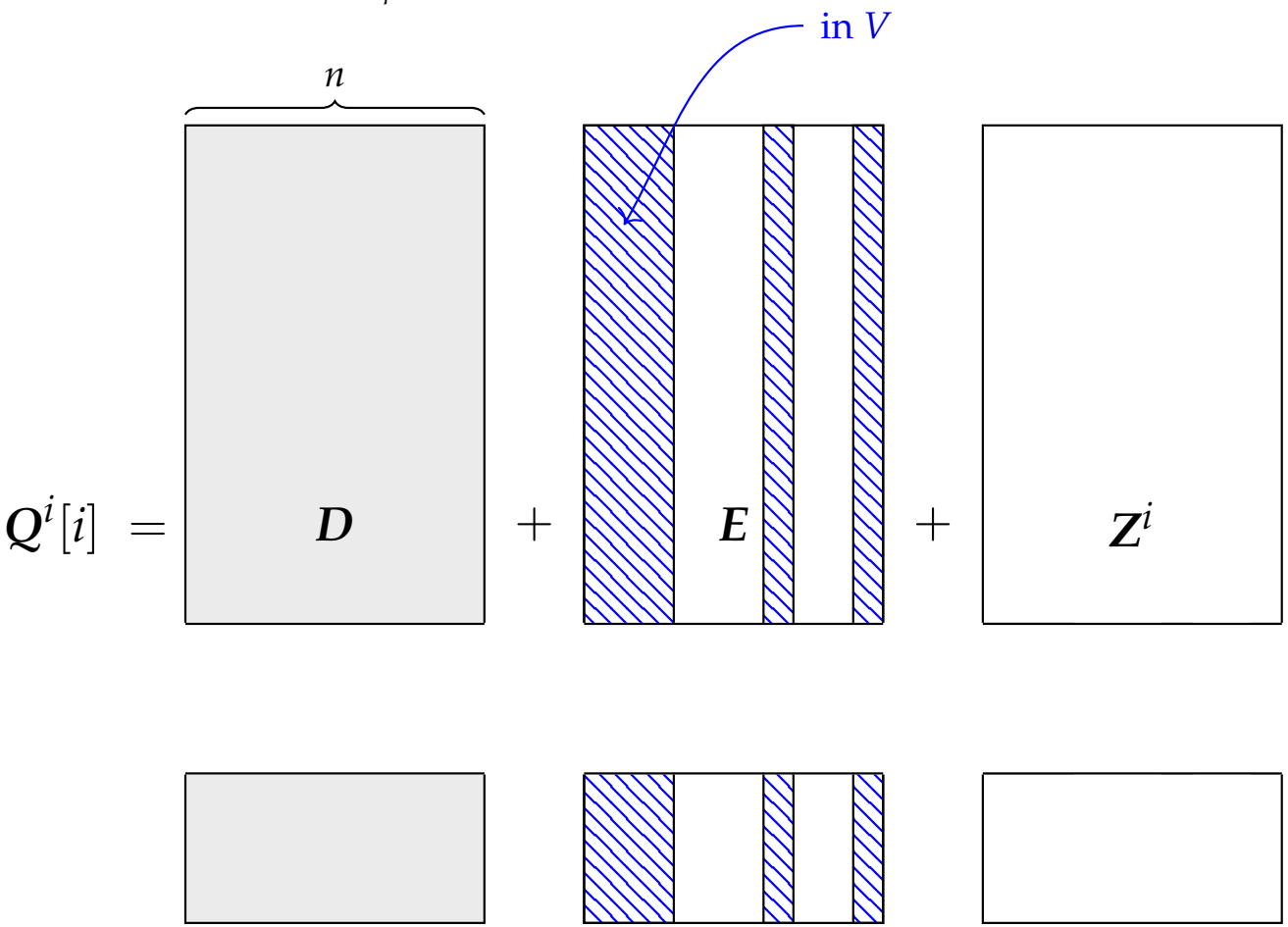
Analysis of a code-based PIR scheme: attack

$\mathbb{F}_{q^s} = V \oplus W, \quad \dim_{\mathbb{F}_q} V = v$



Analysis of a code-based PIR scheme: attack

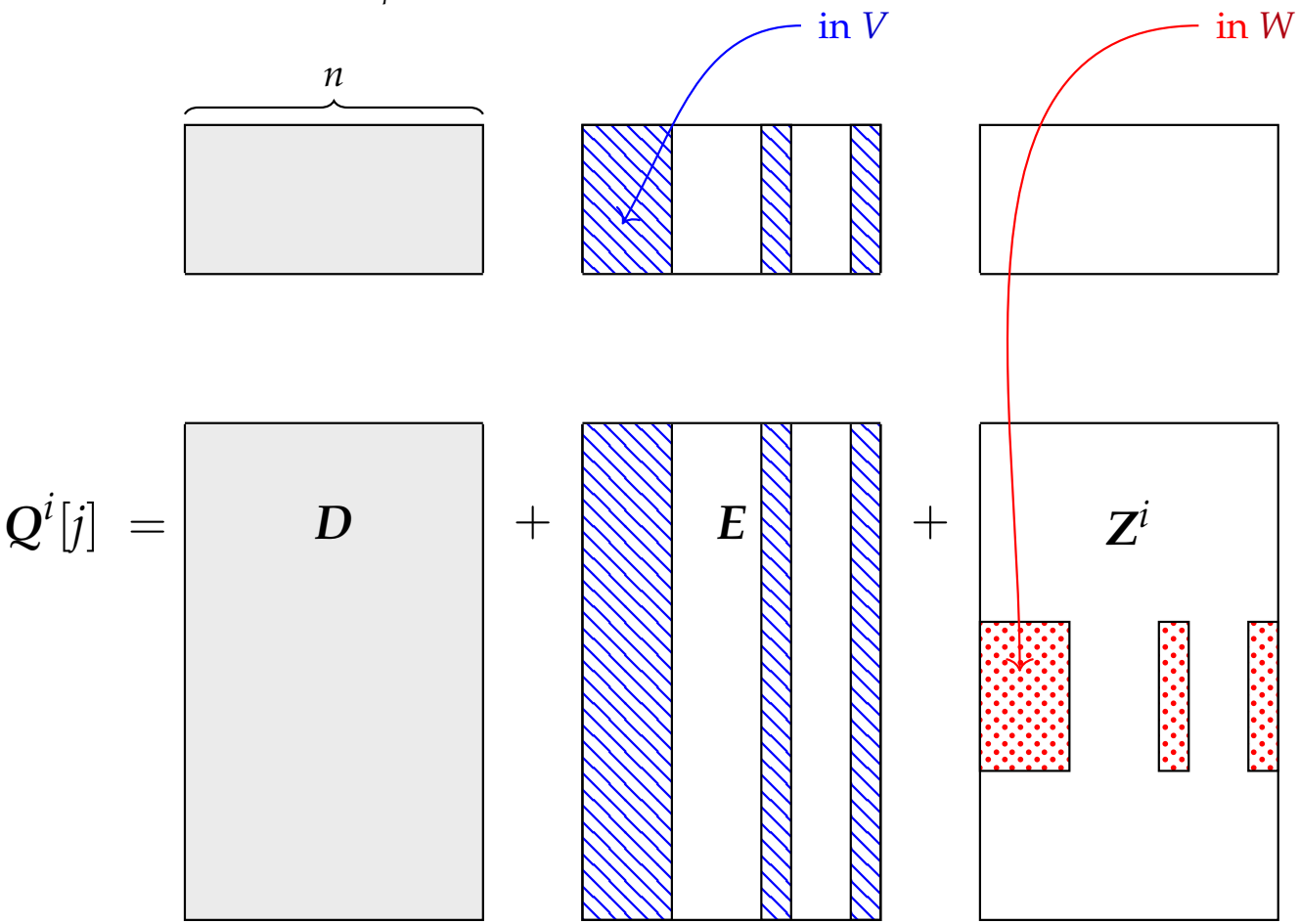
$$\mathbb{F}_{q^s} = V \oplus W, \quad \dim_{\mathbb{F}_q} V = v$$



$$\mathrm{rk}_{\mathbb{F}_q}(Q^i[i]) \leq ks + (n-k)v$$

Analysis of a code-based PIR scheme: attack

$\mathbb{F}_{q^s} = V \oplus W, \quad \dim_{\mathbb{F}_q} V = v$



$\mathrm{rk}_{\mathbb{F}_q}(Q^i[j]) = ns$ w.h.p. if m is large enough

Outline

1. Private information retrieval

2. PIR schemes with low computation
   - Transversal designs and codes
   - A PIR scheme with transversal designs
   - Towards collusion resistance
   - PIR schemes with lifted codes

3. Other constructions of PIR schemes

4. Conclusion

Conclusion

Private information retrieval:

- concentrated a lot of recent research,
- involves various mathematical tools,
- but there remains a lot of work (questionable assumptions, optimal constructions, other contexts)

Conclusion

Private information retrieval:

- concentrated a lot of recent research,
- involves various mathematical tools,
- but there remains a lot of work (questionable assumptions, optimal constructions, other contexts)

Questions?