Topic II: PIR's and codes

SecureCAT | Kick-off workshop

Aguilar de Campoo, Jan. 22nd-26th 2023

E. Martínez-Moro

imUVa Instituto de Investigación en Matemáticas

Universidad de Valladolid

Private Information Retrieval

#SecureCAT #Kick-off workshop.

secureCAT

Coding theory and Algebraic Trends for Cryptography, Distributed Data Storage, Machine Learning and Quantum Information

Supported by Grant TED2021-130358B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR"

PRIVATE INFORMATION RETRIEVAL (PIR)

GIVEN A REMOTE DATABASE $D = (D_1, \ldots, D_M) \in \Sigma^M$

AND AN INDEX $i \in 1, \ldots, M$

CAN WE RETRIEVE THE ENTRY (FILE) D_i

WITHOUT LEAKING ANY INFORMATION ON i ?

TRIVIAL SOLUTION: FULL DOWNLOAD !

1ST INTRODUCED IN

CHOR, GOLDREICH, KUSHILEVITZ, SUDAN. PRIVATE INFORMATION RETRIEVAL. FOCS, 1995.

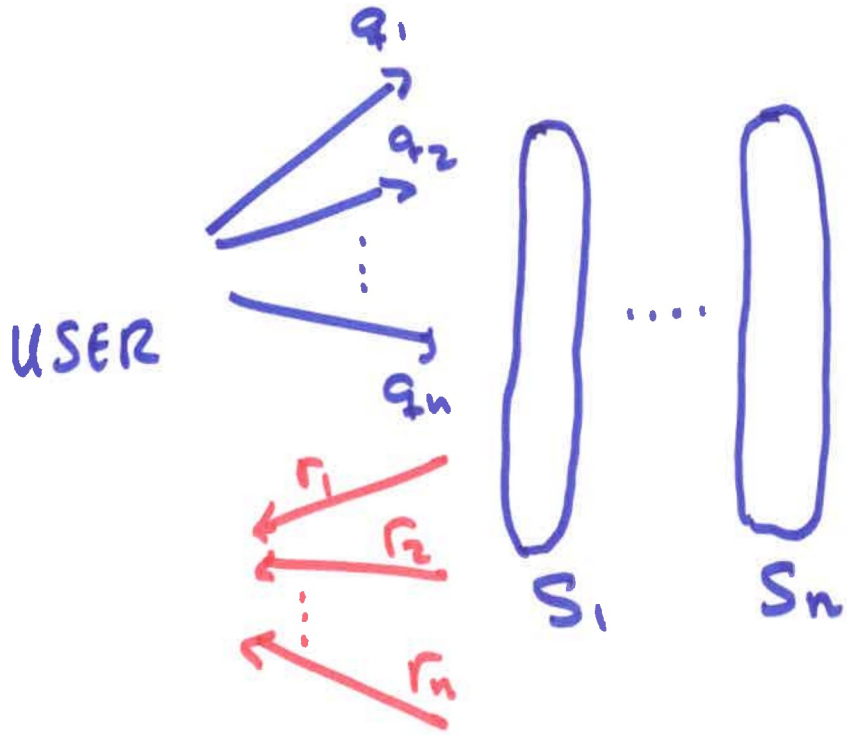
A PRIVATE INFORMATION RETRIEVAL PROTOCOL IS A SET OF ALGORITHMS (Q, Δ, R) S.T.

USER

1) GENERATES A QUERY

$q = (q_1, \ldots, q_n) \leftarrow Q(i)$

AND SENDS q_j TO SERVER S_j



2) EACH SERVER COMPUTES

$r_j = \Delta(q_j, D|_{S_j})$ AND

SENDS IT BACK TO USER.

3) USER RECOVERS

$D_i = R(q, r, i)$

WE SAY THAT T SERVERS COLLUDE (COLLUSION OF SERVERS) IF $\{S_i \mid i \in T\}$, $T \subset [1,n]$ EXCHANGE INFORMATION, QUERIES, DATA ...

$$t = \max \{ |T| \mid T \subseteq [1,n] \text{ IS A COLLUSION} \} \geq 1$$

- INFORMATION-THEORETICAL PRIVACY

$$I(i; q|_T) = 0 \quad \forall T \subseteq [1,n] \quad |T| \leq t.$$

- COMPUTATIONAL PRIVACY.

  BY VARYING i, THE DISTRIBUTION OF $q|_T = Q(i)|_T$ ARE COMPUTATIONALLY INDISTINGUISHABLE.

THEOREM IF $t = n$ (THUS IN PARTICULAR IF $n = 1$)

- FOR I.T. PRIVACY, NO BETTER SOLUTION THAN FULL DOWNLOAD.
- COMPUTATIONAL PRIVACY IS POSSIBLE BUT EXPENSIVE.

PARAMETERS TO TAKE INTO ACCOUNT:

→ COMMUNICATION COMPLEXITY $\uparrow, \downarrow$ U, D

→ COMPUTATION COMPLEXITY (CLIENTS SERVERS)

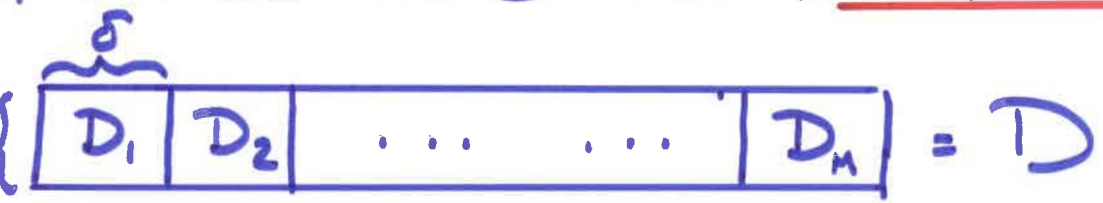
→ SERVER STORAGE

→ SIZE OF COLLECTIONS

SINGLE SERVER PIR SCHEME

HOLZBAUR, HOLLANTI, WACHTER-ZEH. COMPUTATIONAL CODE-BASED SINGLE SERVER PIR. arXiv : 2001.07049 (2020)

ENTRY D_j OF THE DATABASE $\mathcal{D}$ IS AN $(L \times \delta)$ MATRIX OVER $\mathbb{F}_q$.



QUERY GENERATION : THE USER CHOOSES AT RANDOM

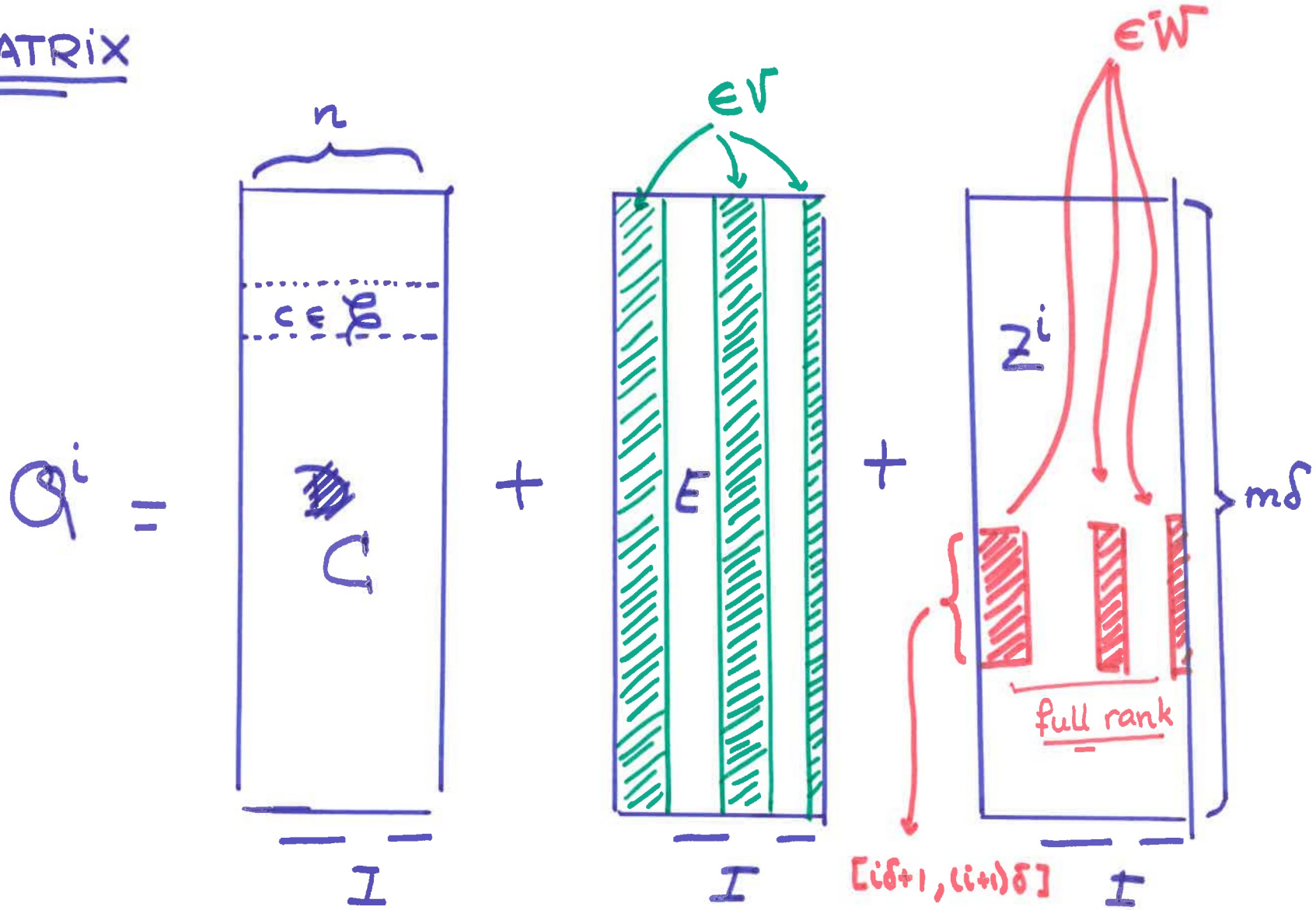
→ A CODE $\mathcal{C} \subseteq \mathbb{F}_{q^s}^n$ OF DIMENSION k

→ AN INFORMATION SET $\mathcal{I} \subset [1,n]$ FOR $\mathcal{C}$

→ A BASIS $\{\gamma_1, \dots, \gamma_s\}$ OF $\mathbb{F}_{q^s} | \mathbb{F}_q$ AND

$$V = \langle \gamma_1, \dots, \gamma_t \rangle_{\mathbb{F}_q} \quad , \quad W = \langle \gamma_{t+1}, \dots, \gamma_s \rangle_{\mathbb{F}_q}$$

QUERY MATRIX



RESPONSE

THE SERVER COMPUTES

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

DECODING

$$A^i = \sum_{r=1}^{m} D_r Q_r^i = \overbrace{\sum_{r=1}^{m} D_r C_r}^{\text{rows in } G} + \overbrace{\sum_{r=1}^{m} D_r (E_r + Z_r^i)}^{\text{zero at } I}$$

$$\sum_{r=1}^{m} D_r E_r + D_i Z_i^i$$

$\sum_{r=1}^{m} D_r E_r$: rows in V^n

$D_i Z_i^i$: rows in W^n

AS THE POSITIONS IN I ARE AN INFORMATION SET OF $\mathcal{C}$ (KNOWN BY THE USER) THEN THE $\mathcal{C}$ MATRIX

$$\sum_{r=1}^{m} D_r \cdot C_r$$

CAN BE RECOVERED. HENCE

$$A^i - \sum_{r=1}^{m} D_r C_r = \sum_{r=1}^{m} D_r E_r + D_i Z_i^i$$

Ψ function separating V and W

$D_i Z_i^i$

Z_i^i full rank

D_i

ANALYSIS

- UPLOAD → $H(Q^i) = m\delta \log_2(q^s) = m\delta s \log_2(q)$
- DOWNLOAD → $H(A^i) = Ln \log_2(q^s) = Lns \log_2(q)$
- RATE →

$$R_{PIR} = \frac{L\delta \log_2(q)}{m\delta ns \log_2(q) + Lns \log_2(q)} = \frac{L}{m\delta + L}\left(1 - \frac{k + \frac{v}{s}(n-k)}{n}\right)$$

$L \gg \delta m \quad R_{PIR} \approx 1 - \frac{k + \frac{v}{s}(n-k)}{n}$

$\mathcal{C} = [n, k]_q$ code

$|v| = k$

Attack

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

and

$$\mathrm{rk}_{\mathbb{F}_q}(Q^i[j]) = ns \quad \text{with high probability if } m \text{ is large enough.}$$

$i \neq j$

A PIR SCHEME BASED ON LIFTED CODES

RS-CODE

$$RS_q(k) = \left\{ ev_{\mathbb{A}^1}(f) = (f(x_1), \ldots, f(x_q)) \mid \deg(f) \leq k-1 \right\}$$

m-LIFTED RS CODE

$$LIFT_q(m,r) = \left\{ ev_{\mathbb{A}^m}(f) \mid f \in \mathbb{F}_q[X] \text{ and } \right.$$
$$\left. \text{for each } L \subseteq \mathbb{A}^m \text{ affine line } \deg(f|_L) \leq r \right\}$$

$f|_L$ is the LOWEST DEGREE UNIVARIATE POLYNOMIAL INTERPOLATING f OVER L

RM ⊂ LIFTED CODES

EXAMPLE: $Ev(X^2Y^2) \in LIFT_4(2,2)$ FOR $q=4$, $m=2$, $r=2$

BUT

$Ev(X^2Y^2) \notin RM_4(2,2)$

FOR CONVENIENCE, $m = 2$

A t-CURVE IS

$$\mathcal{L} = \{ (x, g(x)) \mid g \in \mathbb{F}_q[x],\ \deg(g) \le t \}$$

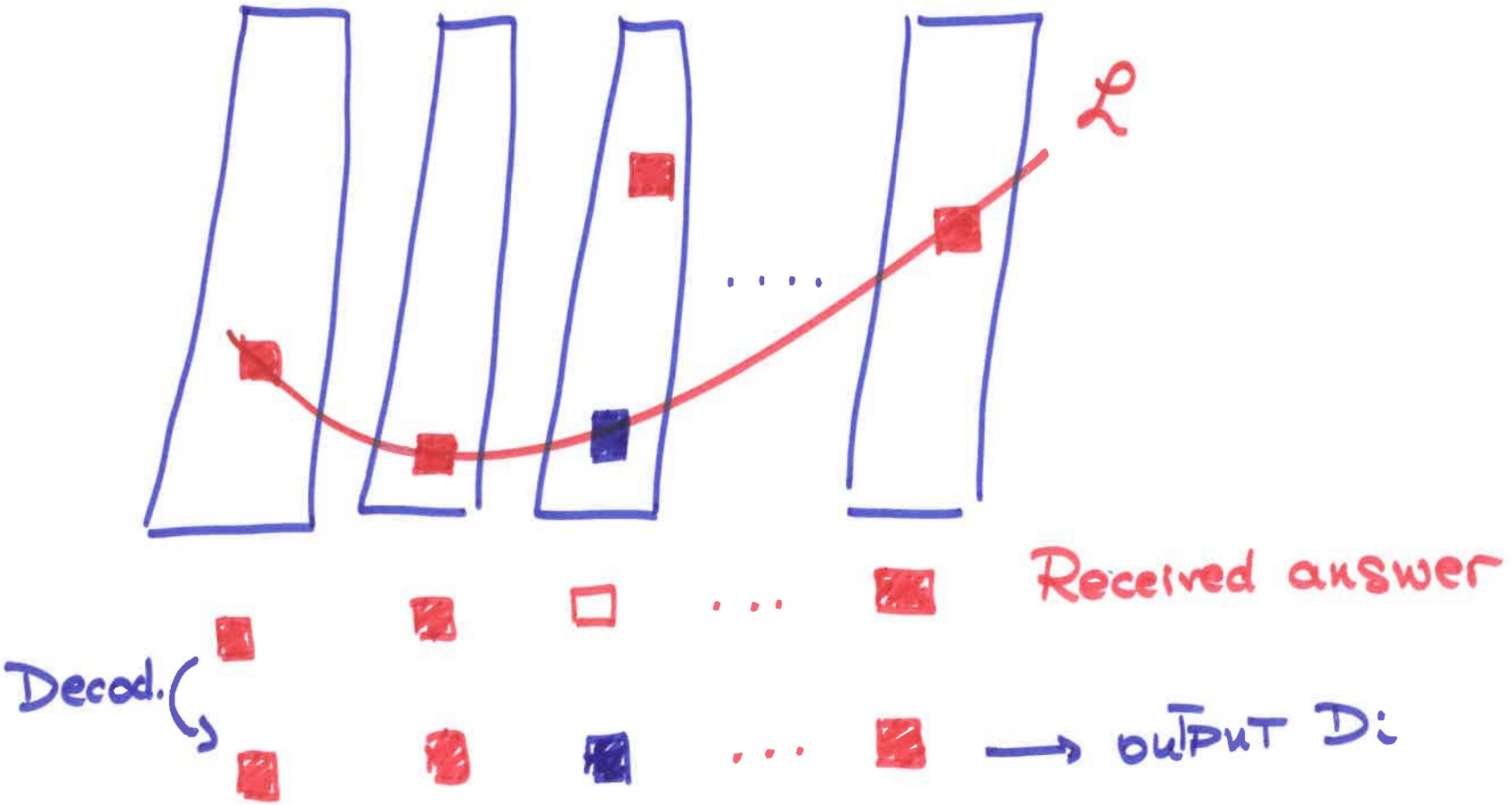
THE WEIGHTED LIFTED R-S CODE OF DEGREE r AND WEIGHT t OVER $\mathbb{F}_q$ IS

$$\mathrm{WLIFT}_q(t,r) = \Big\{ ev_{\mathbb{A}^2}(f) \mid f \in \mathbb{F}_q[X,Y],\ \text{and } \forall\ t\text{-curve } \mathcal{L} \subset \mathbb{A}^2,\ \deg(f|_{\mathcal{L}}) \le r \Big\}$$

THUS, FOR EVERY $c \in \mathrm{WLIFT}_q(t,r)$ AND EVERY t-CURVE $\mathcal{L}$ ONE HAS

$$c|_{\mathcal{L}} \in RS_q(r)$$

DATABASE IS ENCODED WITH $WLIFT_q(t,r)$ AND DISTRIBUTED ACROSS THE SERVERS.



MORE FORMALLY,

$$C = WRM_q^{\dot{\eta}}(d) \subseteq \mathbb{F}_q^{q^2}$$

$$\left\{ ev_{\mathbb{F}_q^2}(P) \mid P \in \mathbb{F}_q[X,Y],\ wdeg_{\dot{\eta}}(P) \le d \right\}$$

$\dot{\eta} = (\eta_1, \eta)$

AN AFFINE η-LINE IS THE SET OF ZEROS OF $Y - \phi(X)$, $\deg(\phi) \le \eta$

- A POLYNOMIAL $f \in \mathbb{F}_q[X,Y]$ WHOSE EVALUATION OVER $\mathbb{F}_q^2$ LIES IN $WRM_q^{\dot{\eta}}(d)$ SATISFIES
  - $ev_{\mathbb{F}_q^2}(f \circ L) \in RS_q(d)$, $(f \circ L)(T) = f(T, \phi(T))$

FOR $L \in \Phi_\eta = \{ L_\phi : t \to (t, \phi(t)) \mid \phi \in \mathbb{F}_q[t] \ \deg(\phi) \le \eta \}$

Th.- [Lavauzelle - Nardi]

Let $\eta \geq 1$, and $\gamma \in (0,1)$, s.t. $q - \lfloor \gamma q \rfloor$ is even.

For every $\delta \leq \frac{1-\gamma}{4}$, the code $WRM_q^{\bar{\eta}}(\lfloor \gamma q \rfloor)$

is $(q-1, \delta, \varepsilon)$-locally correctable where $\varepsilon \leq \frac{2}{1-\gamma}\delta$.

$\angle$ RC conection !

PIR PROTOCOL

$$\mathcal{C} = WRM_q^{\bar{n}}(d) \subseteq \mathbb{F}_q^{q^2}, \quad k = \left\lfloor \frac{d}{n} \right\rfloor \left(d+1-\frac{n}{2}\left(L\frac{d}{n}\right\rfloor+1\right)\right)$$

q SERVERS $(S_t)_{t\in\mathbb{F}_q}$

INITIALISATION

$$D \in \mathbb{F}_q^k \xrightarrow{\text{encoded}} c \in \mathcal{C},$$

FOR EVERY $t \in \mathbb{F}_q$ S_t RECEIVES $c|_{\{t\}\times\mathbb{F}_q}$

QUERIES

ASSUME WE WANT TO RETRIEVE D_i $1 \le i \le k$ AND THAT THE ENCODING MAP IS SYSTEMATIC

$D_i = C_x$ FOR SOME $x = (x_1, x_2) \in \mathbb{F}_q^2$.

- PICK A RANDOM η-LINE $L \in \Phi_\eta$ S.T.

$$L(t_0) = x$$

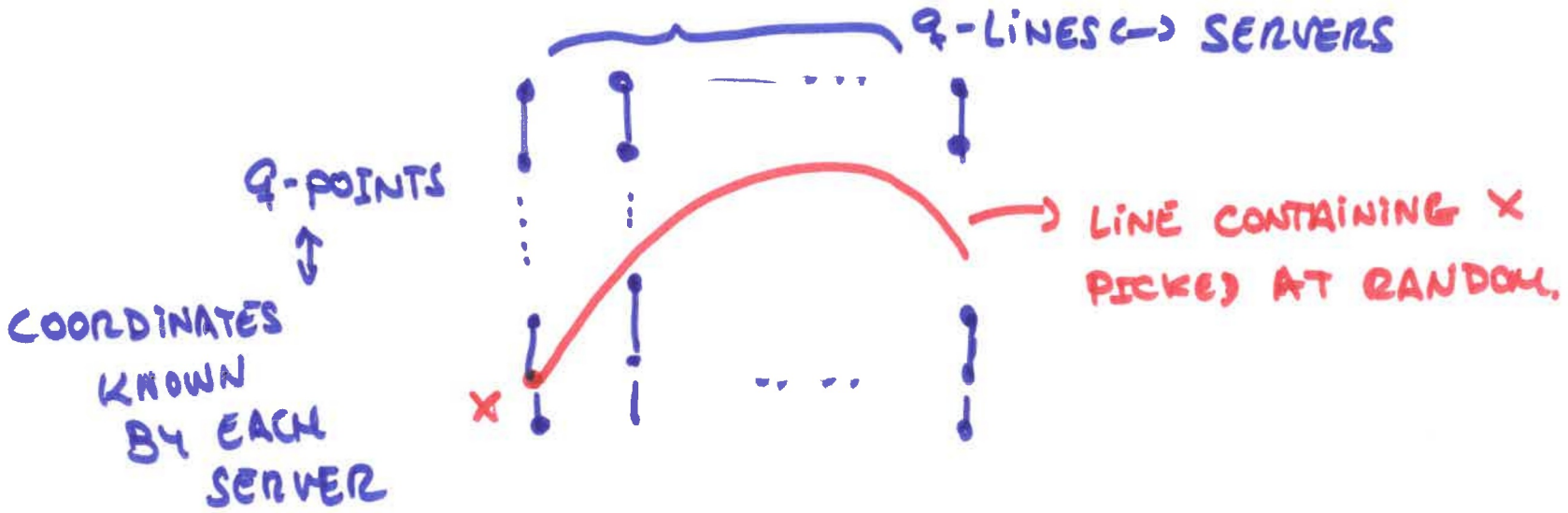
FOR SOME $t_0 \in \mathbb{F}_q$

- SERVER S_{t_0} RECEIVES A RANDOM $y_{t_0} \in \mathbb{F}_q$
- SERVERS S_t $t \neq t_0$ RECEIVE $y_t \in \mathbb{F}_q$ S.T.

$$(t, y_t) = L(t)$$

ANSWERS EVERY SERVER S_t READS $C_{(t,Y_t)}$ AND RETURNS IT.

RECOVERY THE USER COLLECTS $C' = (C_{(t,Y_t)})$ AND RUNS AN ERROR-AND-ERASURE CORRECTING ALGORITHM FOR $RS_q(d)$ WITH INPUT C'. THUS HE/SHE GETS THE CORRECTED SYMBOL.

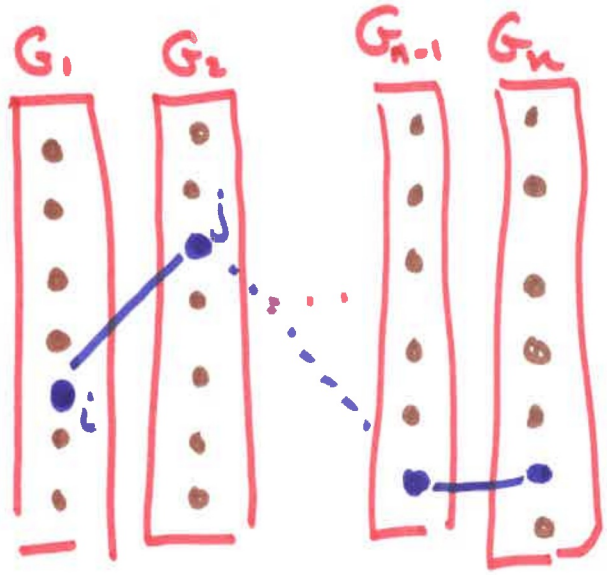


LAVAUZELLE, NARDI, WEIGHTED LIFTED CODES: LOCAL CORRECTABILITIES AND APPLICATIONS TO ROBUST PIR. IEEE Trans. Inf. Th. 67, 1, (2021)

A TRANSVERSAL DESIGN $TD(n,s) = (X, \mathcal{B}, \mathcal{G})$ IS GIVEN BY

- A SET OF POINTS X, $|X| = N = n \cdot s$,
- GROUPS $\mathcal{G} = \{G_j\}_{1 \le i \le n}$ S.T.

$$X = \bigsqcup_{j=1}^{n} G_i \quad \text{and} \quad |G_i| = s$$



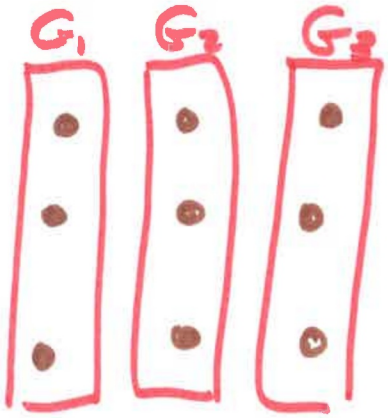
- BLOCKS $B \in \mathcal{B}$ S.T.
  - $B \subset X$, and $|B| = n$
  - For ALL $\{i,j\} \subset X$, $\{i,j\}$ lie:
    either in a single block B
    or in a unique group G

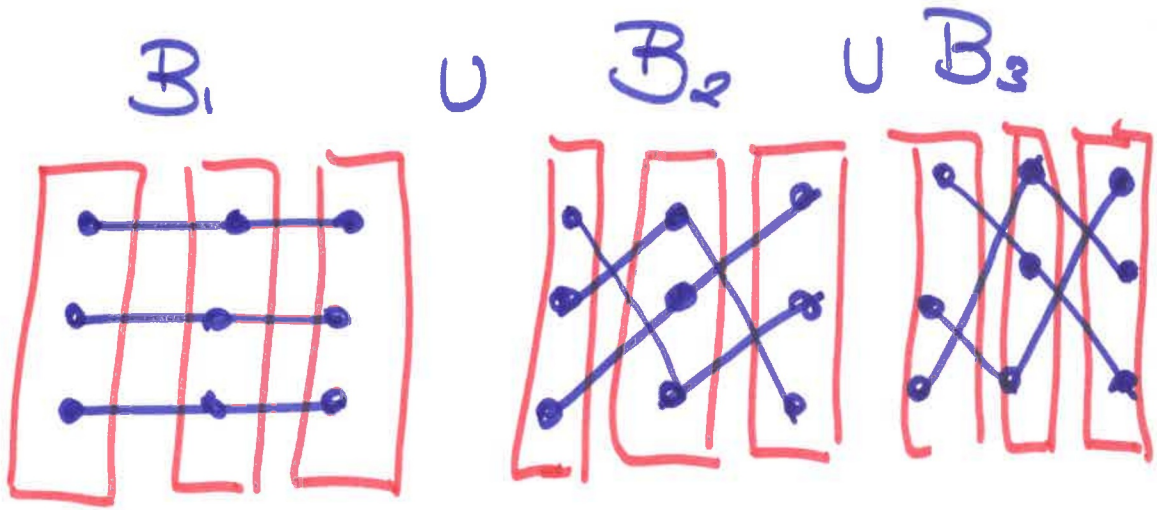
EXAMPLE

$ns = 9$

$s = 3$

3 PARALLEL CLASSES.



$\mathcal{B} =$  $\cup \mathcal{B}_2 \cup \mathcal{B}_3$

GIVEN A TRANSVERSAL DESIGN $TD(n,s) = (X,B,G) = \mathcal{T}$

ITS INCIDENCY MATRIX IS A $ns \times ns$ SQUARE MATRIX

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

THE CODE C BASED ON $\mathcal{T}$ OVER $\mathbb{F}_q$ IS THE $\mathbb{F}_q$-LINEAR CODE ADMITTING M AS PARITY-CHECK MATRIX.

THUS

- → LENGTH = ns
- → DIM = $\dim(\ker M)$
- → EVERY $B \in \mathcal{B}$ → $h \in C^\perp$ S.T. $wt(h_{|G_i}) = 1$

IN THE PREVIOUS EXAMPLE

$$M = \begin{bmatrix} 1&1&1&0&0&0&0&0&0 \\ 0&0&0&1&1&1&0&0&0 \\ 0&0&0&0&0&0&1&1&1 \\ 1&0&0&0&0&1&0&1&0 \\ 0&1&0&1&0&0&0&0&1 \\ 0&0&1&0&1&0&1&0&0 \\ 1&0&0&0&1&0&0&0&1 \\ 0&1&0&0&0&1&1&0&0 \\ 0&0&1&1&0&0&0&1&0 \end{bmatrix}$$

$\text{Rank}_{\mathbb{F}_3} M = 6 \Rightarrow$ ASSOCIATED CODE IS A $[9,3]_3$ CODE.

LET $C \subseteq \mathbb{F}_q^N$ BE A CODE BASED ON A TD$(n,s)^{N,n,s}$

- **INITIALIZATION** USER ENCODES $D \longmapsto c \in C$ AND GIVES $c|_{G_i}$ TO SERVER S_j.

- **QUERY** USER RANDOMLY CHOOSES $B \in \mathcal{B}$ S.T. $i \in B$

$$q_i = Q(i)_j = \begin{cases} \text{UNIQUE ELEMENT } \in B \cap G_j \text{ if } i \notin G_j \\ \text{A RANDOM POINT IN } G_j \text{ OTHERWISE} \end{cases}$$

- **ANSWER** EACH SERVER S_j SENDS BACK c_{q_j}

- **RECOVERY**

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

Th. PREVIOUS PIR IS I-T. PRIVATE.

PROOF.

1) THE ONLY SERVER HOLDING D_i GETS A RANDOM QUERY.

2) FOR EACH OTHER SERVER S_j, THE QUERY q_j GIVES NO INFORMATION ON THE BLOCK B → NO INFO. LEAKS. □

→ COMMUNICATION COMPLEXITY: $\uparrow n \log(s)$ $\downarrow n \log(q)$

→ COMPUTATIONAL COMPLEXITY.

- ONLY ONE READ FOR EACH SERVER (OPTIMAL)
- $\leq n$ ADDITIONS OVER $\mathbb{F}_q$ FOR THE USER.

→ STORAGE OVERHEAD: $(ns-k)\log q$

Questions

- Are there TD's with good $\dim(C)$ depending on (n,s)?

  Affine (geometrical) TD's have been studied.

  $X = \mathbb{F}_q^m$ G_i = hyperplanes

  B = { affine lines l secant to each G_i }

- Collusion resistance?

  Orthogonal arrays $\longrightarrow$ TD's.

Main references

Freij-Hollanti, Ragnar; Gnilke, Oliver W.; Hollanti, Camilla; Horlemann-Trautmann, Anna-Lena; Karpuk, David; Kubjas, Ivo t-private information retrieval schemes using transitive codes. IEEE Trans. Inform. Theory 65 (2019), no. 4, 2107–2118. https://doi.org/10.1007/s10623-018-00591-9

Lavauzelle, Julien; Nardi, Jade Weighted lifted codes: local correctabilities and application to robust private information retrieval. IEEE Trans. Inform. Theory 67 (2021), no. 1, 111–123. https://doi.org/10.1109/TIT.2020.3020752

Lavauzelle, Julien Private information retrieval from transversal designs. IEEE Trans. Inform. Theory 65 (2019), no. 2, 1189–1205. https://doi.org/10.1109/TIT.2018.2861747