Shortened Projective Reed Muller Codes for coded Private Information Retrieval

Myna Vajha, Vinayak Ramkumar and P Vijay Kumar
(Indian Institute of Science, Bangalore)

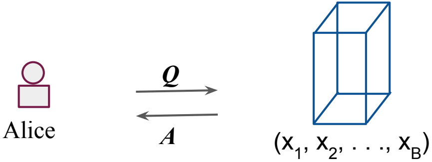
IEEE, International Symposium on Information Theory, 2017.

June 30, 2017

Outline

- Private Information Retrieval (PIR)
- PIR code
- Projective Reed Muller codes as PIR code
- Shortening Algorithm to obtain PIR codes
- Conclusions and Open questions

Private Information Retrieval(PIR): Single Server



- Alice wants to download x_i without revealing any information to server about the index i.

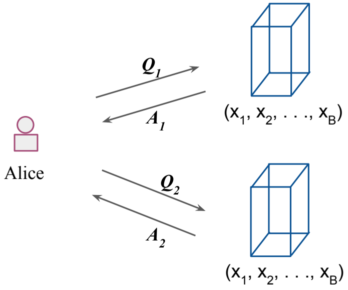
- J is a random variable that represents the index of data in $[1, B]$, and $Q(J)$ be the query sent, then we want $I(Q(J); J) = 0$.

- Number of bits communicated through Query and Answers to achieve PIR is called as communication complexity of PIR.

- It was proved in [1] that communication complexity of $\Omega(B)$ is needed to achieve PIR using a single server.

- B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," Journal of the ACM, 45, 1998

Private Information Retrieval(PIR): Replicated Servers



$\tau = 2$

- It was shown in [1] that the communication complexity can be reduced from $\Omega(B)$ to $O(B^{\frac{1}{3}})$ by introducing a 2-non communicating replicated server model.

- $\tau = \#$ of replicated servers.

---

[1] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," Journal of the ACM, 45, 1998

PIR protocols so far...

τ	Complexity	Year	Authors
2	$O(B^{\frac{1}{3}})$	1995	B. Chor, E. Kushilevitz O. Goldreich, and M. Sudan
τ	$O(B^{\frac{1}{\tau}})$	1995	B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan
τ	$O(B^{\frac{1}{2\tau-1}})$	1997	A. Ambainis
τ	$O(B^{\frac{loglog\,\tau}{\tau\,log\,\tau}})$	2002	A. Beimel, Y. Ishai, E. Kushilevitz, and J.F. Raymond
$\tau \geq 3$	$O(B^{\sqrt{\frac{loglogB}{logB}}})$	2008	S. Yekhanin; K. Efremenko
2	$O(B^{\sqrt{\frac{loglogB}{logB}}})$	2014	Z. Dvir and S. Gopi

Replicated Server PIR

Storage Overhead for replicated server PIR $= \tau \geq 2$.

Can one do better ?

Coded PIR

- Shah, Rashmi, Ramchandran, ISIT 2014.
-

PIR Code

Definition

An (n, k) τ-server PIR code, is an (n, k) linear code such that for every message symbol m_i, $i \in [k]$, there are τ disjoint recovery sets R_{it}, $\forall t \in [\tau]$ i.e. $m_i = \sum_{j \in R_{it}} c_j$, $\forall t \in [\tau]$,

where $\underline{c} = (c_1, \cdots, c_n)$ is a codeword.

A. Fazeli, A. Vardy, and E. Yaakobi, "PIR with low storage overhead: Coding instead of replication," CoRR, vol. abs/1505.06241, 2015.

PIR Code

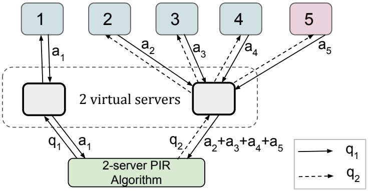


Figure: An Example (5,4) 2-server PIR code.

Storage overhead $= \frac{5}{4}$ $= 1.25$

- $\underline{x}_i = (x_{ij}),\ \forall j \in [B]$ for any $i \in [4]$ is stored in server i.
- Server 5 stores the parity symbols

$$x_{5j} = \sum_{i=1}^{4} x_{ij}.$$

PIR Code

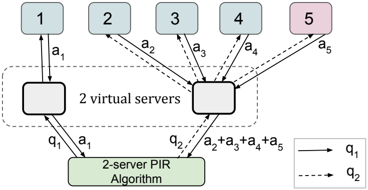


Figure: An Example (5,4) 2-server PIR code.

Storage overhead $= \frac{5}{4} = 1.25$

- $\underline{x}_i = (x_{ij}),\ \forall j \in [B]$ for any $i \in [4]$ is stored in server i.
- Server 5 stores the parity symbols
  $$x_{5j} = \sum_{i=1}^{4} x_{ij}.$$
- To retrieve x_{1j}, generate
  $q_t = Q_B(t, j) \quad t \in [2]$.

PIR Code

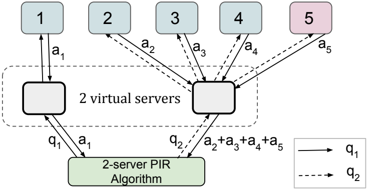


Figure: An Example (5,4) 2-server PIR code.

Storage overhead $= \frac{5}{4} = 1.25$

- $\underline{x}_i = (x_{ij}),\ \forall j \in [B]$ for any $i \in [4]$ is stored in server i.
- Server 5 stores the parity symbols
$$x_{5j} = \sum_{i=1}^{4} x_{ij}.$$
- To retrieve x_{1j}, generate
$q_t = Q_B(t, j) \quad t \in [2]$.
- Send q_1 to server 1 and q_2 to servers $2, 3, 4, 5$. The answer generated by a server $i \in [5]$ on receiving a query q is as shown below:
$$a_i = A(\underline{x_i}, q).$$

PIR Code

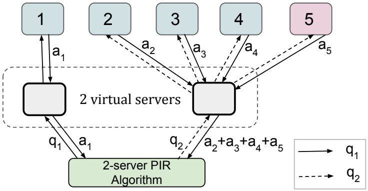


Figure: An Example (5,4) 2-server PIR code.

Storage overhead $= \frac{5}{4} = 1.25$

- $\underline{x}_i = (x_{ij}), \ \forall j \in [B]$ for any $i \in [4]$ is stored in server i.
- Server 5 stores the parity symbols
$$x_{5j} = \sum_{i=1}^{4} x_{ij}.$$
- To retrieve x_{1j}, generate
$q_t = Q_B(t, j) \quad t \in [2]$.
- Send q_1 to server 1 and q_2 to servers $2, 3, 4, 5$. The answer generated by a server $i \in [5]$ on receiving a query q is as shown below:
$$a_i = A(\underline{x}_i, q).$$

- Query and answer functions (Q, A) are determined by the 2-server PIR algorithm.
- Answers that are seen by 2-server PIR protocol are

$$\begin{aligned} a_1 &= A(\underline{x}_1, q_1) \text{ and} \\ a_2 + a_3 + a_4 + a_5 &= A(\underline{x}_2, q_2) + A(\underline{x}_3, q_2) + A(\underline{x}_4, q_2) + A(\underline{x}_5, q_2) \\ &= A(\underline{x}_2 + \underline{x}_3 + \underline{x}_4 + \underline{x}_5, q_2) \text{ (linearity of function A.)} \\ &= A(\underline{x}_1, q_2) \end{aligned}$$

Projective Reed Muller (PRM) Code

- A code vector in binary $PRM(r, m-1)$ code corresponds to evaluations of r-degree homogeneous polynomial in m binary variables at points from $\mathbb{P}^{m-1}(F_2)$.

$$f(x_1, \cdots, x_m) = \sum_{S \subseteq [m], |S| = r} a_S \prod_{i \in S} x_i, \quad a_S \in \mathbb{F}_2$$

$$n = |\mathbb{P}^{m-1}(F_2)| = 2^m - 1, \ k = \binom{m}{r}.$$

- It is clear to see the above polynomials are evaluated to 0 for all $\underline{x}$ such that $w_H(\underline{x}) < r$.
- Can restrict to evaluations at $\underline{x}$ such that $w_H(\underline{x}) \geq r$.

$$n = \sum_{i=r}^{m} \binom{m}{i}, \quad k = \binom{m}{r}.$$

Projective Reed Muller code for PIR

- PRM(2, 3): $r = 2, m = 4$
  - Any code vector corresponds to the evaluation of polynomials of form

    $$f(\underline{x}) = a_{12}x_1x_2 + a_{13}x_1x_3 + a_{14}x_1x_4 + a_{23}x_2x_3 + a_{24}x_2x_4 + a_{34}x_3x_4$$

    of degree 2 in 4 variables at points $\underline{x} = (x_1, x_2, x_3, x_4)$ such that $w_H(\underline{x}) \geq 2$.
  - Message symbol recovery

    $$\begin{aligned} a_{12} &= \sum_{x_1,x_2} f(x_1x_2b_3b_4) \\ &= f(1100) \\ &= f(0110) + f(1010) + f(1110) \\ &= f(0101) + f(1001) + f(1101) \\ &= f(0011) + f(0111) + f(1011) + f(1111). \end{aligned}$$

  - This gives $(n = 11,\ k = 6),\ (\tau = 4)$-server systematic PIR code.

Projective Reed Muller code for PIR

- PRM(2, 3): $r = 2, m = 4$
  - Any code vector corresponds to the evaluation of polynomials of form

    $$f(\underline{x}) = a_{12}x_1x_2 + a_{13}x_1x_3 + a_{14}x_1x_4 + a_{23}x_2x_3 + a_{24}x_2x_4 + a_{34}x_3x_4$$

    of degree 2 in 4 variables at points $\underline{x} = (x_1, x_2, x_3, x_4)$ such that $w_H(\underline{x}) \geq 2$.

  - Message symbol recovery

    $$\begin{aligned} a_{12} &= \sum_{x_1,x_2} f(x_1x_2b_3b_4) \\ &= f(1100) \\ &= f(0110) + f(1010) + f(1110) \\ &= f(0101) + f(1001) + f(1101) \\ &= f(0011) + f(0111) + f(1011) + f(1111). \end{aligned}$$

  - This gives $(n = 11,\ k = 6),\ (\tau = 4)$-server systematic PIR code.

Result

PRM$(r, m-1)$ code is a $(n = \sum_{i=r}^{m} \binom{m}{i},\ k = \binom{m}{r}), (\tau = 2^{m-r})$-server PIR code.

Support Set View point of PRM codes

- We now write $f(\underline{x})$ as $f(\text{Supp}(\underline{x}))$
- Let, R_i for all $i \in \left[\binom{m}{r}\right]$ be the r-element subsets.

$$f(S) = \sum_{\forall R_i \subseteq S} f(R_i). \quad \text{for all } S \subseteq [m] \text{ such that } |S| \geq r.$$

  where $f(R_i) = a_{R_i}$.
- Every such set S corresponds to a coordinate of the code vector.
- For example, PRM$(2,4)$ code has $f(\{1,2,3\}) = f(\{1,2\}) + f(\{1,3\}) + f(\{2,3\})$.
  Setting $a_{12} = a_{13} = a_{23} = 0$, forces $f(1,2,3)$ to be zero and hence can be excluded from the code word.

PIR Codes: any k, τ of form 2^ℓ

- $\tau = 2^\ell = 2^{m-r}$. Choose m such that $k \leq \binom{m}{\ell} = \binom{m}{r}$.
- Shorten $PRM(r, m-1)$ code to obtain the required k. Let,

$$\gamma = \binom{m}{r} - k$$

- Pick γ message symbols that can be represented by r-element sets $\{R_{i_1}, R_{i_2}, \cdots, R_{i_\gamma}\}$ and fix them as 0. This also forces γ code symbols to always be zero.

$$n = \sum_{i=r}^{m} \binom{m}{r} - \gamma'$$

.

- It is clear that $\gamma' \geq \gamma$.
- How to minimize the n i.e., maximize γ' ?

Shortening retains τ

Lemma

On shortening a PRM$(r, m-1)$ code by setting any γ message symbols to zero, the resultant code retains $\tau = 2^{m-r}$ disjoint recovery sets.

Example SPRM code

.

- Consider $k \in (6, 10)$ and $\tau = 8 = 2^{m-r}$. Pick $m = 5$, $r = 2$ i.e., $PRM(2,4)$ code.

k	γ	message	code coordinate sets	γ'	n
10	0	ϕ	ϕ	0	26
9	1	$\{1,2\}$	$\{1,2\}$	1	25
8	2	$\{1,3\}$	$\{1,3\}$	2	24
7	3	$\{2,3\}$	$\{2,3\},\{1,2,3\}$	4	22
6	4	$\{1,4\}$	$\{1,4\}$	5	21
5	5	$\{2,4\}$	$\{2,4\},\{1,2,4\}$	7	19
4	6	$\{3,4\}$	$\{3,4\},\{1,3,4\},\{2,3,4\}$ $\{1,2,3,4\}$	11	15
3	7	$\{1,5\}$	$\{1,5\}$	12	14
2	8	$\{2,5\}$	$\{2,5\},\{1,2,5\}$	14	12
1	9	$\{3,5\}$	$\{3,5\},\{1,3,5\},$ $\{2,3,5\},\{1,2,3,5\}$	18	8
0	10	$\{4,5\}$	$\{4,5\},\{1,4,5\},\{2,4,5\},$ $\{3,4,5\},\{1,2,4,5\},\{1,3,4,5\},$ $\{2,3,4,5\},\{1,2,3,4,5\}$	26	0

- The order in which 2-element message sets are picked above is called co-lexicographic order, where a set $A > B$ iff $\max(A \Delta B) \in A$.

How to get γ'

$$m = 5,\ \ r = 2,\ \ \ell = 3.$$

- $k < \binom{m}{r} = 10$ can be represented by $\ell = 3$ length vector whose weight is $\leq r = 2$.

γ	$\underline{\rho}$	$\mathbb{P}$	γ'	k	n
0	$(0,0,0)$	ϕ	0	10	26
1	$(0,0,1)$	$\{1,2\}$	1	9	25
2	$(0,0,2)$	$\{1,2\},\{1,3\}$	2	8	24
3	$(0,1,0)$	$\{1,2,3\}$	4	7	22
4	$(0,1,1)$	$\{1,2,3\},\{1,4\}$	5	6	21
5	$(0,2,0)$	$\{1,2,3\},\{1,2,4\}$	7	5	19
6	$(1,0,0)$	$\{1,2,3,4\}$	11	4	15
7	$(1,0,1)$	$\{1,2,3,4\},\{1,5\}$	12	3	14
8	$(1,1,0)$	$\{1,2,3,4\},\{1,2,5\}$	14	2	12
9	$(2,0,0)$	$\{1,2,3,4\},\{1,2,3,5\}$	18	1	8

- r-element subsets in $\mathbb{P}$ are picked for shortening.
- $\underline{\rho} = (\rho_{\ell-1}, \cdots, \rho_0)$ where ρ_t represents the number of $r+t$ element sets in $\mathbb{P}$.
- Count the number of distinct subsets of sets in $\mathbb{P}$ with cardinality $\geq r$ to get γ'.

Shortening Algorithm

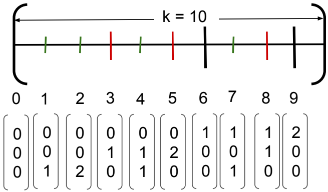
Theorem

For any $\gamma \in \left[0, \binom{m}{\ell}\right)$, γ *can be uniquely represented using a vector* $(\rho_{\ell-1}, \cdots \rho_0)$ *with* $\rho_i \geq 0, \forall i \in [0, \ell-1]$ *and* $\sum_{i=0}^{\ell-1} \rho_i \leq r$ *as*

$$\gamma = \sum_{t=0}^{\ell-1} h(\rho_t, r_t, t) \quad \text{where, } h(p, r, t) = \begin{cases} \sum_{i=0}^{p-1} \binom{r+t-i}{r-i} & p > 0 \\ 0 & p = 0 \end{cases} \text{ and } r_t = r - \sum_{q>t}^{\ell-1} \rho_q.$$

$m = 5, \; r = 2, \; \ell = 3.$

- $k < \binom{m}{r} = 10$ can be represented by $\ell = 3$ length vector whose weight is $\leq r = 2$.

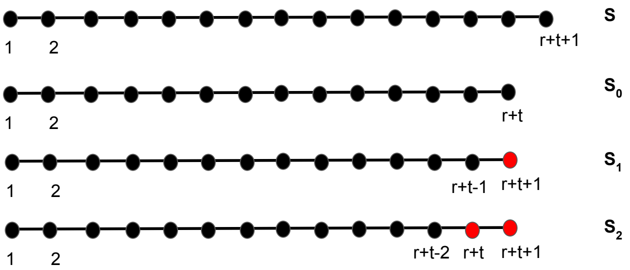


SPRM Codes: Shortening Algorithm

Theorem

For $\gamma = \sum_{i=0}^{\rho_t - 1} \binom{r+t-i}{r-i}$ *for any* $t \in [0, \ell - 1]$ *and* $\rho_t \in [1, r]$, $\gamma' = \sum_{j=0}^{t} \sum_{i=0}^{\rho_t - 1} \binom{r+t-i}{r+j-i}$ *is achievable.*

Case when $\underline{\rho} = (0, \cdots, \rho_t, \cdots, 0)$.



$$\mathbb{P} = \{S_i \mid 0 \leq i < \rho_t\}$$

Shortening Algorithm: any γ

Theorem

For any $\gamma \in \left[0, \binom{m}{\ell}\right)$, represented by vector $(\rho_{\ell-1}, \cdots \rho_0)$ with $\rho_i \geq 0, \forall i \in [0, \ell-1]$ and $\sum_{i=0}^{\ell-1} \rho_i \leq r$. Then,

$$\gamma' = \sum_{t=0}^{\ell-1} h_1(r_t, t) \quad \text{where, } h_1(r,t) = \begin{cases} \sum_{j=0}^{t} \sum_{i=0}^{\rho_t - 1} \binom{r+t-i}{r+j-i} & \rho_t > 0 \\ 0 & \rho_t = 0 \end{cases}$$

is achievable.

$S_0^m = [m]$ and define $\rho_\ell = 0$. For the set S_i^j, j is the number of elements in the set.

$$\begin{aligned} S_i^{r+t-1} &= S_{\rho_t}^{r+t} \setminus \{r_{t-1} + t - i\}, \ \forall i \in [0, r_{t-1} + t - 1] \quad \text{for all } t \in [1, \ell] \\ \mathbb{P} &= \left\{ S_i^{r+t} \mid \forall t \in [0, \ell-1], \ \forall i \in [0, \rho_t - 1] \right\} \end{aligned}$$

Generalized Hamming Weights for PRM codes

- Generalized Hamming Weights (d_i), $\forall i \in \{1, \cdots, k\}$.

$$d_i = \min |\{\text{supp}(D) \mid \forall D \subset C, \ \ \text{rank}(D) = i\}|$$

where, $\text{supp}(D) = \cup_{x \in D} \text{supp}(x)$.

- Shortening of a $PRM(r, m-1)$ by γ gives a sub code of dimension $\binom{m}{r} - \gamma = k - \gamma$.

$$d_{k-\gamma} \leq n - \gamma'$$

where, γ' is given by the shortening algorithm.

Optimal Codes for $\tau = 3, 4$

Theorem

For a (n, k) 3-server systematic PIR code, $n(k, 3) \geq k + \left\lceil \frac{\sqrt{8k+1}+1}{2} \right\rceil$.

- $n(k, \tau) - 1 \geq n(k, \tau - 1)$ as puncturing affects at most one recovery set.
- $n(k, 4) \geq n(k, 3) + 1 \geq k + \left\lceil \frac{\sqrt{8k+1}+1}{2} \right\rceil + 1$
- $PRM(m-2, m-1)$ code is an $(n = k + m + 1,\ k = \binom{m}{2})$ $\tau = 4$-server PIR code. This meets the above lower bound.
- Puncturing $PRM(m-2, m-1)$ at any coordinate gives an $(n = k + m,\ k = \binom{m}{2})$ $\tau = 3$-server PIR code. This meets the lower bound from the theorem.

Contributions

- Optimal systematic PIR codes for $\tau = 3, 4$.
- Upper bounds on generalized hamming weights for binary PRM codes.
- Smaller block lengths in comparison with existing codes.

k \ τ	3*		4*		8		16	
	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2
5	9	10	10	11	19	19	31	31
6	10	11	11	12	21	21	39	40
7	12	12	13	13	22	23	43	43
8	13	13	14	14	24	28	45	54
9	14	14	15	15	25	30	46	60
10	15	17	16	18	26	35	50	61
15	21	23	22	24	36	44	57	80
16	23	24	24	25	37	45	65	84
20	27	30	28	31	42	49	76	92
25	33	35	34	36	52	54	83	108
30	39	42	40	43	58	59	93	118

Block length for some k, τ.
Here n_1 is the block length of the SPRM constructions and n_2 is the block length of the best known codes.

M. Vajha, V. Ramkumar and P. V. Kumar: Binary, Shortened Projective Reed Muller Codes for Coded Private Information

Retrieval, CoRR, vol. abs/1702.05074, 2017. (Accepted to ISIT 2017)

Thanks!