Cryptography Meets Algorithms (15893) Lecture Notes

Lecture 2: Sub-poly Communication 2-server Classical PIR

Scribe: Ashrujit Ghoshal

April 11, 2024

1 Sub-poly Communication 2-server Classical PIR

As introduced in the earlier lecture, Private Information Retreival (PIR) [CKGS98] is a cryptographic mechanism which allows a user holding an index $i \in [n]$ to retrieve the i-th bit from a n-bit database, copies of which are held by one or more (non-colluding) servers, such that the servers do not learn anything about about i. We saw a construction of two-server information-theoretic PIR that has bandwidth $O(n^{1/3})$ in the last lecture. The ultimate goal in this lecture is to see at a PIR construction that has bandwidth sub-polynomial in n. As a first step, we will see a 2-server PIR construction by Woodruff and Yekhanin [WY05] that has bandwidth $O(n^{1/3})$, based on an interpolation approach.

1.1 An Interpolation Approach to 2-Server PIR

As a warm-up, we will first start with a naïve *four-server* construction. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \{0,1\}^n$ be the database. Define $E : [n] \to \{0,1\}^m$ such that $E(1), E(2), \ldots, E(n)$ are n distinct points of Hamming weight 3. Note that such a mapping E exists as long as $\binom{m}{3} \geq n$. Hence, we set $m = O(n^{1/3})$. Define the multivariate polynomial $F(z_1, z_2, \ldots, z_m)$ over $\mathbb{F}_5$ as

$$F(z_1, z_2, \ldots, z_m) = \sum_{i=1}^{n} x_i \prod_{E(i)_\ell = 1} z_\ell \ .$$

First off, observe that since each $E(i)$ has hamming weight 3, F has degree 3. Furthermore, for each i, $F(E(i)) = x_i$.

Therefore, in the PIR protocol a user that has index i, needs to retrieve the value of $F(\mathbf{p})$ for $\mathbf{p} = E(i)$. Let $\mathbf{v}$ be some randomly chosen element of $\mathbb{F}_5^m$. Suppose, the user learns the values $F(\mathbf{p}+\mathbf{v}), F(\mathbf{p}+2\mathbf{v}), F(\mathbf{p}+3\mathbf{v}), F(\mathbf{p}+4\mathbf{v})$. Let $f(\lambda) = F(\mathbf{p}+\lambda\mathbf{v})$. Since the degree of f is 3, and the user knows $f(1), f(2), f(3), f(4)$, they can interpolate f to find $f(0) = F(\mathbf{p}) = x_i$. This observation gives us the following protocol.

1. User picks $\mathbf{v} \in \mathbb{F}_5^m$ uniformly at random
2. To server $j \in [4]$, user sends $\mathbf{p} + j\mathbf{v}$
3. Server j returns $F(\mathbf{p} + j\mathbf{v})$
4. User computes $F(\mathbf{p})$ using interpolation

The correctness of the protocol follows directly from the observation above, while privacy follows because $(\mathbf{p} + j\mathbf{v})$ is distributed uniformly over $\mathbb{F}_5^m$ and does not reveal $\mathbf{p}$. Observe that the communication is somewhat asymmetric: the user sends an element of $\mathbb{F}_5^m$ to the servers while the servers respond with an element of $\mathbb{F}_5$. We will make this symmetric and reduce the number of servers next. Towards that, we prove the following simple lemma. Note that the derivative of a function f at x is denoted $f'(x)$.

Lemma 1. *Let p be a prime. Suppose $x_1, x_2, y_1, y_2, z_1, z_2 \in \mathbb{F}_p$ are such that $x_1 \neq x_2$. Then there exists at most one polynomial $f(\lambda) \in \mathbb{F}_p(\lambda)$ of degree ≤ 3 such that $f(x_1) = y_1, f(x_2) = y_2, f'(x_1) = z_1, f'(x_2) = z_2$.*

Proof. Assume there exist two such polynomials f_1, f_2. Consider the polynomial $f = f_1 - f_2$. Clearly, $f(x_1) = f(x_2) = 0 = f'(x_1) = f'(x_2)$. Therefore $(\lambda - x_1)^2(\lambda - x_2)^2$ divides $f(\lambda)$. Since the degree of $f(\lambda)$ is at most 3, this implies $f(\lambda) = 0$. □

Therefore, if the user were given the value of $f(1), f(2), f'(1), f'(2)$ then they can compute $f(0)$. This observation allows us to give the following 2-server protocol and even reduce the finite field to $\mathbb{F}_3$. (Recall $\mathbf{p}$ is such that $\mathbf{p} = E(i)$ where i is the index whose value the user wants to retrieve, and $f(\lambda) = F(\mathbf{p} + \lambda\mathbf{v})$).

1. User picks $\mathbf{v} \in \mathbb{F}_3^m$ uniformly at random
2. To server $j \in [2]$, user sends $\mathbf{p} + j\mathbf{v}$
3. Server j returns $F(\mathbf{p} + j\mathbf{v}), \frac{\partial F}{\partial z_1}\big|_{(\mathbf{p}+j\mathbf{v})}, \ldots, \frac{\partial F}{\partial z_m}\big|_{(\mathbf{p}+j\mathbf{v})}$
4. For $j \in [2]$, client computes $f'(j) = \sum_{\ell=1}^m \frac{\partial F}{\partial z_\ell}\big|_{(\mathbf{p}+j\mathbf{v})}\mathbf{v}_\ell$ (where $\mathbf{v}_\ell$ is the value at the ℓ-th index of $\mathbf{v}$). Use $f(1), f(2), f'(1), f'(2)$ to compute $f(0)$ and output the answer.

Correctness follows since $f'(\lambda)\big|_j = \sum_{\ell=1}^m \frac{\partial F}{\partial z_\ell}\big|_{(\mathbf{p}+j\mathbf{v})}\mathbf{v}_\ell$ using the chain rule. Privacy follows for the same reason as in the previous protocol. Note that the communication here is $O(m) = O(n^{1/3})$ but symmetric. Next, we will see a protocol where the communication is significantly reduced using ideas from coding theory.

1.2 2-Server PIR using Matching Vector Families

In this section, we will see the 2-server PIR construction by Dvir and Gopi [DG16]. We first define what a matching vector family is.

Definition 2. *Let $S \in \mathbb{Z}_m \setminus \{0\}$ and let $\mathcal{F} = (\mathcal{U}, \mathcal{V})$ where $\mathcal{U} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$, $\mathcal{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ and for all $i \in [n]$, $\mathbf{u}_i, \mathbf{v}_i \in \mathbb{Z}_m^k$. Then $\mathcal{F}$ is called an S-matching vector family of size n and dimension k if for all $i, j \in [n]$,*

$$\langle \mathbf{u}_i, \mathbf{v}_j \rangle \begin{cases} = 0 & \text{if } i = j \\ \in S & \text{if } i \neq j \end{cases}$$

Matching vector family constructions will not be the focus of this class (if you want to learn more about matching vector families and their applications to coding theory, refer to [DGY11] and chapter 4 of [Y$^+$12]), and we will directly use the following result from [Gro00] in our PIR construction.

Proposition 3 ([Gro00])**.** *There is an explicitly constructible S-matching vector family in $\mathbb{Z}_6^k$ of size $n \geq \left(\Omega\left(\frac{(\log k)^2}{\log\log k}\right)\right)$ where $S = \{1, 3, 4\} \subset \mathbb{Z}_6$.*

Note that $n = \left(\Omega\left(\frac{(\log k)^2}{\log\log k}\right)\right)$ implies $k = \exp(O(\sqrt{\log n \log\log n}))$.

We will work with polynomials over the ring $\mathcal{R} = \mathcal{R}_{6,6} = \mathbb{Z}_6[\gamma]/(\gamma^6 - 1)$. We will denote the vector $(\gamma^{z_1}, \gamma^{z_2} \ldots, \gamma^{z_k})$ by $\gamma^{\mathbf{z}}$ where $\mathbf{z} = (z_1, z_2, \ldots, z_k) \in \mathbb{Z}_6^k$. Further for $\mathbf{y} = (y_1, y_2, \ldots, y_k)$,$\mathbf{z} = (z_1, z_2, \ldots, z_k)$, we denote by $\mathbf{y}^{\mathbf{z}}$ the monomial $\prod_{i=1}^k y_i^{z_i}$.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be the database and $(\mathcal{U}, \mathcal{V})$ be a $\{1, 3, 4\}$-matching vector family of dimension k and size n. Define $F(\mathbf{y}) \in \mathcal{R}[\mathbf{y}] = \mathcal{R}[y_1, y_2, \ldots, y_k]$ given by

$$F(\mathbf{y}) = F(y_1, y_2, \ldots, y_k) = \sum_{\ell=1}^{n} x_\ell \mathbf{y}^{\mathbf{u}_\ell} .$$

Here is the protocol. Let $i \in [n]$ be the index the user wants to read.

1. The user picks $\mathbf{z}$ uniformly at random from $\mathbb{Z}_6^k$

2. For $j = 1, 2$, the user sends $\mathbf{z} + (j-1)\mathbf{v}_i$ to server j

3. Server j sends back $F(\gamma^{\mathbf{z}+(j-1)\mathbf{v}_i})$ and

$$F^{(1)}(\gamma^{\mathbf{z}+(j-1)\mathbf{v}_i}) := \begin{bmatrix} y_1 \frac{\partial F}{\partial y_1}\big|_{\gamma^{\mathbf{z}+(j-1)\mathbf{v}_i}} \\ y_2 \frac{\partial F}{\partial y_2}\big|_{\gamma^{\mathbf{z}+(j-1)\mathbf{v}_i}} \\ \vdots \\ y_k \frac{\partial F}{\partial y_k}\big|_{\gamma^{\mathbf{z}+(j-1)\mathbf{v}_i}} \end{bmatrix}$$

4. Let

$$M := \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \\ 1 & \gamma & \gamma^3 & \gamma^4 \\ 0 & \gamma & 3\gamma^3 & 4\gamma^4 \end{bmatrix}$$

Compute the first entry of the matrix

$$M^{-1} \begin{bmatrix} F(\gamma^{\mathbf{z}}) \\ \langle F^{(1)}(\gamma^{\mathbf{z}}), \mathbf{v}_i \rangle \\ F(\gamma^{\mathbf{z}+\mathbf{v}_i}) \\ \langle F^{(1)}(\gamma^{\mathbf{z}+\mathbf{v}_i}), \mathbf{v}_i \rangle \end{bmatrix} .$$

Return 0 if it is 0 and 1 otherwise.

Proof of correctness: Define

$$G(j) := F(\gamma^{\mathbf{z}+(j-1)\mathbf{v}_i}) = \sum_{\ell=1}^{n} x_\ell \gamma^{\langle \mathbf{z}, \mathbf{u}_\ell \rangle + (j-1)\langle \mathbf{v}_i, \mathbf{u}_\ell \rangle} .$$

Using the fact that $\gamma^6 = 1$ we can write

$$G(j) = \sum_{m=0}^{5} c_m \gamma^{(j-1)m} ,$$

where each $c_m \in \mathcal{R}$ is given by

$$c_m = \sum_{\ell : \langle \mathbf{u}_\ell, \mathbf{v}_i \rangle = m} x_\ell \gamma^{\langle \mathbf{z}, \mathbf{u}_\ell \rangle} .$$

Since

$$\langle \mathbf{u}_\ell, \mathbf{v}_i \rangle \bmod 6 \begin{cases} = 0 & \text{if } \ell = i \\ \in \{1, 3, 4\} & \text{if } \ell \neq i \end{cases}$$

we can conclude that $c_0 = x_i \gamma^{\langle \mathbf{u}_i, \mathbf{z} \rangle}$ and $c_2 = c_5 = 0$. Therefore,

$$G(j) = c_0 + c_1 \gamma^{(j-1)} + c_3 \gamma^{3(j-1)} + c_4 \gamma^{4(j-1)} \ .$$

Next, consider the polynomial

$$g(T) = c_0 + c_1 T + c_3 T^3 + c_4 T^4 \in \mathcal{R}[T] \ .$$

By definition,

$$g(\gamma^{j-1}) = G(j) = F(\gamma^{\mathbf{z}+(j-1)\mathbf{v}_i})$$

Further, consider this inner product:

$$\left\langle F^{(1)}(\gamma^{\mathbf{z}+(j-1)\mathbf{v}_i}), \mathbf{v}_i \right\rangle$$

This is equal to

$$\begin{aligned} &\sum_{\ell=1}^{n} x_\ell \langle \mathbf{u}_\ell, \mathbf{v}_i \rangle \gamma^{\langle \mathbf{z}, \mathbf{u}_\ell \rangle + (j-1)\langle \mathbf{v}_i, \mathbf{u}_\ell \rangle} \\ &= \sum_{m=0}^{5} m \left(\sum_{\ell : \langle \mathbf{u}_\ell, \mathbf{v}_i \rangle = m \bmod 6} x_\ell \gamma^{\langle \mathbf{z}, \mathbf{u}_\ell \rangle} \right) \gamma^{(j-1)m} = \sum_{m=0}^{5} m c_m \gamma^{(j-1)m} \end{aligned}$$

Therefore

$$\begin{bmatrix} F(\gamma^{\mathbf{z}}) \\ \langle F^{(1)}(\gamma^{\mathbf{z}}), \mathbf{v}_i \rangle \\ F(\gamma^{\mathbf{z}+\mathbf{v}_i}) \\ \langle F^{(1)}(\gamma^{\mathbf{z}+\mathbf{v}_i}), \mathbf{v}_i \rangle \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \\ 1 & \gamma & \gamma^3 & \gamma^4 \\ 0 & \gamma & 3\gamma^3 & 4\gamma^4 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_3 \\ c_4 \end{bmatrix} = M \begin{bmatrix} c_0 \\ c_1 \\ c_3 \\ c_4 \end{bmatrix}$$

Note that determinant of M is non-zero. Further since $c_0 = x_i \gamma^{\mathbf{u}_i, \mathbf{z}}$ and $x_i \in \{0, 1\}$, we have that $c_0 = 0$ if and only if $x_i = 0$. Therefore the first entry of

$$M^{-1} \begin{bmatrix} F(\gamma^{\mathbf{z}}) \\ \langle F^{(1)}(\gamma^{\mathbf{z}}), \mathbf{v}_i \rangle \\ F(\gamma^{\mathbf{z}+\mathbf{v}_i}) \\ \langle F^{(1)}(\gamma^{\mathbf{z}+\mathbf{v}_i}), \mathbf{v}_i \rangle \end{bmatrix}$$

is 0 if and only if $x_i = 0$.

References

[CKGS98] Benny Chor, Eyal Kushilevitz, Oded Goldreich, and Madhu Sudan. Private information retrieval. *Journal of the ACM (JACM)*, 45(6):965–981, 1998.

[DG16] Zeev Dvir and Sivakanth Gopi. 2-server pir with subpolynomial communication. *Journal of the ACM (JACM)*, 63(4):1–15, 2016.

[DGY11] Zeev Dvir, Parikshit Gopalan, and Sergey Yekhanin. Matching vector codes. *SIAM Journal on Computing*, 40(4):1154–1178, 2011.

[Gro00] Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit ramsey graphs. *Combinatorica*, 20(1):71–86, 2000.

[WY05] David Woodruff and Sergey Yekhanin. A geometric approach to information-theoretic private information retrieval. In *20th Annual IEEE Conference on Computational Complexity (CCC'05)*, pages 275–284. IEEE, 2005.

[Y⁺12] Sergey Yekhanin et al. Locally decodable codes. *Foundations and Trends® in Theoretical Computer Science*, 6(3):139–255, 2012.