AICR · OPEN PROBLEM

Polylogarithmic Single-Server Shuffle PIR with Negligible Security Error

Information-theoretic, one server, polynomially many anonymous queries

Status. Statement: AI-written from Yuval Ishai, Mahimna Kelkar, Daniel Lee and Yiping Ma, *Information-Theoretic Single-Server PIR in the Shuffle Model*, IACR ePrint 2024/930. Not yet checked by a human. Settled in three neighbouring corners: $O(n^{\gamma})$ communication with inverse-polynomial error and polynomially many queries; polylogarithmic communication with $n^{O(\log n)}$ queries; and negligible error with $O(n/\log n)$ communication. The polylogarithmic-and-negligible corner with polynomially many queries is open, and is ruled out for one broad class of constructions by the source's own lower bound.

Category. *Information-Theoretic Cryptography.*

Conjecture (informal). *In the shuffle model, many clients send their PIR queries to a single server through an anonymous channel, so the server sees the whole multiset of queries but not who sent which. The source shows that this alone buys something impossible in the standard model: a single-server PIR with sublinear communication and* information-theoretic *security, no computational assumption anywhere. But its security error is only inverse-polynomial. The source's main open question is whether one can have both — per-query communication polylogarithmic in the database size n, and a security error negligible in n — while still requiring only polynomially many honest queries. Its own lower bound says no construction in its inner–outer paradigm with an* additive *inner layer can do this; the authors conjecture a Reed–Muller inner layer can.*

1 The setting

In the shuffle model for PIR [IKOS06, GIK+24], a single server holds a database $x \in \Sigma^n$ and many stateless clients each retrieve one entry. A client's query is split into k *sub-queries*, each sent as a separate anonymous message; all messages from all clients arrive at the

server in a uniformly random order. Security asks that the shuffled multiset of sub-queries be statistically close whichever indices the clients wanted — so the shuffler is doing the work that a second non-colluding server does in multi-server PIR, without holding a copy of the database.

The source's main construction (its Theorem 1.1) gives, for every constant $0 < \gamma < 1$, a doubly efficient scheme with $O(n^{\gamma})$ per-query communication and computation and ε-statistical security for any inverse polynomial $\varepsilon = 1/p_1(n)$, provided the number of honest queries is at least some polynomial $p_2(n) = O(n^{1+4/\gamma} \cdot p_1(n)^8)$. It is built by an *inner–outer* paradigm: two standard multi-server PIR schemes are composed, an outer one whose sub-queries are distributed among the shuffled messages and an inner one that answers each. Instantiating the inner layer with a $(\log n)$-server CNF-based protocol pushes communication down to polylog(n), but only when the number of clients is super-polynomial, $n^{O(\log n)}$.

Two things bound the negligible-error corner. First, the source's Theorem 6.5: for schemes in the inner–outer paradigm whose inner layer is a constant-server *additive* PIR, if the number of clients and the sub-query-vector counts of both layers are all polynomially bounded, then the security error is at least $1/p_2(n)$ for some polynomial p_2 — inverse-polynomial security is tight there. Second, its Remark 11: negligible error *is* achievable if one settles for $O(n/\log n)$ communication, by a split-and-mix matrix scheme with super-linearly many clients. So the open region is precisely small communication together with negligible error.

The source records the gap as its main open question and states a conjectured route: instantiate both the outer and the inner PIR with the Reed–Muller construction at a polylogarithmic security threshold and polylogarithmic communication. That construction is not additive, so Theorem 6.5 does not apply to it.

2 Notation and parameters

- n is the database size, Σ a finite alphabet, $x \in \Sigma^n$ the database.
- $k = k(n)$ is the number of sub-queries one client's query is split into; all are sent to the same server as separate anonymous messages.
- $C = C(n)$ is a lower bound on the number of honest client

queries; $\Pi = \{\Pi_c\}_{c\in\mathbb{N}}$ is the shuffler, with Π_c a distribution over the symmetric group S_c. The *perfect* shuffler takes each Π_c uniform.
- $\varepsilon = \varepsilon(n)$ is the statistical security error.
- Per-query communication is the total size of one client's k sub-queries plus the k answers it receives.
- $\mathrm{SD}(\cdot,\cdot)$ is statistical distance.

3 The conjecture

Definition 1 (PIR in the shuffle model, [IKLM24, Definition 5.1]). A single-server PIR protocol over Σ in the shuffle model is a tuple $\mathsf{ShPIR} = (\mathsf{Setup}, \mathsf{Query}, \mathsf{Answer}, \mathsf{Recon})$ where $\mathsf{Setup}(x) \to P_x$ is deterministic and run by the server on $x \in \Sigma^n$; $\mathsf{Query}(i; n) \leftarrow (q_1, \ldots, q_k)$ is randomized and run by the client on an index $i \in [n]$; $\mathsf{Answer}(P_x, q_\ell) \to a_\ell$ is deterministic, the same algorithm for every sub-query; and $\mathsf{Recon}(a_1, \ldots, a_k) \to x_i$ is deterministic. Correctness requires, for every n, every $x \in \Sigma^n$ and every $i \in [n]$, that Recon outputs x_i with probability 1.
For security, fix n, a shuffler Π and a query count C^*, and for $I = (i_1, \ldots, i_{C^*}) \in [n]^{C^*}$ let $\widetilde{D}_{n,\Pi,C^*}(I)$ be the distribution of $\pi(q)$, where q is the concatenation of $\mathsf{Query}(i_1; n), \ldots, \mathsf{Query}(i_{C^*}; n)$ (independent draws) and $\pi \leftarrow \Pi_{kC^*}$. Then ShPIR is (Π, C, ε)-*secure* if for every n, every $C^* \geq C(n)$ and all $I, I' \in [n]^{C^*}$,

$$\mathrm{SD}\big(\widetilde{D}_{n,\Pi,C^*}(I),\ \widetilde{D}_{n,\Pi,C^*}(I')\big) \leq \varepsilon(n).$$

Conjecture 1 (Polylogarithmic shuffle PIR with negligible error). *There is a single-server PIR protocol in the shuffle model (Definition 1) over $\Sigma = \{0,1\}$, a polynomial C, a negligible function ε and a constant d such that, for every n, the protocol on databases of size n has per-query communication at most $(\log n)^d$ and is (Π, C, ε)-secure for the perfect shuffler Π.*

Remark 1 (The route the source conjectures). The source does not merely pose Conjecture 1; it names a candidate. Its closing section states: "We conjecture that polylogarithmic communication

per client with negligible security can be achieved by instantiating both OPIR and IPIR with the Reed-Muller PIR construction with a polylogarithmic security threshold and a polylogarithmic communication complexity." Establishing Conjecture 1 *via that instantiation* is therefore the source's own conjecture, and is a stronger statement than Conjecture 1 itself; a resolution by any other construction settles the open question but not the authors' guess at how.

Remark 2 (Why the polynomial query bound is the crux). Dropping the requirement that C be polynomial makes the statement already known: the source achieves polylogarithmic communication with a $(\log n)$-server CNF-based inner layer once the number of clients is $n^{O(\log n)}$. Dropping instead the polylogarithmic communication makes it known too, by the source's Remark 11, at $O(n/\log n)$ communication. Conjecture 1 keeps both.

Remark 3 (What the source's lower bound already excludes). Theorem 6.5 of the source rules out negligible error for inner–outer constructions $\mathsf{ShPIR}(\Phi,\Psi)$ in which Ψ is a constant-server additive PIR and the client count together with the numbers K_Φ, K_Ψ of possible sub-query vectors are all polynomially bounded. A proof of Conjecture 1 must therefore leave that class — which is what the Reed–Muller proposal does. A refutation, conversely, would have to extend Theorem 6.5 beyond additive inner layers, and the source flags instantiating the inner PIR with Reed–Muller as "an interesting open problem to consider" precisely because its bound says nothing there.

4 Bibliography

[IKLM24] Yuval Ishai, Mahimna Kelkar, Daniel Lee and Yiping Ma. *Information-theoretic single-server PIR in the shuffle model.* IACR ePrint 2024/930. The source. Definition 5.1 is on page 15; Theorem 1.1 on page 2; Theorem 6.5 and the Reed–Muller remark on page 29; the open question in Section 7 ("Conclusion and Open Questions"), page 30.

[IKOS06] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky and Amit Sahai. *Cryptography from anonymity.* 47th Annual Symposium on Foundations of Computer Science (FOCS), 2006, pages 239–248. Introduces cryptography in the shuffle model.

[BIK05] Amos Beimel, Yuval Ishai and Eyal Kushilevitz. *General constructions for information-theoretic private information retrieval.* Journal of Computer and System Sciences, 2005. The two-server information-theoretic PIR the source's toy protocol starts from.

[DG15] Zeev Dvir and Sivakanth Gopi. *2-server PIR with sub-polynomial communication.* ACM Symposium on Theory of Computing (STOC), 2015. The standard-model constant-server benchmark the source compares its polylogarithmic variant against.

[GIK+24] Adrià Gascón, Yuval Ishai, Mahimna Kelkar, Baiyu Li, Yiping Ma and Mariana Raykova. *Computationally secure aggregation and private information retrieval in the shuffle model.* IACR ePrint 2024/870. The computational-security companion line the source points to for concrete efficiency.

[LMW23] Wei-Kai Lin, Ethan Mook and Daniel Wichs. *Doubly efficient private information retrieval and fully homomorphic RAM computation from ring LWE.* ACM Symposium on Theory of Computing (STOC), 2023. The standard-model doubly efficient PIR the source contrasts itself with; inherently computational.