CONJURA · OPEN PROBLEM

A Balanced $\sqrt{n}$ Information-Theoretic Preprocessing PIR

One server, client-specific preprocessing, no computational assumption

Status. Statement: AI-written from Yuval Ishai, Elaine Shi and Daniel Wichs, *PIR with Client-Side Preprocessing: Information-Theoretic Constructions and Lower Bounds*, IACR ePrint 2024/976. Not yet checked by a human. The client-space–bandwidth product is settled up to $n^{o(1)}$ by the source's own matching bounds; what is open is whether all three of client space, bandwidth and per-query server computation can sit at $\widetilde{O}(n^{1/2})$ simultaneously. The best known point has client space $\widetilde{O}(n^{2/3})$ and server computation $\widetilde{O}(n^{2/3})$.

Category. *Information-Theoretic Cryptography.*

Conjecture (informal). *Classical private information retrieval with one server needs public-key cryptography for sublinear bandwidth, and linear server work per query. Both barriers fall if the client first downloads a short* hint *in a preprocessing phase — and, remarkably, the source shows they fall with no cryptographic assumption at all, provided the hint is client-specific. It also proves a matching lower bound: a client hint of t bits forces t consecutive queries to consume $n/2$ bandwidth between them, so client space times per-query bandwidth is essentially n. The symmetric point on that curve is $\sqrt{n}$ of each. The source's own scheme reaches it in two of the three cost measures but not the third: it needs $\widetilde{O}(n^{2/3})$ client space and $\widetilde{O}(n^{2/3})$ server computation per query. Whether one scheme can put client space, bandwidth* and *per-query server computation all at $\widetilde{O}(n^{1/2})$ is the open problem.*

1 The setting

In preprocessing PIR the server first encodes the n-bit database x, and the client downloads a hint; afterwards the client makes queries, each retrieving one bit x_i while hiding i. Two flavours are distinguished by who keeps state. In the *public* preprocessing model

the client stores nothing, and there the single-server case is hopeless information-theoretically: the lower bound of Di Crescenzo, Malkin and Ostrovsky [DMO00] rules out sublinear bandwidth however much the server preprocesses. The source therefore works in the *client-specific* preprocessing model, where the client keeps a hint, and shows that information-theoretic single-server preprocessing PIR is possible there.

Its lower bound (Theorem 1.1) is the shape of the trade-off: for any $t \geq 0$ there is no information-theoretic preprocessing PIR with client space below $t/2$ in which some consecutive run of t queries consumes less than $n/2$ bandwidth in total; if a computationally secure scheme did this, one-way functions would exist. The bound holds against arbitrary server-side encoding with no limit on server space, against both parties keeping arbitrary dynamic state after the preprocessing, and against unbounded polynomial computation and any number of round trips — it constrains only the size of the initial client hint. It is tight up to $n^{o(1)}$ if one does not care about server computation: with $t \cdot n^{o(1)}$ client space one can get $O(n/t)$ amortized bandwidth per query.

The remaining tension is with *sublinear server computation*. The source's Theorem 1.4 gives an information-theoretic scheme with $\widetilde{O}(n^{1/3})$ online bandwidth, $\widetilde{O}(n^{1/2})$ offline bandwidth, $\widetilde{O}(n^{1/2})$ client computation, $\widetilde{O}(n^{2/3})$ server computation per query and $\widetilde{O}(n^{2/3})$ client space, after an $O(n)$-bandwidth streaming preprocessing pass. Against this sit its own Theorem 1.1 and the client-space–server-computation bound of Corrigan-Gibbs and Kogan [CK20]. A gap remains in the sublinear-server-computation regime, and the source names closing it at the balanced point as a fascinating open problem.

2 Notation and parameters

- n is the database length in bits; $x \in \{0,1\}^n$.
- *Client space* is the number of bits the client stores after the preprocessing phase (the hint), and thereafter.
- *Bandwidth* per query counts all bits exchanged for that query, in either direction, across any number of round trips. Where the source separates them, *online* bandwidth is on the critical path for obtaining the answer and *offline* bandwidth is background maintenance.

- *Server computation* per query is the number of bits of the (preprocessed) database the server reads while answering.
- Costs are per query and worst case unless stated as amortized. $\widetilde{O}(\cdot)$ hides a multiplicative $\log^{\alpha(n)} n$ for an arbitrary super-constant α, as in the source.
- Security is information-theoretic: the distribution of the server's view is independent of the queried index, for computationally unbounded servers, up to $\mathrm{negl}(n)$.

3 The conjecture

Definition 1 (Single-server preprocessing PIR with client–side state). A single-server preprocessing PIR scheme consists of a preprocessing protocol and a query protocol. In preprocessing, the server holds $x \in \{0,1\}^n$, may encode it arbitrarily and store the encoding, and the client interacts with the server and stores a hint. In each subsequent query the client holds an index $i \in [n]$, interacts with the server (both parties may update their state), and outputs a bit. The scheme is *correct* if that bit equals x_i with probability $1 - \mathrm{negl}(n)$ for every x, every i, and every sequence of preceding queries. It is *information-theoretically secure* if for every x and every two query sequences of the same length, the distributions of the server's entire view are within $\mathrm{negl}(n)$ statistical distance, against a computationally unbounded server.

Conjecture 1 (A balanced $\sqrt{n}$ scheme). *There is a single-server preprocessing PIR scheme (Definition 1) with information-theoretic security in which, for databases of length n and after a one-time preprocessing phase, all three of*

(a) the client space,
(b) the per-query bandwidth (online plus offline), and
(c) the per-query server computation

are bounded by $\widetilde{O}(n^{1/2})$, for a scheme supporting any polynomial number of queries.

Remark 1 (The conjecture sits exactly on the known lower bound). Client space $\widetilde{O}(n^{1/2})$ with per-query bandwidth $\widetilde{O}(n^{1/2})$

has product $\widetilde{O}(n)$, so Conjecture 1 does not contradict the source's Theorem 1.1: with $t = \widetilde{O}(n^{1/2})$ client space, t consecutive queries are allowed to consume about n bandwidth in total, which is exactly $\widetilde{O}(n^{1/2})$ each. The statement is therefore a question about achievability at the boundary, not about beating the bound. The genuinely new demand is the third clause: server computation $\widetilde{O}(n^{1/2})$ per query, where the source's Theorem 1.4 spends $\widetilde{O}(n^{2/3})$.

Remark 2 (Either direction resolves it). The source states the problem as "whether it is possible to close this gap". Both directions are open and both count as a resolution: exhibiting a scheme as in Conjecture 1, or proving that no information-theoretic single-server preprocessing PIR can have client space, bandwidth and per-query server computation all $\widetilde{O}(n^{1/2})$ — which, by the second half of the source's Theorem 1.1, would have to be proved unconditionally rather than by exhibiting a one-way function.

Remark 3 (What is not being asked). Conjecture 1 is not about the preprocessing phase's own cost: the source's Theorem 1.4 spends $O(n)$ bandwidth and server computation there and $\widetilde{O}(n)$ client computation, and that is not what the $\widetilde{O}(n^{1/2})$ bounds refer to. Nor is it about client computation per query, which the source's scheme already has at $\widetilde{O}(n^{1/2})$. It is also distinct from the source's separate two-server open question, which asks for a two-server scheme with the same efficiency but *adaptive* correctness; that one is about the preprocessing leaking the choice of random sets to one server, not about the cost profile.

4 Bibliography

[ISW24] Yuval Ishai, Elaine Shi and Daniel Wichs. *PIR with client-side preprocessing: information-theoretic constructions and lower bounds.* IACR ePrint 2024/976. The source. Theorem 1.1 (lower bound) is on page 2, Theorem 1.4 and the open problem on page 7, Table 1 on page 3.

[CK20] Henry Corrigan-Gibbs and Dmitry Kogan. *Private information retrieval with sublinear online time.* EUROCRYPT 2020. The client-space versus server-computation trade-off the source's gap is measured against; its Theorem 23 is one of the two lower bounds bracketing the open region.

[DMO00] Giovanni Di Crescenzo, Tal Malkin and Rafail Ostrovsky. *Single database private information retrieval implies oblivious transfer.* EUROCRYPT 2000. The barrier that forces the source into the client-specific preprocessing model.

[BIM00] Amos Beimel, Yuval Ishai and Tal Malkin. *Reducing the servers computation in private information retrieval: PIR with preprocessing.* CRYPTO 2000. Two-server information-theoretic preprocessing PIR in the public preprocessing model.

[WY05] David P. Woodruff and Sergey Yekhanin. *A geometric approach to information-theoretic private information retrieval.* IEEE Conference on Computational Complexity, 2005. The other prior two-server information-theoretic preprocessing scheme.

[Yeo23] Kevin Yeo. *Lower bounds for (batch) PIR with private preprocessing.* EUROCRYPT 2023. Extends the prior lower bounds to the batched setting.