Poster: (Symmetric) PIR over Arbitrary Sized Records

Ryan Henry
PhD Student
rhenry@cs.uwaterloo.ca

Yizhou Huang
MMath Student
y226huang@cs.uwaterloo.ca

Ian Goldberg
Associate Professor
iang@cs.uwaterloo.ca

Cheriton School of Computer Science
University of Waterloo
Waterloo, ON, Canada N2L 3G1

(Poster Abstract)

Private information retrieval (PIR) is a cryptographic technique that helps users retrieve information from a database in a way that is highly respectful of privacy: the user submits a query encoding a set of keywords [1], the indices of certain records [2], or some simple SQL statements [3], and the database server processes and responds to the query without learning any nontrivial information about what information the user is after.

The trivial way to solve the PIR problem is to have the database server respond to every query with the entire database, and then let the user carry out her own local keyword searches, index lookups, or SQL statement evaluations. This solution is simple and information-theoretically secure, but it is not very interesting and it is highly impractical for large databases since the communication cost is linear in the length of the database. To exclude this and related trivial solutions, the PIR literature only considers protocols whose total communication cost is strictly less than (and scales sublinearly with) the length of the database. At first blush, this might seem like two contradictory goals; after all, the database server knows precisely what sequence of bits it receives from and sends to each user. If some database bits are not somehow "included" in the response, then given a sufficiently clever algorithm and sufficient computational resources, it seems likely that the database server could deduce something about which database bits the user is *not* requesting. It turns out that there are several ways to solve this problem if the server prepares each query response by performing some computation involving every bit in the database. (Therefore, all PIR protocols use sublinear communication, but at least linear computation).

Most existing PIR protocols belong to one of two categories: 1) the *computational PIR* (CPIR) protocols, which use ideas from public key cryptography to encode queries (and their responses) so that only the user can decode them, and 2) *information-theoretic PIR* (IT-PIR) protocols, which use ideas from secret sharing and coding theory to split each query among multiple independent PIR servers in such a way that the servers cannot deduce any information about the query unless more than a threshold of them collude. *Symmetric PIR* (SPIR) is a variant of (either IT- or C)PIR with the added privacy constraint that users should not learn any information about database records that they did not explicitly request. This work is concerned with the efficiency of IT-PIR and, by extensions, of SPIR protocols constructed from IT-PIR. We focus on Goldberg's robust IT-PIR [4] and extend it with support for *multi-block queries* that can dramatically improve "throughput" in Goldberg's scheme.

Existing work models an N-bit (or N-word) database $\mathbf{D}$ as an r-by-s matrix: each of the r rows in $\mathbf{D}$ is a (retrievable) *block* of s bits (or of s words). In Goldberg's IT-PIR, the communication-optimal setting occurs when $r = s = \sqrt{N}$. However, as the wife of complexity theorist and PIR researcher Bill Gasarch reportedly opined: "A database is NOT an n-bit string" [5]! Insisting that $r = s = \sqrt{N}$ is quite restrictive in practice; one cannot generally rely on all records in a database being of a fixed length, nor on the number of records being somehow related to their lengths. Using a suboptimal choice of parameters can address part of this problem, but it also serves to increase the communication cost of the already costly protocol. A second possible workaround is to

pack multiple records into a block (together with some padding) to handle records that are smaller than $\sqrt{N}$ words, and require users to submit multiple queries to retrieve records that are larger than $\sqrt{N}$ words. Of course, to avoid leaking information to the database servers, it becomes necessary for *all users* to submit the *maximum possible number* of queries needed for *any* record, regardless of the size of the actual record they seek.

We propose a new, communication-efficient way for users to retrieve multiple blocks simultaneously in Goldberg's IT-PIR, and use our new multi-block query technique to support variable-length database records. Multi-block queries trade off some Byzantine robustness to improve throughput without affecting user privacy; in fact, multi-block IT-PIR queries are information-theoretically indistinguishable from standard, single-block IT-PIR queries if the number of colluding database servers does not exceed the user-defined privacy threshold. By taking advantage of the recent Cohn-Heninger multi-polynomial list decoding algorithm [6], we show that reasonable parameter choices allow the user to retrieve several blocks with the same communication cost as a single-block query and still maintain Byzantine robustness up to the original Guruswami-Sudan list decoding bound [7].

We use our new multi-block IT-PIR to construct four new symmetric PIR (SPIR) protocols that each support arbitrary-sized database records. Decoupling the block size from the sizes of individual records in this way lets us fix the block size to its communication-optimal value without artificially restricting the contents and layout of the records. This makes it feasible to host a diverse set of records that may contain, for example, multimedia files of varying sizes in an IT-PIR database. Moreover, three of our four new SPIR constructions are trivial to augment with efficient zero-knowledge proofs about the records a user requests, thus making it easy to implement flexible *pricing* [8], [9] and *access control* [8], [10], [11] structures over the records. The resulting SPIR protocols are therefore well suited to privacy-preserving e-commerce applications, such as privacy-friendly sales of e-books, music, movies, or smart phone and tablet apps.

References

[1] B. Chor, N. Gilboa, and M. Naor, "Private Information Retrieval by Keywords," Technion, Israel, Technical report TR CS0917, 1997.

[2] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private Information Retrieval," *Journal of the ACM*, vol. 45, no. 6, pp. 965–981, November 1998.

[3] F. G. Olumofin and I. Goldberg, "Privacy-Preserving Queries over Relational Databases," in *Privacy Enhancing Technologies*, Berlin, Germany, July 2010, pp. 75–92.

[4] I. Goldberg, "Improving the Robustness of Private Information Retrieval," in *Proceedings of IEEE S&P 2007*, Oakland, California, May 2007, pp. 131–148.

[5] B. Gasarch, December 2011, Comment posted to "Is cryptographic theory practically relevant?". Random Bits. [Online; accessed April 6, 2012] https://jonkatz.wordpress.com/2011/12/16/is-cryptographic-theory-practically-relevant/.

[6] H. Cohn and N. Heninger, "Approximate Common Divisors via Lattices," *IACR Cryptology ePrint Archive*, vol. 2011/437, August 2011.

[7] V. Guruswami and M. Sudan, "Improved Decoding of Reed-Solomon and Algebraic-Geometry Codes," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 1757–1767, September 1999.

[8] R. Henry, F. Olumofin, and I. Goldberg, "Practical PIR for Electronic Commerce," in *ACM CCS 2011*, Chicago, Illinois, October 2011, pp. 677–690.

[9] J. Camenisch, M. Dubovitskaya, and G. Neven, "Unlinkable Priced Oblivious Transfer with Rechargeable Wallets," in *Proceedings of FC 2010*, Tenerife, Canary Islands, January 2010, pp. 66–81.

[10] ——, "Oblivious Transfer with Access Control," in *Proceedings of ACM CCS 2009*, Chicago, Illinois, November 2009, pp. 131–140.

[11] J. Camenisch, M. Dubovitskaya, G. Neven, and G. M. Zaverucha, "Oblivious Transfer with Hidden Access Control Policies," in *Proceedings of PKC 2011*, Taormina, Italy, March 2011, pp. 192–209.