IT-PIR on MPC for Lightweight Devices

Privacy is a fundamental human right [1]. When we search for (or consume) any information on social media, video-streaming sites, etc., service providers can track our queries and interests. They can infer personally identifiable [13] and sensitive information like political preference, sexual orientation, etc., which users may not like to share. Private information retrieval (PIR) [5] is a powerful cryptographic primitive that solves this ubiquitous problem of safeguarding the privacy of users' access patterns to remote, untrusted databases. The information-theoretic PIR (IT-PIR) is the most efficient variant [16] of PIR, which is robust against a quantum adversary but does not consider communication between participating servers. On the other hand, secure multiparty computation (MPC) is another reliable cryptographic technique that can enable multiple mutually-distrusted parties to compute a function on private data having communication between servers. Also, the so-called *n-out-of-n* variant of both IT-PIR and MPC have a similar threat model. Therefore, we aim to achieve *first-ever* IT-PIR functionalities on an MPC framework relaxing existing non-communication assumptions (between IT-PIR servers) to gain *optimal-rate* communication and computation cost for *lightweight* IT-PIR clients, and to utilize higher bandwidth and computing power available at servers. Furthermore, PIR and related tools have been used to build systems for anonymous messaging that are immune from surveillance by governments or corporations. These systems are, however, require expensive computation and communication and have limited scaling abilities. We plan to extend our novel IT-PIR-on-MPC architecture to the so-called symmetric PIR [7, 9] and reverse PIR [14] to support more efficient anonymous messaging.

There are two major classes of PIR: (i) IT-PIR provides perfect privacy (against an unbounded adversary) with non-collusion assumption and non-communication setting among servers [5, 8], and (ii) Computational PIR (CPIR) provides privacy (only against a bounded adversary) with assumptions of computational intractability [10]. Concerning computation and communication overhead at both client- and server-side, IT-PIR is more efficient than CPIR [16]. Besides, CPIR schemes are based on cryptographic tools that are currently considered secure on classic computers but may be compromised by future algorithmic breakthroughs or quantum computers. Given the long-term value of communication privacy, the long-term unbounded security of IT-PIR is a desirable feature. We want to reimagine the IT-PIR setting by relaxing the non-interactive nature between servers and plan to achieve the optimal-rate upstream and downstream communication and computation cost at the capacity-constrained client-side.

In our protocol, we propose that IT-PIR servers form an MPC architecture where each server holds an (n, n)-secret share of the database, i.e., the security holds until at least one of n servers remains honest. We like to investigate how we can achieve the optimal upload cost (i.e., logarithmic in the number of rows of the database) for the client. A potential strategy could be motivated by [3], where the query constructed as a path from the root to the leaf node (on which row index the client is interested in). After receiving the query, we need to inspect how each server can talk by MPC-style interactions with other servers to expand the small query to a full-length CGKS-style [5] (n, n)-secret-share query. That full-length query will be multiplied with the share of the database at each server to generate the response. In this step, we have a couple of interesting research questions to solve: (i) we need to explore the trade-off between the number of optimal interactions between servers, server-side communication, and computation cost, (ii) we have to investigate the most efficient way to multiply (n, n)-additive shares to generate the response at each server. Afterward, the client accumulates all responses to reconstruct the sought record. We plan to explore an optimal download technique for the client motivated by [17]. Here, we reduce the (upstream) communication and computation burden of IT-PIR clients by transferring the task to heavy-weight servers.

Whistleblowers and protesters need anonymity when they broadcast messages to the public to avoid getting caught by dystopian authorities. We like to extend the capacity of state-of-the-art anonymous broadcasting protocols like Riposte [6]. As a building block of the anonymous ecosystem, at first, we plan to inspect how to construct an efficient (n, n)-symmetric PIR protocol that permits a user to fetch only the allowed number of records. We want to investigate the most efficient zero-knowledge proof (ZKP) scheme to enable IT-PIR-on-MPC servers to perform the sanity check of the received query from the client. For efficiency, we need the proof of correctness to scale with the size of the small query instead of the full-length query [2]. Next, we plan to investigate how to build an efficient (n, n)-reverse PIR a.k.a. PIR writing protocol that enables a user to write on the database privately. We like to explore the feasibility of developing reverse PIR by a demultiplexer technique on top of our basic IT-PIR-on-MPC framework. In our anonymous platform, we aim to achieve traffic analysis resistance by applying DC-net-based [4] small epochs, disruption defense by the (n, n)-symmetric PIR, scalability to billions of users by the (n, n)-PIR writing, communication and computation efficiency at the user-side by the underlying IT-PIR-on-MPC architecture. The existing work [6, 11, 12, 15] has brought these closer to practice, but significant gaps remain. We expect that our proof-of-concept implementation will result in order-of-magnitude performance improvements and enable practical application of these concepts even to lower-power mobile devices, such as used by activists.

References

[1] UN General Assembly. Universal declaration of human rights (217 [iii] a), 1948.

[2] William Black and Ryan Henry. There are 10 types of vectors (and polynomials): Efficient zero-knowledge proofs of “one-hotness” via polynomials with one zero. In *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society*, WPES’19, page 37–49, New York, NY, USA, 2019. Association for Computing Machinery.

[3] Elette Boyle, Niv Gilboa, and Yuval Ishai. Function secret sharing: Improvements and extensions. In *Proceedings of CCS 2016*, pages 1292–1303, Vienna, Austria (October, 2016).

[4] David Chaum. Blind signatures for untraceable payments. In *Advances in Cryptology: Proceedings of CRYPTO 1982*, pages 199–203, Santa Barbara, CA, USA (August, 1982).

[5] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private information retrieval. In *Proceedings of FOCS 1995*, pages 41–50, Milwaukee, WI, USA (October, 1995).

[6] Henry Corrigan-Gibbs, Dan Boneh, and David Mazières. Riposte: An anonymous messaging system handling millions of users. In *Proceedings of IEEE S&P 2015*, pages 321–338, San Jose, CA, USA (May, 2015).

[7] Yael Gertner, Shafi Goldwasser, and Tal Malkin. A random server model for private information retrieval or how to achieve information theoretic PIR avoiding database replication. In *Proceedings of RANDOM 1998*, volume 1518 of *LNCS*, pages 200–217, Barcelona, Spain (October, 1998).

[8] Ian Goldberg. Improving the robustness of private information retrieval. In *Proceedings of IEEE S&P 2007*, pages 131–148, Oakland, CA, USA (May, 2007).

[9] Ryan Henry, Femi Olumofin, and Ian Goldberg. Practical PIR for electronic commerce. In *Proceedings of CCS 2011*, pages 677–690, Chicago, IL, USA (October, 2011).

[10] Eyal Kushilevitz and Rafail Ostrovsky. Replication is not needed: Single database, computationally-private information retrieval. In *Proceedings of FOCS 1997*, pages 364–373, Miami Beach, FL, USA (October, 1997).

[11] Albert Kwon, David Lazar, Srinivas Devadas, and Bryan Ford. Riffle: an efficient communication system with strong anonymity. *Proceedings on Privacy Enhancing Technologies*, 2016(2):115 – 134, 2016.

[12] David Lazar and Nickolai Zeldovich. Alpenhorn: Bootstrapping secure communication without leaking metadata. In *Proceedings of the 12th USENIX*

Conference on Operating Systems Design and Implementation, OSDI'16, page 571–586, USA, 2016. USENIX Association.

[13] Arvind Narayanan and Vitaly Shmatikov. Myths and fallacies of "personally identifiable information". *Commun. ACM*, 53(6):24–26 (June, 2010).

[14] Rafail Ostrovsky and Victor Shoup. Private information storage (Extended abstract). In *Proceedings of STOC 1997*, pages 294–303, El Paso, TX, USA (May, 1997).

[15] Ania M. Piotrowska, Jamie Hayes, Tariq Elahi, Sebastian Meiser, and George Danezis. The Loopix anonymity system. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 1199–1216, Vancouver, BC (August, 2017). USENIX Association.

[16] Radu Sion and Bogdan Carbunar. On the practicality of private information retrieval. In *Proceedings of NDSS 2007*, San Diego, CA, USA (March, 2007).

[17] **Syed Mahbub Hafiz** and Ryan Henry. A bit more than a bit is more than a bit better: Faster (essentially) optimal-rate many-server PIR. *Proceedings on Privacy Enhancing Technologies*, 2019(4):112 – 131, 2019.