Tutorial: Private Information Retrieval

Ryan Henry
Indiana University
School of Informatics and Computing
Bloomington, IN 47405, USA
henry@indiana.edu

ABSTRACT

Private information retrieval (PIR) is a cryptographic primitive that facilitates the seemingly impossible task of letting users fetch records from untrusted and remote database servers without revealing to those servers which records are being fetched. The research literature on PIR is vast; in the over two decades since its 1995 introduction by Chor, Goldreich, Kushilevitz, and Sudan, the cryptography, privacy, and theoretical computer science research communities have studied PIR intensively and from a variety of perspectives. Alas, despite a series of significant advances, most privacy practitioners and theoreticians alike fall into one of two camps: (i) those who believe that PIR is so inefficient and abstruse as to make it all-but-useless in practice, or (ii) those who remain blissfully unaware that PIR even exists. Indeed, to date not even one of the numerous PIR-based applications proposed in the research literature has been deployed at scale to protect the privacy of users "in the wild".

This tutorial targets both of the above camps, presenting a bird's-eye overview of the current state of PIR research. Topics covered will span the spectrum from purely theoretical through imminently applicable and all the high points in between, thereby providing participants with an awareness of what modern PIR techniques have (and do not have) to offer, dispelling the myth of PIR's inherent impracticality, and hopefully inspiring participants to identify practical use cases for PIR within their own niche areas of expertise. This introductory tutorial will be accessible to anyone comfortable with college-level mathematics (basic linear algebra and some elementary probability and number theory).

KEYWORDS

Private information retrieval; coding theory; applied cryptography; trusted hardware; function secret sharing

CCS'17, Oct. 30–Nov. 3, 2017, Dallas, TX, USA.

 ISBN 978-1-4503-4946-8/17/10.
DOI: http://dx.doi.org/0.1145/3133956.3136069

DESCRIPTION OF TUTORIAL

This tutorial provides participants with a broad overview of the current state-of-the-art with respect to a cryptographic primitive called *private information retrieval* (PIR). Information retrieval (IR)—roughly, the activity of locating and retrieving relevant information from a corpus (i.e., database) D of files or records—is among the most fundamental and ubiquitous tasks in computing. In "classical" IR settings, a client encodes its information needs in a query q, while the holder of D (whether a remote server or a service running on the client's local host) ranks the elements of D by relevance to the client's needs as expressed in q, and then it returns to the client one or more of the "top" matches.

PIR extends the classical IR setting with an interesting twist: it insists that the holder of D learns nothing about the information needs expressed in q (including the ranking it induces and the matches it returns). While intuition might insist that it is clearly *not* possible to realize such privacy for IR, the research literature contains numerous PIR protocols that do just that (albeit only for some very basic forms of IR), with known constructions based on information- and coding-theoretic primitives (e.g., secret sharing [3, 4] and error-correcting codes [9]), both public- and private-key encryption (e.g., partially homomorphic encryption [7] and function secret sharing [1, 11], anonymity [6, 10], and trusted hardware (e.g., TPMs [12] and SGX). As for what types of IR can be realized as PIR, the literature considers PIR constructions supporting physical position-based queries [3], keyword-based queries [2], simple SQL-based queries [8], statistical queries [11], and index-based queries [5].

Yet despite the ubiquity of IR, the ever-increasing quantity and sensitivity of information being digitized and made available online, and the veritable wealth of PIR techniques in the literature, *PIR has never been deployed at scale to protect the privacy of users "in the wild"*. This tutorial sets out with the lofty goal of changing that. It will cover topics spanning the spectrum from purely theoretical through imminently applicable and all the high points in between, thereby providing participants with an awareness of what modern PIR techniques have (and do not have) to offer, dispelling the myth of PIR's inherent impracticality, and hopefully inspiring participants to identify practical use cases for PIR within their own niche areas of expertise.

SCOPE OF TUTORIAL

The duration of the tutorial will be **three hours**, with each hour being devoted to a different "module" about PIR.

The first half-hour will motivate the PIR problem, illustrating why widely deployed primitives (like encryption and anonymous communications) cannot be relied upon to solve many, or even most, IR-based privacy problems. Following this, the second half-hour will be devoted to formally defining PIR as a cryptographic primitive, using the so-called "trivial PIR" protocol as an ideal (from a privacy perspective) yet impractical reference against which to base the definitions. Throughout the development of formal PIR definitions, participants will learn about some of the fundamental possibility and impossibility results these definitions imply. We will also take this opportunity to clarify the relationships PIR has to oblivious transfer (OT), oblivious RAM (ORAM), and secure multiparty computation (MPC).

The second hour will consist of a discussion about all of the major "flavours" of PIR in the research literature, including information-theoretic PIR (IT-PIR), computational PIR (C-PIR), function secret sharing-based PIR (FSS-PIR), anonymity-based PIR (A-PIR), trusted hardware-based PIR (TH-PIR), and various hybrids thereof. Throughout this discussion, participants will learn about the high-level idea underlying each flavour of PIR as well as their inherent strengths and weaknesses. In addition, we will walk through low-level details for a handful of representative PIR protocols.

The final hour will focus on a hodge-podge of practical considerations like Byzantine robustness, techniques for "batching" queries or batch-coding databases, how to realize expressive query types (e.g., keyword-, SQL-, and index-based PIR queries) or advanced functionality (e.g., ACLs and pricing) atop basic PIR, and open-source implementations.

INTENDED AUDIENCE

This tutorial is intended for a broad audience, including researchers and practitioners from academia, government, and industry. The material covered focuses primarily on technical aspects of PIR and will be most accessible to computer scientists and software developers—the tutorial will not address legal, ethical, sociological, or economic considerations; nonetheless, social scientists and policymakers are welcome and encouraged to participate.

Prerequisite knowledge: This is an *introductory level* tutorial and will be accessible to anyone comfortable with college-level mathematics (specifically, basic linear algebra and some elementary probability and number theory). It does not assume any prior knowledge about PIR or other advanced cryptographic primitives, although participants having some familiarity with cryptographic definitions and hardness assumptions will find a few of the "deep-dive" topics easier to grok.

REFERENCES

[1] Elette Boyle, Niv Gilboa, and Yuval Ishai. Function secret sharing. In *Advances in Cryptology: Proceedings of EUROCRYPT 2015 (Part II)*, volume 9057 of *LNCS*, pages 337–367, Sofia, Bulgaria (April 2015).

[2] Benny Chor, Niv Gilboa, and Moni Naor. Private information retrieval by keywords. Technical Report CS 0917, Technion-Israel Institute of Technology, Haifa, Israel (February 1997).

[3] 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).

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

[5] Syed Mahbub Hafiz and Ryan Henry. Querying for queries: Indexes of queries for efficient and expressive IT-PIR (October–November 2017).

[6] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Cryptography from anonymity. In *Proceedings of FOCS 2006*, pages 239–248, Berkeley, CA, USA (October 2006).

[7] 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).

[8] Femi G. Olumofin and Ian Goldberg. Privacy-preserving queries over relational databases. In *Proceedings of PETS 2010*, volume 6205 of *LNCS*, pages 75–92, Berlin, Germany (July 2010).

[9] Nihar B. Shah, K. V. Rashmi, and Kannan Ramchandran. One extra bit of download ensures perfectly private information retrieval. In *Proceedings of ISIT 2014*, pages 856–860, Honolulu, HI, USA (June–July 2014).

[10] Raphael R. Toledo, George Danezis, and Ian Goldberg. Lower-cost ϵ-private information retrieval. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*, volume 2016(4), pages 184–201, Darmstadt, Germany (October 2016).

[11] Frank Wang, Catherine Yun, Shafi Goldwasser, Vinod Vaikuntanathan, and Matei Zaharia. Splinter: Practical private queries on public data. In *Proceedings of NSDI 2017*, pages 299–313, Boston, MA, USA (March 2017).

[12] Peter Williams and Radu Sion. Usable PIR. In *Proceedings of NDSS 2008*, San Diego, CA, USA (February 2008).

BIOGRAPHICAL SKETCH

Ryan Henry is an assistant professor in the computer science department at Indiana University in Bloomington, Indiana. His research explores the systems challenges of applied cryptography, with an emphasis on using cryptography to build secure systems that preserve the privacy of their users. In addition to designing and analyzing privacy-enhancing systems, Professor Henry is interested in practical matters like implementing and working toward the deployment of such systems, as well as more theoretical matters like devising number-theoretic attacks against non-standard cryptographic assumptions and developing new models and theories to understand just how efficient "heavy-weight" cryptographic primitives can be. He received his MMath (2010) and Ph.D. (2014) from the University of Waterloo, where he held a Vanier Canada Graduate Scholarship (Vanier CGS), the most prestigious graduate scholarship in Canada. He has published several papers on PIR at top research venues (e.g., CCS, NDSS, and PETS), is a contributor to Percy++ (an open-source implementation of PIR protocols in C++), and two of his three active NSF grants heavily involve PIR research.

Acknowledgement: This material is based upon work supported by the National Science Foundation under Grant No. 1718475.