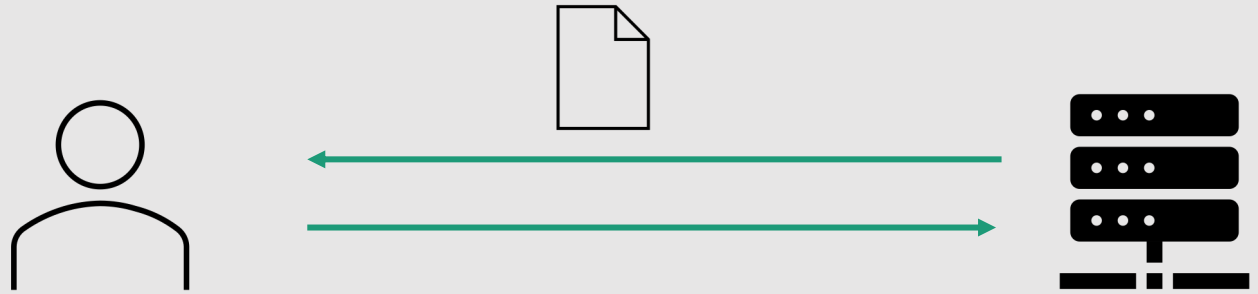
Welcome to WIP!

Workshop in PIR

Organizing Committee:
Will Scott (Protocol Labs)
David Wu (UT Austin)
Shannon Veitch (ETH Zurich)
Elaine Shi (CMU)
Sebastian Angel (U Penn)
Ling Ren (UIUC)

@ PETS 2024, Bristol, UK

Private Information Retrieval



Client wants to look something up in an online database, without revealing the information being retrieved.

Solutions involve: distributed trust, trusted hardware, homomorphic encryption, ...

Applications

- Private DNS lookups
- Compromised password lookup
- Anonymous communication (contact discovery)
- Certificate Transparency auditing
- Private streaming
- Private search

Extensions

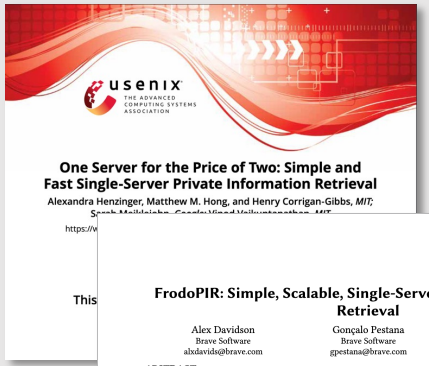
- Index vs Keyword PIR (and more advanced queries)
- Batching
- Malicious security, robustness
- Offline/Online
- Symmetric PIR
- Conceptual extensions (PSI, PSU)

Publishing in PIR

Security/Privacy

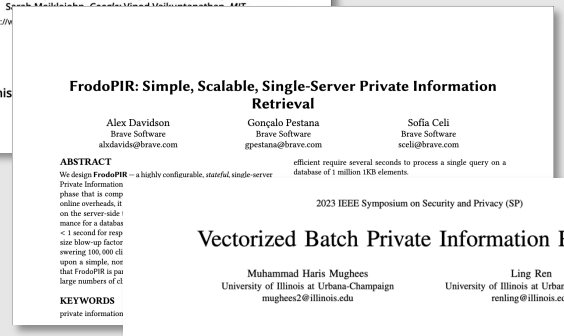
usenix THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

acm SIGSAC


usenix THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

One Server for the Price of Two: Simple and Fast Single-Server Private Information Retrieval

Alexandra Henzinger, Matthew M. Hong, and Henry Corrigan-Gibbs, *MIT*;


FrodoPIR: Simple, Scalable, Single-Server Private Information Retrieval

Alex Davidson
Brave Software
alxdavids@brave.com

Gonçalo Pestana
Brave Software
gpestana@brave.com

Sofia Celi
Brave Software
sceli@brave.com

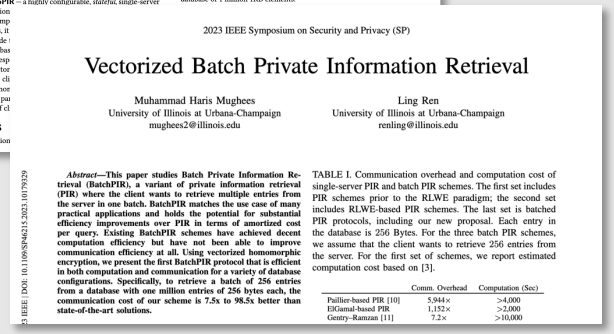
ABSTRACT

We design **FrodoPIR** – a highly configurable, *stateful*, single-server

efficient require several seconds to process a single query on a database of 1 million 1KB elements.

KEYWORDS

private information


2023 IEEE Symposium on Security and Privacy (SP)

Vectorized Batch Private Information Retrieval

Muhammad Haris Mughees
University of Illinois at Urbana-Champaign
mughees2@illinois.edu

Ling Ren
University of Illinois at Urbana-Champaign
renling@illinois.edu

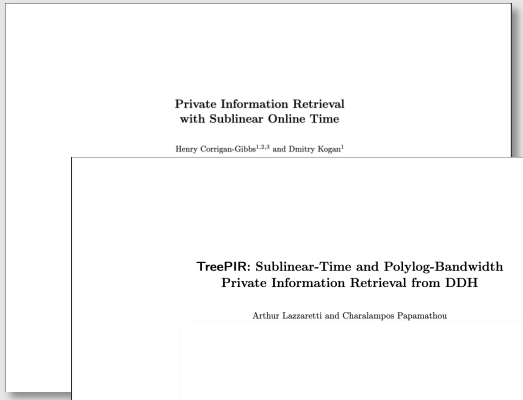
***Abstract*—This paper studies Batch Private Information Retrieval (BatchPIR), a variant of private information retrieval (PIR) where the client wants to retrieve multiple entries from the server in one batch. BatchPIR matches the use case of many practical applications and holds the potential for substantial efficiency improvements over PIR in terms of amortized cost per query. Existing BatchPIR schemes have achieved decent computation efficiency but have not been able to improve communication efficiency at all. Using vectorized homomorphic encryption, we present the first BatchPIR protocol that is efficient in both computation and communication for a variety of database configurations. Specifically, to retrieve a batch of 256 entries from a database with one million entries of 256 bytes each, the communication cost of our scheme is 7.5x to 98.5x better than state-of-the-art solutions.**

TABLE I. Communication overhead and computation cost of single-server PIR and batch PIR schemes. The first set includes PIR schemes prior to the RLWE paradigm; the second set includes RLWE-based PIR schemes. The last set is batched PIR protocols, including our new proposal. Each entry in the database is 256 Bytes. For the three batch PIR schemes, we assume that the client wants to retrieve 256 entries from the server. For the first set of schemes, we report estimated computation cost based on [3].

	Comm. Overhead	Computation (Sec)
Paillier-based PIR [10]	5,944×	>4,000
ElGamal-based PIR	1,152×	>2,000
Gentry–Ramzan [11]	7.2×	>10,000

Crypto/IACR

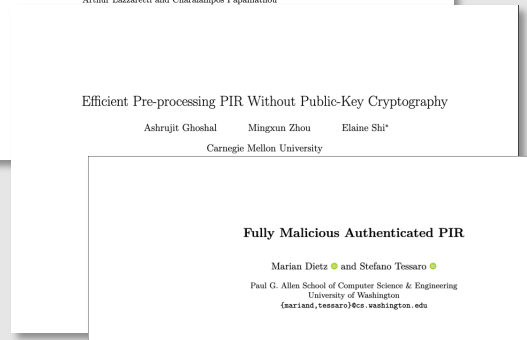
International Association for Cryptologic Research


Private Information Retrieval with Sublinear Online Time

Henry Corrigan-Gibbs[1,2,3] and Dmitry Kogan[1]

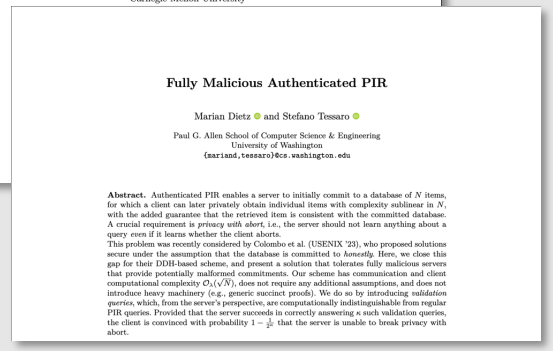
TreePIR: Sublinear-Time and Polylog-Bandwidth Private Information Retrieval from DDH

Arthur Lazzaretti and Charalampos Papamathou


Efficient Pre-processing PIR Without Public-Key Cryptography

Ashrujit Ghoshal Mingxun Zhou Elaine Shi*

Carnegie Mellon University


Fully Malicious Authenticated PIR

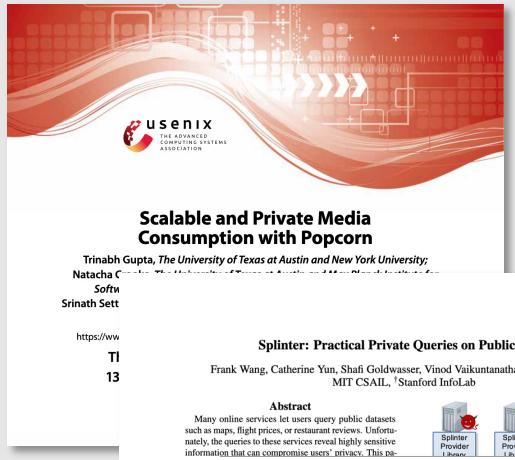
Marian Dietz and Stefano Tessaro

Paul G. Allen School of Computer Science & Engineering
University of Washington
{mariand,tessaro}@cs.washington.edu

Abstract. Authenticated PIR enables a server to initially commit to a database of N items, for which a client can later privately obtain individual items with complexity sublinear in N, with the added guarantee that the retrieved item is consistent with the committed database. A crucial requirement is *privacy with abort*, i.e., the server should not learn anything about a query even if it learns whether the client aborts.

This problem was recently considered by Colombo et al. (USENIX '23), who proposed solutions secure under the assumption that the database is committed to *honestly*. Here, we close this gap for their DDH-based scheme, and present a solution that tolerates fully malicious servers that provide potentially malformed commitments. Our scheme has communication and client computational complexity $\mathcal{O}_\lambda(\sqrt{N})$, does not require any additional assumptions, and does not introduce heavy machinery (e.g., generic succinct proofs). We do so by introducing *validation queries*, which, from the server's perspective, are computationally indistinguishable from regular PIR queries. Provided that the server succeeds in correctly answering s such validation queries, the client is convinced with probability $1 - \frac{1}{2^s}$ that the server is unable to break privacy with abort.

Systems (NSDI, SOSP)

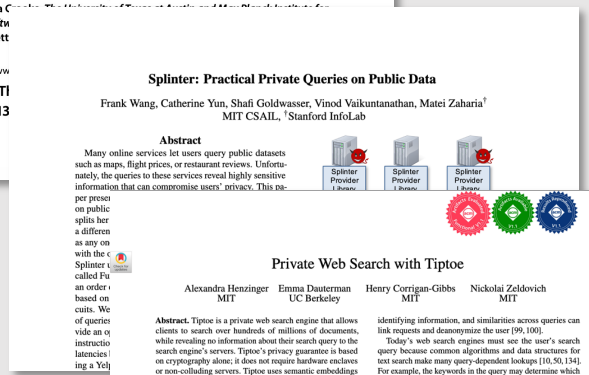

usenix THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

Scalable and Private Media Consumption with Popcorn

Trinabh Gupta, *The University of Texas at Austin and New York University*;

Natacha

Srinath Sett


Splinter: Practical Private Queries on Public Data

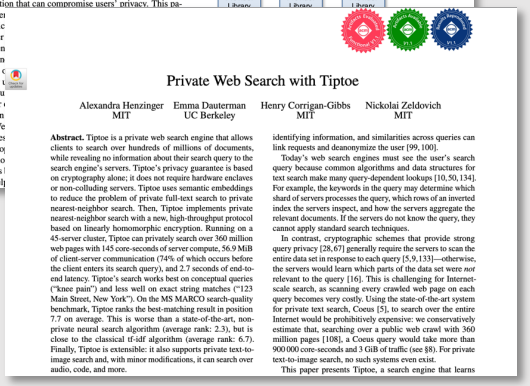
Frank Wang, Catherine Yun, Shafi Goldwasser, Vinod Vaikuntanathan, Matei Zaharia†

MIT CSAIL, †Stanford InfoLab

Abstract

Many online services let users query public datasets such as maps, flight prices, or restaurant reviews. Unfortunately, the queries to these services reveal highly sensitive information that can compromise users' privacy. This pa-

Splinter Provider


Private Web Search with Tiptoe

Alexandra Henzinger
MIT

Emma Dauterman
UC Berkeley

Henry Corrigan-Gibbs
MIT

Nickolai Zeldovich
MIT

Abstract. Tiptoe is a private web search engine that allows clients to search over hundreds of millions of documents, while revealing no information about their search query to the search engine's servers. Tiptoe's privacy guarantee is based on cryptography alone; it does not require hardware enclaves or non-colluding servers. Tiptoe uses semantic embeddings to reduce the problem of private full-text search to private nearest-neighbor search. Then, Tiptoe implements private nearest-neighbor search with a new, high-throughput protocol based on linearly homomorphic encryption. Running on a 45-server cluster, Tiptoe can privately search over 360 million web pages with 145 core-seconds of server compute, 56.9 MiB of client-server communication (74% of which occurs before the client enters its search query), and 2.7 seconds of end-to-end latency. Tiptoe's search works best on conceptual queries ("knee pain") and less well on exact string matches ("123 Main Street, New York"). On the MS MARCO search-quality benchmark, Tiptoe ranks the best-matching result in position 7.7 on average. This is worse than a state-of-the-art, non-private neural search algorithm (average rank: 2.3), but is close to the classical tf-idf algorithm (average rank: 6.7). Finally, Tiptoe is extensible: it also supports private text-to-image search and, with minor modifications, it can search over audio, code, and more.

identifying information, and similarities across queries can link requests and deanonymize the user [99, 100].

Today's web search engines must see the user's search query because common algorithms and data structures for text search make many query-dependent lookups [10, 50, 134]. For example, the keywords in the query may determine which shard of servers processes the query, which rows of an inverted index the servers inspect, and how the servers aggregate the relevant documents. If the servers do not know the query, they cannot apply standard search techniques.

In contrast, cryptographic schemes that provide strong query privacy [28, 67] generally require the servers to scan the entire data set in response to each query [5, 9, 133]—otherwise, the servers would learn which parts of the data set were *not* relevant to the query [16]. This is challenging for Internet-scale search, as scanning every crawled web page on each query becomes very costly. Using the state-of-the-art system for private text search, Coeus [5], to search over the entire Internet would be prohibitively expensive: we conservatively estimate that, searching over a public web crawl with 360 million pages [108], a Coeus query would take more than 900 000 core-seconds and 3 GiB of traffic (see §8). For private text-to-image search, no such systems even exist.

This paper presents Tiptoe, a search engine that learns

WIP Goals

- Provide a *central place* for collaboration on PIR.
- Share early work, directions, and theories for successful PIR construction.
- Provide a point of coordination for PIR researchers and grow a more vibrant community working on the PIR problem.

As a first iteration of WIP, this is very much a work in progress, and we look forward to having your feedback as to what you'd like to gain from this workshop

Program & Information

https://github.com/private-retrieval/wip

WIP: Meta Discussion

- Google form for feedback: https://forms.gle/45Lpvoitdhd6evNQ9