Blindstore

A Privacy-Preserving Data Store

github.com/blindstore/blindstore

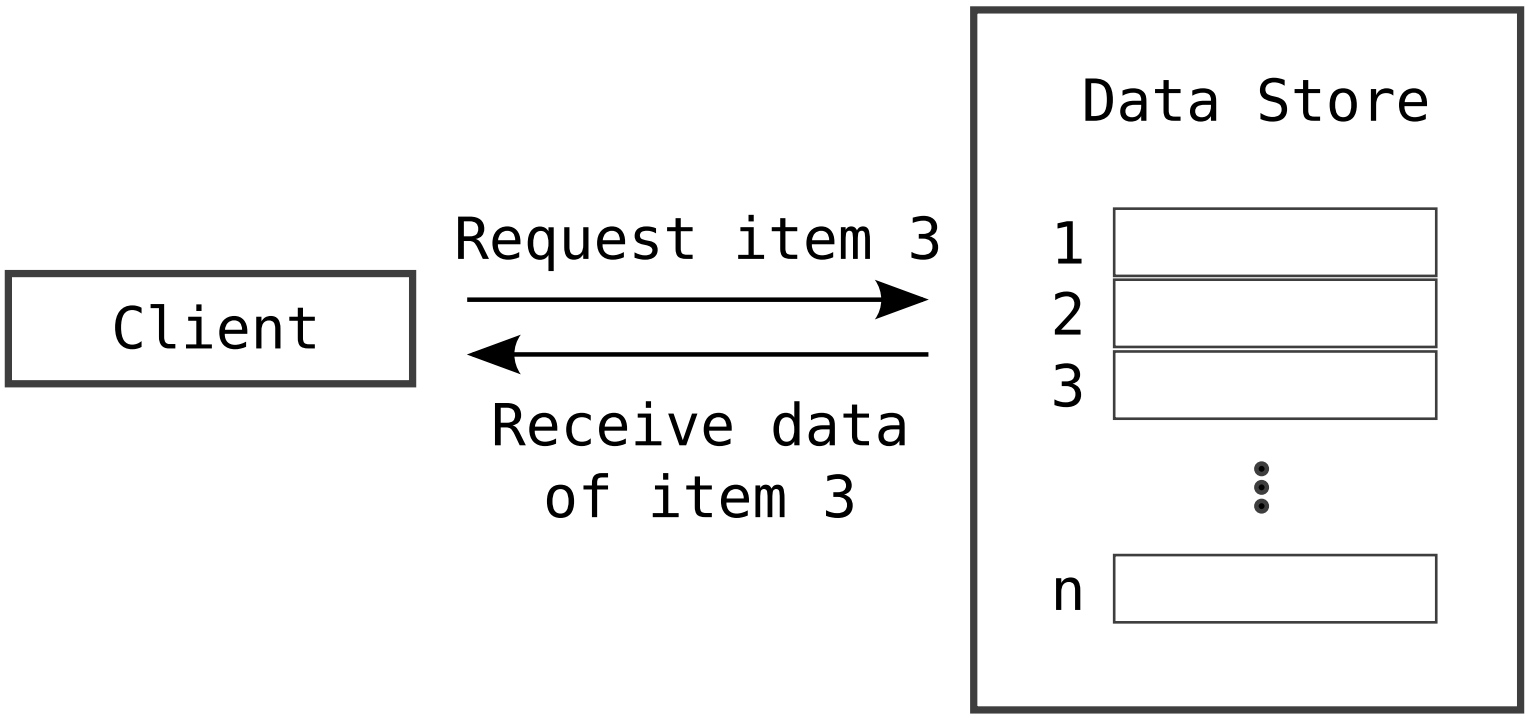
Benjamin Lipp, for the Blindstore team

31st Chaos Communication Congress, Hamburg

December 29th, 2014

Privacy-Preserving Data Store?

- Query an item from the data store and receive the answer:



- But: The data store does not know what you wanted to know

How is it possible?

"Private Information Retrieval"

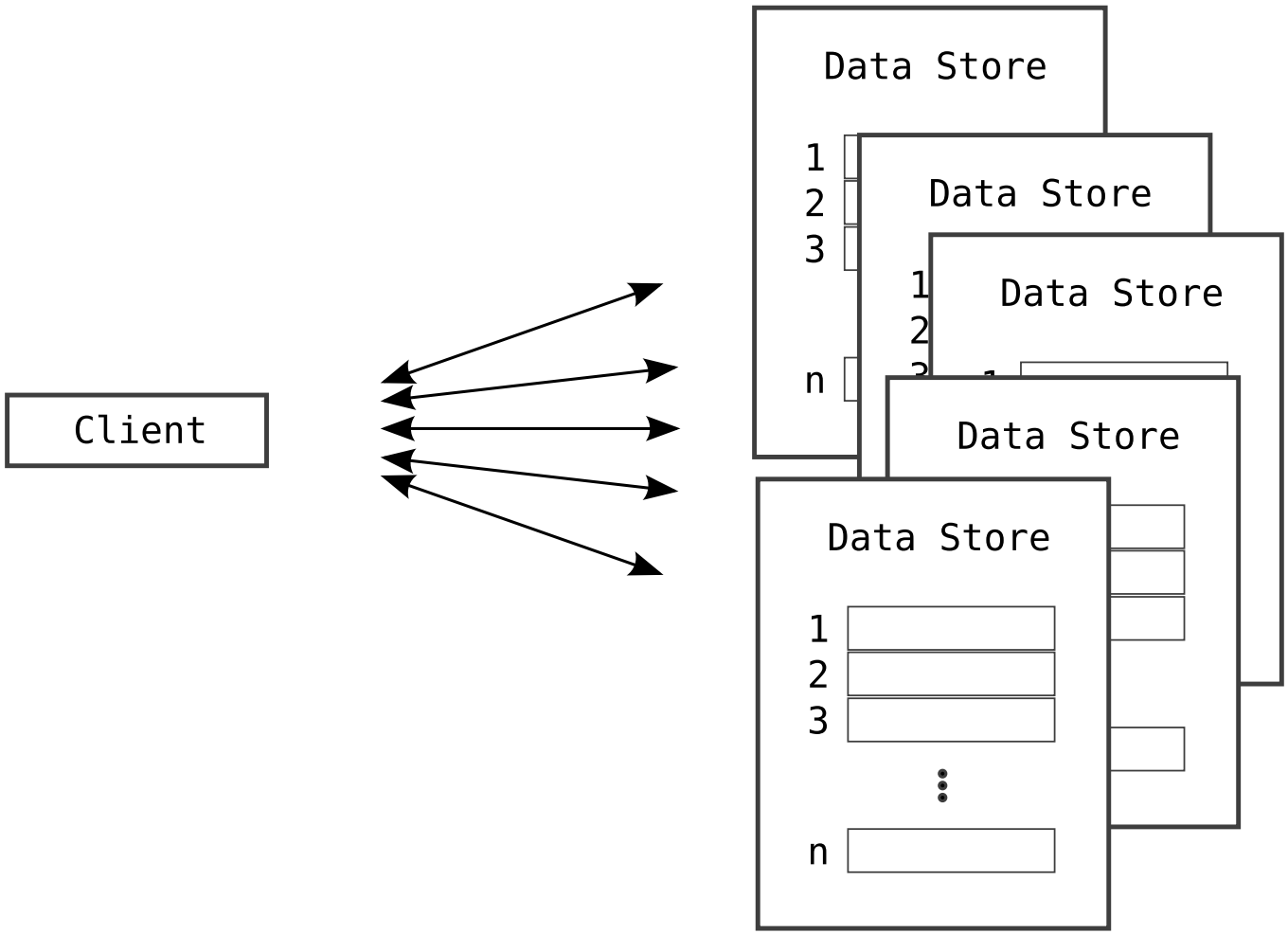
Naive approach:

- Download everything
- Obviously expensive in communication

Use other methods to lower this resource usage.

Information-theoretic PIR

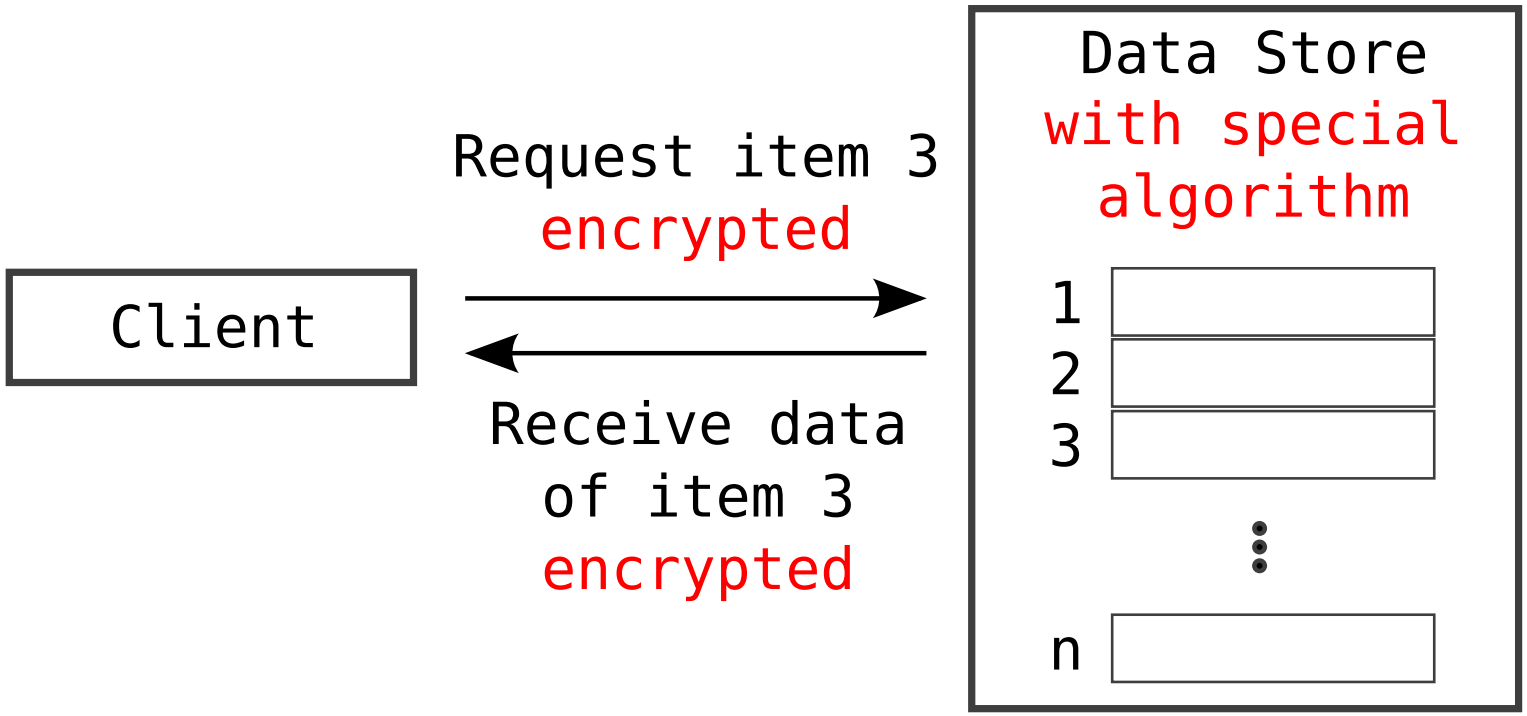
Distribute the question to multiple services:



This works (for example) if at least 2 stores don't collude.

Computational PIR

- Algorithm that runs on the whole database
- Homomorphic encryption



- Decrypt with private key.

(Fully) Homomorphic Encryption

Very roughly, just to get the idea:

Calculations, without knowing the actual content.

- $c_1 = \text{encrypt}(p_1)$
- $c_2 = \text{encrypt}(p_2)$
- $\text{decrypt}(c_1 + c_2) = p_1 + p_2$

Blindstore . . .

- goes for the computational approach.
- is a library providing functions for client and server.
- implements a scheme presented in a paper in 2013.
  - *Single-Database Private Information Retrieval from Fully Homomorphic Encryption* by Xun Yi, Mohammed Golam Kaosar, Russell Paulet, and Elisa Bertino
- Complexity, where n is the database size:
  - Communication: $O(\log(n))$
  - Computation: $O(n \log(n))$
- done in C++ for performance reasons.
- performance: 3 seconds to retrieve 1 record from a database of 1024 records with 1 kilobyte record size, measured on an i5 @ 2.50 GHz

What can PIR be useful for?

Two major ideas:

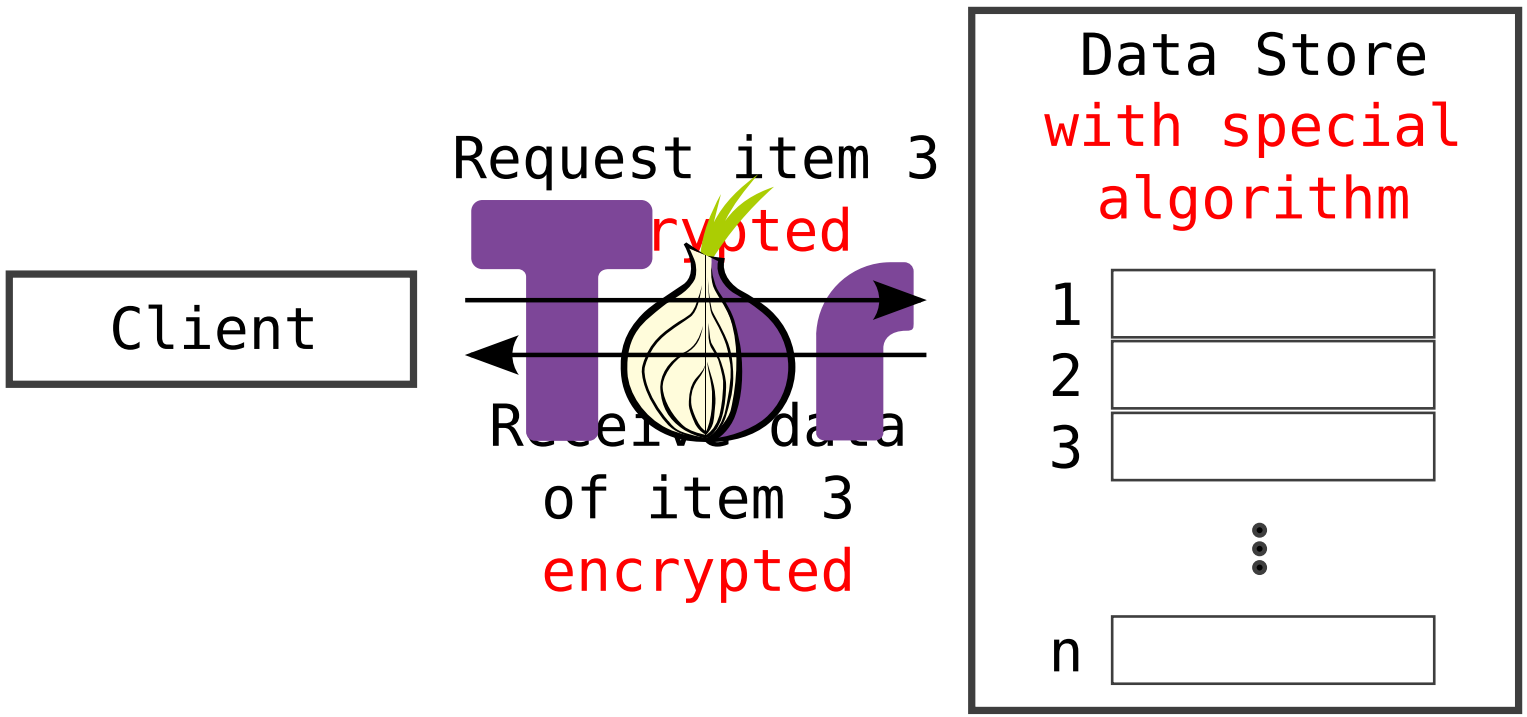
1. The queries of an authenticated user are hidden.
2. Query usage patterns/statistics of a whole community are hidden.

Examples:

- DNS server
- Peer-to-peer signaling/presence discovery

How is it different from TOR?

- TOR hides the user's identity, not the query.
- However, TOR could be used to anonymously connect to Blindstore.



How is it different from CryptDB?

- CryptDB stores data encrypted and thus is only useful for one party.
- Blindstore stores data unencrypted.

The Future

- increase performance: vectorisation, parallelisation, distribution
- choice of privacy level
- build a hashtable instead of just block retrieval by index number

The End, and: Related Work

More details needed?

Assembly `$(pwgen 10 1)` in the Security Cluster
Call 2103

Project Infrastructure

Website: `https://blindstore.github.io/`
Code, Issues: `https://github.com/blindstore/blindstore`
Discussion: `https://gitter.im/blindstore/blindstore`
Mailing list: `https://groups.google.com/group/blindstore`

Related work on site

"DP5: PIR for Privacy-preserving Presence"
by Ian Goldberg, George Danezis and Nikita Borisov

Go to the talk! Today at 17:15, Saal 1