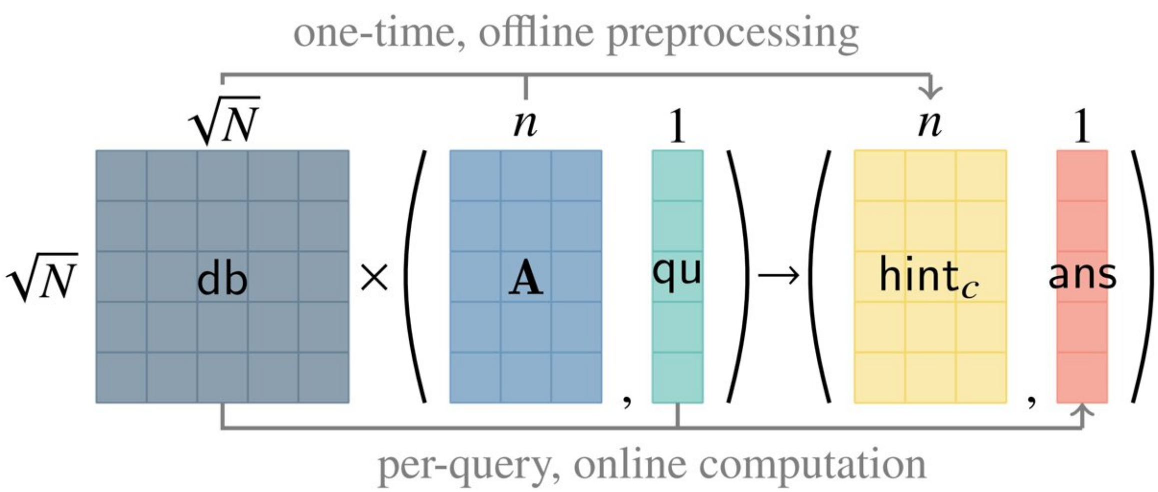
A faster way to preserve privacy online

December 7 2022, by Adam Zewe



The server computation in SimplePIR. Each cell represents a Z? element, and × denotes matrix multiplication. The server performs the bulk of its work in a one-time preprocessing step. Thereafter, the server can answer each client's query with a lightweight online phase. Credit: *One Server for the Price of Two: Simple and Fast Single-Server Private Information Retrieval* (2022).

Searching the internet can reveal information a user would rather keep private. For instance, when someone looks up medical symptoms online, they could reveal their health conditions to Google, an online medical database like WebMD, and perhaps hundreds of these companies' advertisers and business partners.

For decades, researchers have been crafting techniques that enable users to search for and retrieve information from a database privately, but these methods remain too slow to be effectively used in practice.

MIT researchers have now developed a scheme for private information retrieval that is about 30 times faster than other comparable methods. Their technique enables a user to search an online database without revealing their query to the server. Moreover, it is driven by a simple algorithm that would be easier to implement than the more complicated approaches from previous work.

Their technique could enable private communication by preventing a messaging app from knowing what users are saying or who they are talking to. It could also be used to fetch relevant online ads without advertising servers learning a users' interests.

"This work is really about giving users back some control over their own data. In the long run, we'd like browsing the web to be as private as browsing a library. This work doesn't achieve that yet, but it starts building the tools to let us do this sort of thing quickly and efficiently in practice," says Alexandra Henzinger, a computer science graduate student and lead author of a paper introducing the technique.

Preserving privacy

The first schemes for private information retrieval were developed in the 1990s, partly by researchers at MIT. These techniques enable a user to communicate with a remote server that holds a database, and read records from that database without the server knowing what the user is reading.

To preserve privacy, these techniques force the server to touch every single item in the database, so it can't tell which entry a user is searching for. If one area is left untouched, the server would learn that the client is not interested in that item. But touching every item when there may be millions of database entries slows down the query process.

To speed things up, the MIT researchers developed a protocol, known as Simple PIR, in which the server performs much of the underlying cryptographic work in advance, before a client even sends a query. This preprocessing step produces a data structure that holds compressed information about the database contents, and which the client downloads before sending a query.

In a sense, this data structure is like a hint for the client about what is in the database.

"Once the client has this hint, it can make an unbounded number of queries, and these queries are going to be much smaller in both the size of the messages you are sending and the work that you need the server to do. This is what makes Simple PIR so much faster," Henzinger explains.

But the hint can be relatively large in size. For example, to query a 1-gigabyte database, the client would need to download a 124-megabyte hint. This drives up communication costs, which could make the technique difficult to implement on real-world devices.

To reduce the size of the hint, the researchers developed a second technique, known as Double PIR, that basically involves running the Simple PIR scheme twice. This produces a much more compact hint that is fixed in size for any database.

Using Double PIR, the hint for a 1 gigabyte database would only be 16 megabytes.

"Our Double PIR scheme runs a little bit slower, but it will have much lower communication costs. For some applications, this is going to be a desirable tradeoff," Henzinger says.

Hitting the speed limit

They tested the Simple PIR and Double PIR schemes by applying them to a task in which a client seeks to audit a specific piece of information about a website to ensure that website is safe to visit. To preserve privacy, the client cannot reveal the website it is auditing.

The researchers' fastest technique was able to successfully preserve privacy while running at about 10 gigabytes per second. Previous schemes could only achieve a throughput of about 300 megabytes per second.

They show that their method approaches the theoretical speed limit for private information retrieval—it is nearly the fastest possible scheme one can build in which the server touches every record in the database, adds Corrigan-Gibbs.

In addition, their method only requires a single server, making it much simpler than many top-performing techniques that require two separate servers with identical databases. Their method outperformed these more complex protocols.

"I've been thinking about these schemes for some time, and I never thought this could be possible at this speed. The folklore was that any single-server scheme is going to be really slow. This work turns that whole notion on its head," Corrigan-Gibbs says.

While the researchers have shown that they can make PIR schemes much faster, there is still work to do before they would be able to deploy their techniques in real-world scenarios, says Henzinger. They would like to cut the communication costs of their schemes while still enabling them to achieve high speeds. In addition, they want to adapt their techniques to handle more complex queries, such as general SQL queries, and more demanding applications, such as a general Wikipedia search. And in the long run, they hope to develop better techniques that can preserve privacy without requiring a server to touch every database item.

"I've heard people emphatically claiming that PIR will never be practical. But I would never bet against technology. That is an optimistic lesson to learn from this work. There are always ways to innovate," senior author Vinod Vaikuntanathan, an EECS professor and principal investigator in CSAIL, says.

"This work makes a major improvement to the practical cost of private information retrieval. While it was known that low-bandwidth PIR schemes imply public-key cryptography, which is typically orders of magnitude slower than private-key cryptography, this work develops an ingenious method to bridge the gap. This is done by making a clever use of special properties of a public-key encryption scheme due to Regev to push the vast majority of the computational work to a precomputation step, in which the server computes a short 'hint' about the database," says Yuval Ishai, a professor of computer science at Technion (the Israel Institute of Technology), who was not involved in the study.

"What makes their approach particularly appealing is that the same hint can be used an unlimited number of times, by any number of clients. This renders the (moderate) cost of computing the hint insignificant in a typical scenario where the same database is accessed many times."

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

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology