Protecting web users' privacy

March 23 2017, by Larry Hardesty

Most website visits these days entail a database query — to look up airline flights, for example, or to find the fastest driving route between two addresses. Credit: Massachusetts Institute of Technology

Most website visits these days entail a database query—to look up airline flights, for example, or to find the fastest driving route between two addresses.

But online database queries can reveal a surprising amount of information about the people making them. And some travel sites have been known to jack up the prices on flights whose routes are drawing an unusually high volume of queries.

At the USENIX Symposium on Networked Systems Design and Implementation next week, researchers from MIT's Computer Science and Artificial Intelligence Laboratory and Stanford University will present a new encryption system that disguises users' database queries so that they reveal no private information.

The system is called Splinter because it splits a query up and distributes it across copies of the same database on multiple servers. The servers return results that make sense only when recombined according to a procedure that the user alone knows. As long as at least one of the servers can be trusted, it's impossible for anyone other than the user to determine what query the servers executed.

"The canonical example behind this line of work was public patent databases," says Frank Wang, an MIT graduate student in electrical engineering and computer science and first author on the conference paper. "When people were searching for certain kinds of patents, they gave away the research they were working on. Stock prices is another example: A lot of the time, when you search for stock quotes, it gives away information about what stocks you're going to buy. Another example is maps: When you're searching for where you are and where you're going to go, it reveals a wealth of information about you."

Honest broker

Of course, if the site that hosts the database is itself collecting users' data without their consent, the requirement of at least one trusted server is difficult to enforce.

Wang, however, points to the increasing popularity of services such as DuckDuckGo, a search engine that uses search results from other sites, such as Bing and Yahoo, but vows not to profile its customers.

"We see a shift toward people wanting private queries," Wang says. "We can imagine a model in which other services scrape a travel site, and maybe they volunteer to host the information for you, or maybe you subscribe to them. Or maybe in the future, travel sites realize that these services are becoming more popular and they volunteer the data. But right now, we're trusting that third-party sites have adequate protections, and with Splinter we try to make that more of a guarantee."

Division of labor

Splinter uses a technique called function secret sharing, which was first described in a 2015 paper by a trio of Israeli computer scientists. One of them, Elette Boyle, earned her PhD at MIT studying with RSA Professor of Computer Science and Engineering Shafi Goldwasser, a 2013 recipient of the Turing Award, the highest award in computer science. Goldwasser, in turn, is one of Wang's co-authors on the new paper, along with Vinod Vaikuntanathan, an MIT associate professor of electrical engineering and computer science (EECS); Catherine Yun, an EECS graduate student; and Matei Zaharia, an assistant professor of computer science at Stanford.

Systems for disguising database queries have been proposed in the past, but function secret sharing could make them as much as 10 times faster. In experiments, the MIT and Stanford researchers found that Splinter could return a result from a database with millions of entries—including a duplicate of the Yelp database for selected cities—in about a second.

With function secret sharing, a database query is converted into a set of complementary mathematical functions, each of which is sent to a different database server. On each server, the function must be applied to every record in the database; otherwise, a spy could determine what data the user is interested in. Every time the function is applied to a new record, it updates a value stored in memory. After it's been applied to the last record, the final value is returned to the user. But that value is meaningless until it's combined with the values reported by the other servers.

Splinter represents several key elaborations on previous work on function secret sharing. Whereas earlier research focused on concealing simple binary-comparison and addition operations, Splinter executes more complex operations typical of database queries, such as finding a specified number of records with the highest or lowest values for some variable—such as the 10 lowest fares for a particular flight itinerary. The MIT and Stanford researchers had to devise cryptographic functions that could perform all the comparing and sorting required for ranking results without betraying any information.

Practical considerations

Splinter has also been engineered to run efficiently on real database systems. Most modern computer chips, for instance, are hardwired to implement the encryption scheme known as AES. Hardwiring makes AES hundreds of times faster than it would be if it were implemented in software, but AES has some idiosyncrasies that make it less than ideal for function secret sharing. Through a clever combination of software processes and AES encryption, the MIT and Stanford researchers were able to make Splinter 2.5 times as efficient as it would be if it used the AES circuits alone.

"There's always this gap between something being proposed on paper and actually implementing it," Wang says. "We do a lot of optimization to get it to work, and we have to do a lot of tricks to get it to support actual database queries."

Provided by Massachusetts Institute of Technology

Citation: Protecting web users' privacy (2017, March 23) retrieved 26 April 2024 from https://phys.org/news/2017-03-web-users-privacy.html