Developers get open source boost for data privacy protection

September 8 2019, by Nancy Cohen

Credit: CC0 Public Domain

Google has announced that it is releasing the open-source version of a differential privacy library that helps power some of its own products.

Google invested in differential privacy protection, for example, in

gauging how popular a specific restaurant's dish is on Google Maps, said *Engadget*. Now the tool might help other developers reach a desired level of differential privacy defense. "By releasing its homegrown differential privacy tool, Google will make it easier for any company to boost its privacy bona fides," said *Wired*.

Back up. What is differential privacy? This is data science parlance. Here is how Lily Hay Newman described it in *Wired*: It strategically adds random noise to user information stored in databases so that companies can still analyze it without being able to single people out.

Previously, Andy Greenberg had covered it in *Wired* as "a mathematical tool," managing to mine user data while all the same protecting user privacy. How? Enter that addition of "random noise" to the person's information "before it's uploaded to the cloud."

Nick Statt in *The Verge* offered another snapshot of differential privacy as "a cryptographic approach to data science, particularly with regard to analysis, that allows someone relying on software-aided analysis to draw insights from massive datasets while protecting user privacy."

In May, Nicolas Sartor in *Aircloak* wrote about how it was related to anonymization, a term which many more are already familiar. "When dealing with data anonymization, one inevitably encounters differential privacy. Many privacy researchers regard it as the 'gold standard' of anonymization. Well-known tech companies such as Apple or Google are using it for certain data analyses and market it to raise public awareness underlining their focus on data protection."

As for the Google announcement, Newman reported on what developers get: (1) a set of open source differential privacy libraries that offer equations and models needed to set boundaries and constraints on identifying data and (2) an interface to make it easier for more

developers to implement the protections.

Miguel Guevara, Product Manager, Privacy and Data Protection Office, posted something on the Google Developers blog on Thursday that makes it clear not only developers—but businesses and in turn the people they serve— can stand to gain from strong privacy protections, while the open source library was designed to meet the needs of developers.

"Whether you're a city planner, a small business owner, or a software developer, gaining useful insights from data can help make services work better and answer important questions. But, without strong privacy protections, you risk losing the trust of your citizens, customers, and users."

Guevera said "From medicine, to government, to business, and beyond, it's our hope that these open-source tools will help produce insights that benefit everyone." Guevara offered an example of how the analysis might be implemented by researchers in health care.

"Differentially-private data analysis...enables organizations to learn from the majority of their data while simultaneously ensuring that those results do not allow any individual's data to be distinguished or re-identified...For example, if you are a health researcher, you may want to compare the average amount of time patients remain admitted across various hospitals in order to determine if there are differences in care. Differential privacy is a high-assurance, analytic means of ensuring that use cases like this are addressed in a privacy-preserving manner."

The GitHub page said the project has "a C++ library of ε-differentially private algorithms, which can be used to produce aggregate statistics over numeric data sets containing private or sensitive information. In addition, we provide a stochastic tester to check the correctness of the

algorithms."

Actually, that stochastic tester is what *Help Net Security*'s Zeljka Zorz found to be the most important things about the release. She said it was to help spot glitches and problems in implementation "that could make the differential privacy property no longer hold. This will allow developers to make sure their implementation works as it should."

Her comment had resonance in light of what Newman in *Wired* said about experts strongly discouraging developers from attempting to "roll your own" differential privacy scheme, or design one from scratch. "Google hopes that its open source tool will be easy enough to use that it can be a one-stop shop for developers who might otherwise get themselves into trouble."

CNET quoted Bryant Gipson, an engineering manager at Google, in an interview. "The aim of this is to provide a library of primary algorithms that you could build any type of differential privacy solution on top of."

More information: developers.googleblog.com/2019 … d-organizations.html