What is differential privacy and how can it protect your data?

March 19 2018, by Tianqing Zhu

Tech companies can use differential privacy to collect and share aggregate data about user habits, while maintaining individual privacy. Credit: Tim Snell/Flickr, CC BY-ND

It's no secret that big tech companies like Facebook, Google, Apple and Amazon are increasingly infiltrating our personal and social interactions

to collect vast amounts of data on us every day. At the same time, privacy violations in cyberspace regularly make front page news.

So how should privacy be protected in a world where data is gathered and shared with increasing speed and ingenuity?

Differential privacy is a new model of cyber security that proponents claim can protect personal data far better than traditional methods.

The maths it is based on was developed 10 years ago, and the method has been adopted by Apple and Google in recent years.

What is differential privacy?

Differential privacy makes it possible for tech companies to collect and share aggregate information about user habits, while maintaining the privacy of individual users.

For example, say you wanted to show the most popular routes people take walking through a park. You track the routes of 100 people who regularly walk through the park, and whether they walk on the path or through the grass.

But instead of sharing the specific people taking each route, you share the aggregate data collected over time. People viewing your results might know that 60 out of 100 people prefer to take a short-cut through the grass, but not which 60 people.

Why do we need it?

Many of the world's governments have strict policies about how tech companies collect and share user data. Companies who do not follow the

rules can face huge fines. A Belgian court recently ordered Facebook to stop collecting data on users' browsing habits on external websites, or face fines of €250,000 a day.

For many companies, especially multinationals operating in different jurisdictions, this leaves them in a delicate position when it comes to the collection and use of customer data.

On the one hand, these companies need users' data so they can provide high-quality services that benefit users, such as personalised recommendations. On the other hand, they may face charges if they collect too much user data, or if they try to move data from one jurisdiction to another.

Traditional privacy-preserving tools such as cryptography can't resolve this dilemma since it prevents tech companies from accessing the data at all. And anonymity reduces the value of data – an algorithm can't serve you personalised recommendations if it doesn't know what your habits are.

How does it work?

Let's continue the example of walking routes through a park. If you know the identities of those included in the study, but you don't know who took which route then you might assume that privacy is protected. But that may not be the case.

Say someone viewing your data wants to ascertain if Bob prefers to walk through the grass or on the path. They have obtained background information about the other 99 people in the study, which tells them that 40 people prefer to walk on the path and 59 prefer to walk through the grass. Therefore, they can deduce that Bob, who is the 100th person in the database, is the 60th person who prefers to walk through the grass.

This type of attack is a called a differentiated attack, and it is quite hard to defend against as you cannot control how much background knowledge someone can obtain. Differential privacy aims to defend against this type of attack.

Someone deducing your walking route might not sound too serious, but if you replace walking routes with HIV test results, then you can see there is potential for a serious invasion of privacy.

The differential privacy model guarantees that even if someone has complete information about 99 of 100 people in a data set, they still cannot deduce the information about the final person.

The primary mechanism to achieve that is to add random noise to the aggregate data. In the path example, you may say the number of people who prefer to cross the grass is 59 or 61, rather than exact number of 60. The inaccurate number can preserve the privacy of Bob, but it will have very little impact on the pattern: around 60% people prefer to take a short-cut.

The noise is carefully designed. When Apple employed differential privacy in iOS 10, it added noise to individual user inputs. That means it can track, for example, the most frequently used emojis, but the emoji usage of any individual user is masked.

Cynthia Dwork, the inventor of the differential privacy, has proposed wonderful mathematical proofs on how much noise is enough to achieve the requirement of differential privacy.

What are its practical applications?

Differential privacy can be applied to everything from recommendation systems to location-based services and social networks. Apple uses

differential privacy to gather anonymous usage insights from devices like iPhones, iPads and Macs. The method is user-friendly, and legally in the clear.

Differential privacy would also allow a company like Amazon access your personalised shopping preferences while hiding sensitive information about your historical purchase list. Facebook could use it to collect behavioural data for targeted advertising, without violating a country's privacy policies.

How could it be used in the future?

Different countries have varying privacy policies, and sensitive documents currently have to be manually checked before they move from one country to another. This is time-consuming and expensive.

Recently, a team from Deakin University developed differential privacy technology to automate privacy processes within cloud-sharing communities across countries.

They propose using mathematical formulas to model the privacy laws of each country that could be translated to "middleware" (software) to ensure the data conforms. Employing differential privacy in this way could protect the privacy of users and resolve a data sharing headache for tech companies.

This article was originally published on The Conversation. Read the original article.

Provided by The Conversation

Citation: What is differential privacy and how can it protect your data? (2018, March 19)