People want data privacy but don't always know what they're getting

22 October 2020, by Gabriel Kaptchuk, Elissa M. Redmiles and Rachel Cummings

Credit: Unsplash/CC0 Public Domain

The Trump administration's move to ban the popular video app TikTok has stoked fears about the Chinese government collecting personal information of people who use the app. These fears underscore growing concerns Americans have about digital privacy generally.

Debates around privacy might seem simple: Something is private or it's not. However, the technology that provides digital privacy is anything but simple.

Our data privacy research shows that people's hesitancy to share their data stems in part from not knowing who would have access to it and how organizations that collect data keep it private. We've also found that when people are aware of data privacy technologies, they might not get what they expect.

Differential privacy explained

While there are many ways to provide privacy for people who share their data, differential privacy has recently emerged as a leading technique and is being rapidly adopted.

Imagine your local tourism committee wanted to find out the most popular places in your area. A simple solution would be to collect lists of all the locations you have visited from your mobile device, combine it with similar lists for everyone else in your area, and count how often each location was visited. While efficient, collecting people's sensitive data in this way can have dire consequences. Even if the data is stripped of names, it may still be possible for a data analyst or a hacker to identify and stalk individuals.

Differential privacy can be used to protect everyone's personal data while gleaning useful information from it. Differential privacy disguises individuals' information by randomly changing the lists of places they have visited, possibly by removing some locations and adding others. These introduced errors make it virtually impossible to compare people's information and use the process of elimination to determine someone's identity. Importantly, these random changes are small enough to ensure that the summary statistics—in this case, the most popular places—are accurate.

In practice, differential privacy isn't perfect. The randomization process must be calibrated carefully. Too much randomness will make the summary statistics inaccurate. Too little will leave people vulnerable to being identified. Also, if the randomization takes place after everyone's unaltered data has been collected, as is common in some versions of differential privacy, hackers may still be able to get at the original data.

When differential privacy was developed in 2006, it was mostly regarded as a theoretically interesting tool. In 2014, Google became the first company to start publicly using differential privacy for data collection.

Since then, new systems using differential privacy have been deployed by Microsoft, Google and the U.S. Census Bureau. Apple uses it to power

machine learning algorithms without needing to see your data, and Uber turned to it to make sure their internal data analysts can't abuse their power. Differential privacy is often hailed as the solution to the online advertising industry's privacy issues by allowing advertisers to learn how people respond to their ads without tracking individuals.

Reasonable expectations?

But it's not clear that people who are weighing whether to share their data have clear expectations about, or understand, differential privacy.

In July, we, as researchers at Boston University, the Georgia Institute of Technology and Microsoft Research and the Max Planck Institute, surveyed 675 Americans to evaluate whether people are willing to trust differentially private systems with their data.

We created descriptions of differential privacy based on those used by companies, media outlets and academics. These definitions ranged from nuanced descriptions that focused on what differential privacy could allow a company to do or the risks it protects against, descriptions that focused on trust in the many companies that are now using it and descriptions that simply stated that differential privacy is "the new gold standard in data privacy protection," as the Census Bureau has described it.

Americans we surveyed were about twice as likely to report that they would be willing to share their data if they were told, using one of these definitions, that their data would be protected with differential privacy. The specific way that differential privacy was described, however, did not affect people's inclination to share. The mere guarantee of privacy seems to be sufficient to alter people's expectations about who can access their data and whether it would be secure in the event of a hack. In turn, those expectations drive people's willingness to share information.

Troublingly, people's expectations of how protected their data will be with differential privacy are not always correct. For example, many differential privacy systems do nothing to protect user data from lawful law enforcement searches, but 20% of respondents expected this protection.

The confusion is likely due to the way that companies, media outlets and even academics describe differential privacy. Most explanations focus on what differential privacy does or what it can be used for, but do little to highlight what differential privacy can and can't protect against. This leaves people to draw their own conclusions about what protections differential privacy provides.

Building trust

To help people make informed choices about their data, they need information that accurately sets their expectations about privacy. It's not enough to tell people that a system meets a "gold standard" of some types of privacy without telling them what that means. Users shouldn't need a degree in mathematics to make an informed choice.

Identifying the best ways to clearly explain the protections provided by differential privacy will require further research to identify which expectations are most important to people who are considering sharing their data. One possibility is using techniques like privacy nutrition labels.

Helping people align their expectations with reality will also require companies using differential privacy as part of their data collecting activities to fully and accurately explain what is and isn't being kept private and from whom.

This article is republished from The Conversation under a Creative Commons license. Read the

Provided by The Conversation

APA citation: People want data privacy but don't always know what they're getting (2020, October 22) retrieved 16 October 2021 from https://techxplore.com/news/2020-10-people-privacy-dont-theyre.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.