What if you could better control what mobile apps do with your data?

August 17 2020, by Daniel Tkacik

Credit: CC0 Public Domain

Every year, mobile app developers make billions of dollars selling data they collect from the mobile apps on cell phones, and they aren't making it easy for consumers to prevent it.

While both Apple iOS and Android have introduced a growing collection of privacy permission settings that, in theory, give users more control over their data, studies have shown that users are still overwhelmed and are unable to take advantage of them. In particular, the privacy controls fail to distinguish between different purposes for which data is collected.

"Right now, a user can't tell an app, "Yes, you may access my location for travel directions, but no, you may not sell that same data to third-party analytics companies," says CyLab's Daniel Smullen, a software engineering Ph.D. student. Smullen is the lead author of a new study that explores how, using machine learning, it may be possible to offer users the best of both worlds: the ability to better control their data without overwhelming them with an unmanageable number of privacy decisions.

The study, presented by Carnegie Mellon CyLab researchers at last month's Privacy Enhancing Technologies Symposium, is part of the Personalized Privacy Assistant Project, an effort led by CyLab and Institute for Software Research professor Norman Sadeh to develop a digital "privacy assistant" that can help people manage the ever-increasing number of privacy choices they are presented with.

"The current choices are black or white: Either you allow the app to access your location or you don't, but you can't selectively control how the data will be used," says Sadeh. "Yet many users have been shown to have significant reservations about some possible uses of their data while being comfortable with others. These preferences vary from one user to another."

Smullen's and Sadeh's study shows that using machine learning, it is possible to develop privacy assistants that can accurately infer many of the privacy decisions a user would want to make simply by asking them a small number of questions such as questions about their privacy

preferences.

"The privacy assistants are not intended to take decisions away from users but rather to simplify the number and types of decisions they have to make," says Smullen. "We evaluated configurations where the assistants make privacy setting recommendations users can review before deciding whether or not to accept them."

Sadeh says the results show that operating systems such as iOS and Android could do much better when it comes to allowing users to effectively control how data collected about them is being used.

"Users do not need to be overwhelmed with an unmanageable number of privacy decisions to regain control over their data," says Sadeh. "There is a better path."

In the study, the researchers show that mobile app users can be organized into privacy preference "profiles" based on a small number of questions–typically five to six in most cases. In rare instances—roughly 10% of the time—the machine learning doesn't have enough data to make a prediction and will ask one or two follow-up questions.

Smullen says the difference in user burden in this new model versus the status quo is huge. Right now, if an average user has 70 apps installed on their phone and needs to decide on three privacy settings per app, they have to configure over 200 settings themselves. But with this machine-learning-enabled privacy assistant, users simply need to answer just a handful of questions in order for the privacy assistant to provide accurate recommendations for most of their settings.

To understand and model the relationship between various attributes about users and their privacy preferences, the researchers surveyed nearly 1000 Android users in the U.S. about their preferences to allow or

deny permissions for a wide variety of apps in the Google Play store. The survey also gathered some data about the users themselves, such as their age, their education level, and their location of residence.

"The surprising finding here is that when you include permissions that take the purpose of data collection into account, you actually end up with more powerful predictive models than if you ignored purpose information," says Smullen. "In other words, we are able to support a permission model that is both more accurate, by taking into account purpose information, and more manageable, because it lends itself to more accurate recommendations."

More information: www.petsymposium.org/2020/file … popets-2020-0011.pdf

Provided by Carnegie Mellon University