Mobile app behavior often appears at odds with privacy policies

November 15 2016

Credit: Peter Griffin/Public Domain

How a mobile app says it will collect or share a user's personal information with third parties often appears to be inconsistent with how the app actually behaves, a new automated analysis system developed by Carnegie Mellon University has revealed.

An analysis of almost 18,000 popular free apps from the Google Play

store found almost half lacked a privacy policy, even though 71 percent of those appear to be processing personally identifiable information and would thus be required to explain how under state laws such as the California Online Privacy Protection Act (CalOPPA).

Even those apps that had policies often had inconsistencies. For instance, as many as 41 percent of these apps could be collecting location information and 17 percent could be sharing that information with third parties without stating so in their privacy policy.

"Overall, each app appears to exhibit a mean of 1.83 possible inconsistencies and that's a huge number," said Norman Sadeh, professor of computer science in CMU's Institute for Software Research. The number of discrepancies is not necessarily surprising to privacy researchers, he added, "but if you're talking to anyone else, they're likely to say 'My goodness!'"

Sebastian Zimmeck, a post-doctoral associate who designed and implemented this system with Sadeh, will present their findings Nov. 17-19 at the AAAI Fall Symposium on Privacy and Language Technologies in Arlington, Va.

A number of federal and state laws require mobile apps to have privacy policies, such as the Children's Online Privacy Protection Act (COPPA) for mobile apps directed at children that collect personally identifiable information. However, CalOPPA requires privacy policies for any mobile app that collects personally identifiable information, regardless of whether it is directed at children or adults. Delaware has a similar law. Those state laws effectively serve as a minimum privacy threshold because app publishers usually don't market state-specific apps.

Sadeh's group is collaborating with the California Office of the Attorney General to use a customized version of its system to check for

compliance with CalOPPA and to assess the effectiveness of CalOPPA and "Do Not Track" legislation.

"Just because the automated system finds a possible privacy requirement inconsistency in an app does not mean that a problem necessarily exists," Sadeh emphasized.

It's possible that when the system looks for the privacy policy it fails to find it—the system currently looks at the app store, the company's website and also scans the code of the app. It is also possible that when it analyzes the app's source code it may make mistaken assumptions about how the code handles personal information.

"That's why a human would need to validate findings by the automated system before any enforcement or corrective action took place," Sadeh said.

Nevertheless, with substantially more than a million apps already in Google Play and the number growing by the day, such a system might help developers detect problems with their apps before they are marketed and could help make spotting violators of laws a more manageable task for regulators and privacy activists.

The automated system uses natural language processing and machine learning to analyze the text of privacy policies. It then examines the app's computer code to see whether its behavior suggests it shares personal information and therefore should have a privacy policy. It also checks whether the app's data collection and sharing behavior is consistent with an existing privacy policy.

This approach is far faster than any human review. Two years ago, for instance, 1,200 apps were reviewed in a week's time by the joint efforts of 26 international privacy enforcement agencies. The Carnegie Mellon

system, by comparison, was able to review almost 18,000 in about 31 hours, or about 6 seconds per app.

"With a few servers, we should be able to scan all the free apps in the Google Play store every month," Sadeh said.

It would still require a lot of manpower to systematically review those apps flagged by the automated system, however. Instead, the system could be used to assign scores to apps and help regulators focus on the seemingly most egregious ones. This could result in letters or emails being sent to developers asking for clarification.

"Some discrepancies are to be expected because not all developers are sophisticated about privacy," Sadeh said.

For instance, one common error is to build an app that uses Google Maps, but fail to mention the processing of location information in the related privacy policy.

"Whenever you're using Google Maps," he noted, "you're effectively sharing personal information with Google."

Provided by Carnegie Mellon University