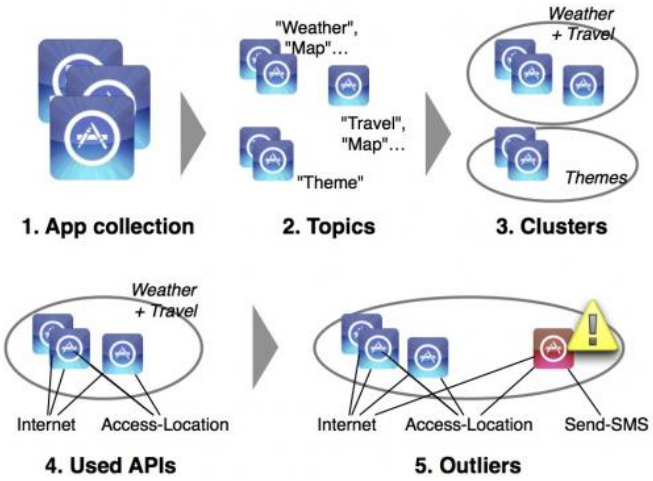
Wisdom of app stores: Early identification of malicious Android apps from Google Play

18 February 2014



Starting from a collection of "good" apps (1), Chabada identifies their description topics (2) to form clusters of related apps (3). For each cluster, it identifies the data and services used, grouped by related permissions (4), and can then identify outliers using resources uncommon for that cluster (5). Credit: Andreas Zeller

Apps on mobile phones can be dangerous data thieves. To expose those spies in your pocket, computer scientists from Saarland University present a new method to analyze apps en masse. To do this, they compare the actual behavior of the respective app with its published functional description. The researchers from Saarbruecken applied their approach to over 22,000 apps on the Google Play platform, detecting several malware instances. The American search engine corporation has already shown interest.

"How do I know that the new installed app behaves as described?" asks Andreas Zeller, professor of software engineering at Saarland University. So far experts have identified so-called malicious apps by checking their behavior against patterns of known attacks. "But what if the attack is brand-new?" asks Zeller.

His group seems to have found a new method to answer all these questions. Zeller summarizes the basic idea as follows: "Apps whose functionality is described in the app store should behave accordingly. If that is not the case, they are suspect."

His research group has named the software based on this idea "Chabada". For every app, it analyzes the description of its functionality that can be read in the app store. With methods from natural language processing, it identifies the main topics, for example "music". After that, Chabada clusters applications by related topics. For instance, the cluster "travel" consists of all apps that deal with traveling in some way. Using program analysis, Chabada detects which data and services are accessed by the apps. Travel apps normally access the current location and a server to load a map. So a travel app secretly sending text messages is suspicious.

The researchers applied this approach on 22,521 apps from the Google Play Store. With a purpose-built script, they had downloaded the 150 most popular apps in the 30 categories from Google Play during spring and winter of last year. Chabada then analyzed them. Finally, the computer scientists from Saarbruecken investigated the 160 most significant outliers to verify Chabada's selection. The result: It had detected 56 percent of the existing spy apps, without knowing their behavior patterns beforehand.

How important the researchers' efforts are is shown by a news item published by the Russian software company "Doctor Web" at the end of June last year. It reported that the company had discovered various malicious apps on the "Google Play" platform. Downloaded onto a smartphone, the malware installed other programs, which secretly sent text messages to expensive premium services. Although Doctor Web, according to its own statement, informed Google immediately, the

malicious apps were still available for download for several days. Doctor Web estimates that in this way up to 25,000 smartphones were used fraudulently. "In the future Chabada could serve as a kind of gatekeeper, ensuring that malicious apps will never make it into an app store", Zeller explains.

The computer scientists from Saarbruecken will present their new approach at the International Conference on Software Engineering (ICSE) in Hyderabad, India at the end of May. Already in March, Google security researchers will be meeting with the Saarbruecken team. Google has also already invited Zeller and his colleagues to have Chabada analyze the whole Google App Store.

Provided by Saarland University