Automated Analysis of Privacy Requirements for Mobile Apps
Sebastian Zimmek, Zi Qi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven M. Bellovin, Joel Reidenberg

Background

We introduce a system to analyze Android apps’ compliance with privacy requirements.

- We define privacy requirement compliance to mean that apps need a privacy policy and must behave according to it.
- In addition, the policy by itself is required to follow requirements (e.g., on notifying a user on access, edit, and deletion rights).

In detail, apps that process Personally Identifiable Information (PII) are generally required to:

1. have a privacy policy (either on its Google Play page or inside the app);
2. include notices about policy changes;
3. notify users of data collection practices;
4. disclose how data is shared with third parties.

Policy Analysis

71% (6,198/8,696) of apps appear to have no privacy policy despite processing PII.

The system classifies descriptions of practices in privacy policies based on machine learning.

- Keyword sets are used to identify practices: data type keywords and action keywords.
- Sentences in policies are extracted based on data type keywords (e.g., all sentences that contain the term “location”).
- Using action keywords unigram and bigram feature vectors are constructed from the extracted sentences (e.g., “share location”).
- The unigram and bigram features are leveraged by Support Vector Machine (SVM) and Logistic Regression (Log. Reg.) classifiers.

Inconsistency (n=9,050)

- 2,455 apps have one potential privacy requirement non-compliance, 2,460 have two, and only 1,461 adhere completely to their policy (out of n = 9,050 apps).
- Each app exhibits a mean of 1.83 instances of potential privacy requirement non-compliance.
- Non-compliance does not necessarily mean that a law is violated.

Classification results for a policy test set (n=40) and the occurrence of positive classifications (Pos) in a set of n=9,050 policies.

Compliance with Privacy Requirements

Identifying privacy requirement non-compliance for a test set of app/policy pairs (n=40) and the percentages of potential non-compliance (Inconsistency) for n=9,050 app/policy pairs.