

Which Apps have Privacy Policies?

Peter Story, Sebastian Zimmeck, and Norman Sadeh
Usable Privacy Policy Project, Carnegie Mellon University

Overview

Which factors are associated with whether smartphone apps link to a privacy policy? We downloaded the metadata of over 1 million apps from the Google Play Store and trained a logistic regression model to predict whether apps have privacy policy links.

Introduction

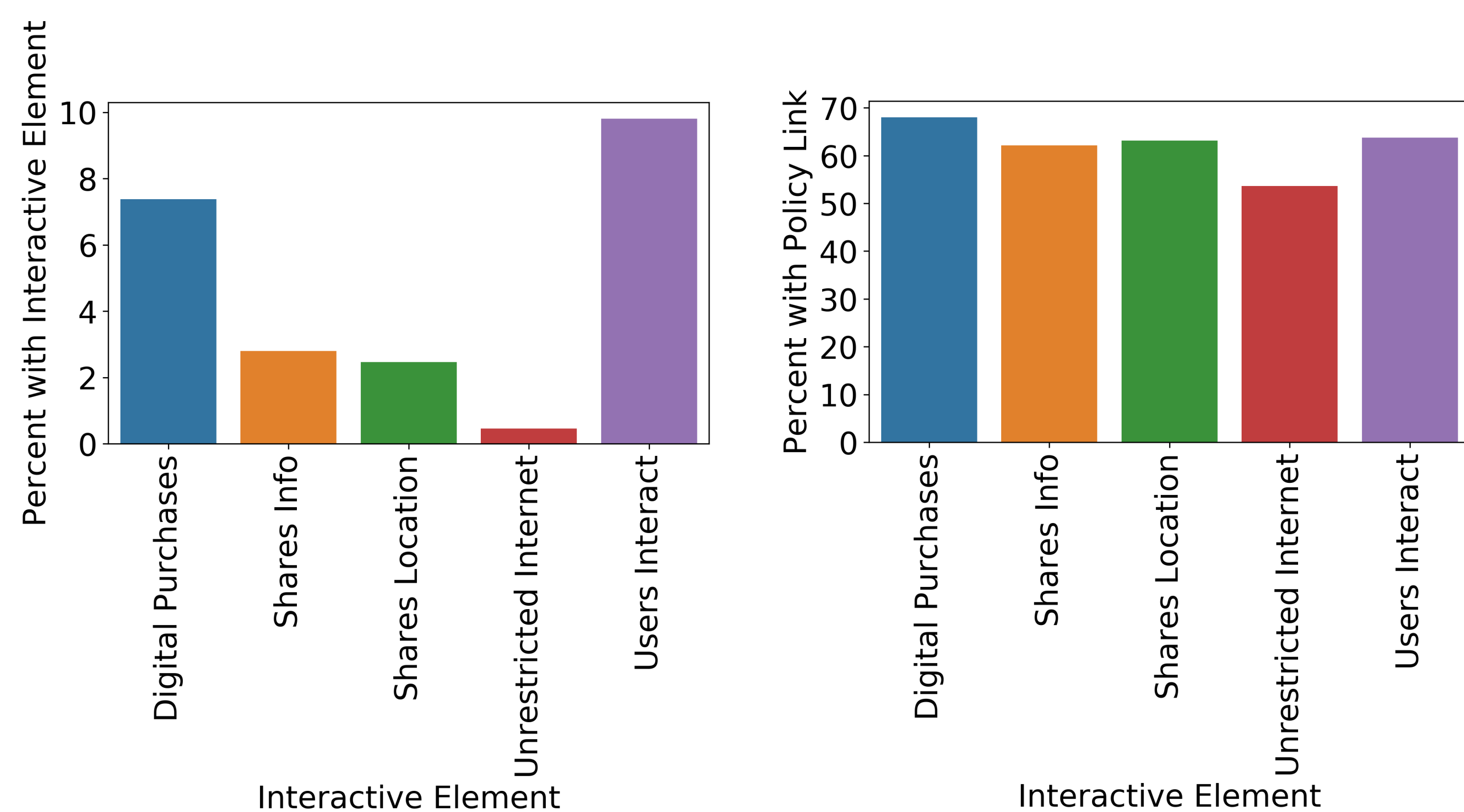
The Google Play Store contains over a million apps. Apps can collect many different types of information from users. Apps can use this information in a variety of ways. One way to learn about what an app does with the information it collects is to read the app developer's privacy policy. However, we found that less than half of apps (45.2%) link to their policy from the Play Store.

Data Collection

We used a recursive crawling technique to download the metadata associated with a large number of apps in the U.S. Play Store. We ran one crawl between August 28th and September 2nd 2017 (n=1,423,450) which found that 41.7% of apps had links to privacy policies. We ran another crawl between November 29th and December 2nd 2017 (n=1,163,622) which found that 45.2% of apps had links to privacy policies on their Play Store pages.

The number of apps decreased between these crawls and the percent of apps with policy links increased. One possible explanation is that these changes were caused by Google removing apps which collected “Personal and Sensitive Information” but did not have links to privacy policies [1].

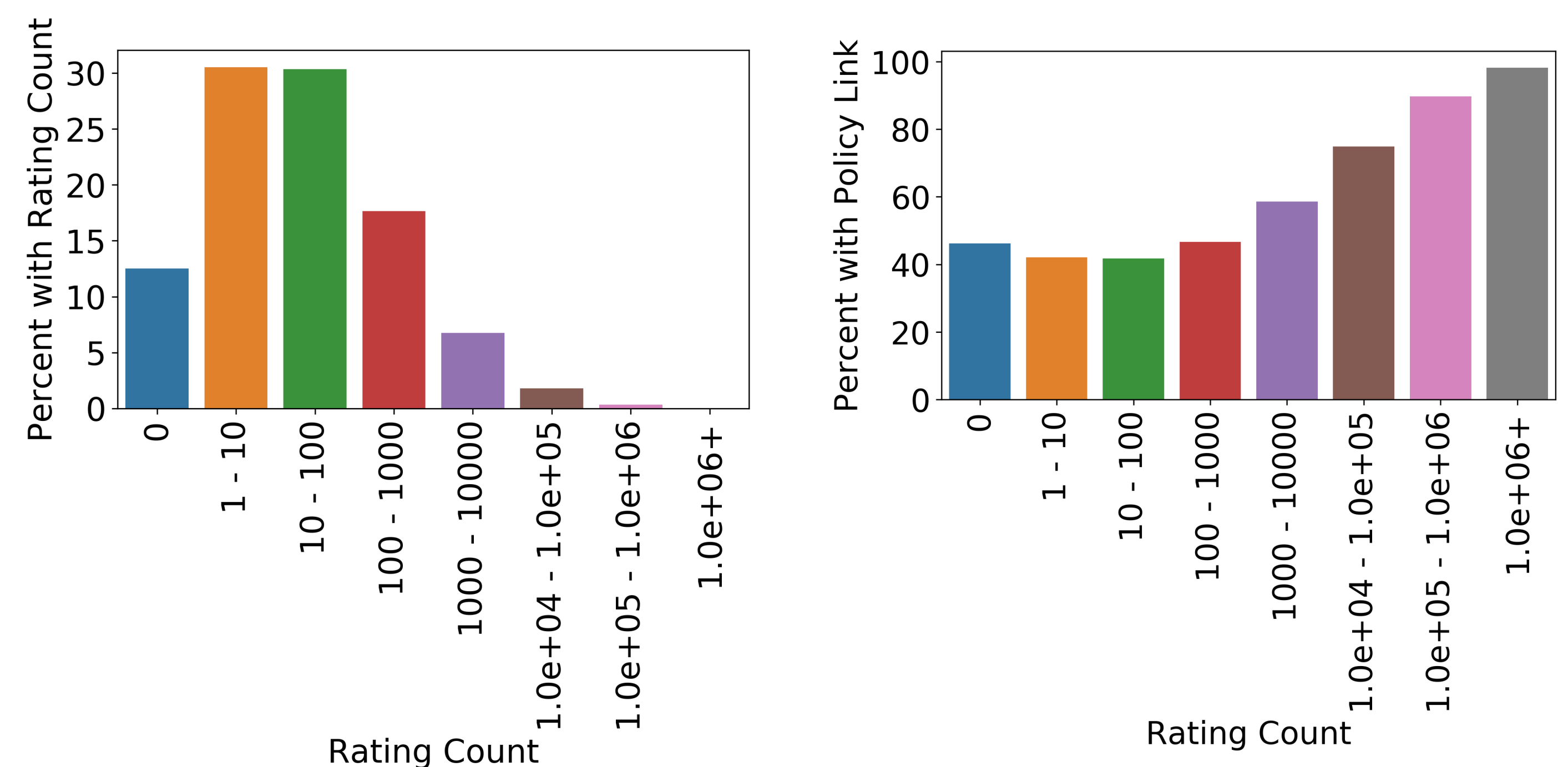
The latest crawl was used to generate our graphs and to train the logistic regression analysis.



Percent of apps with each interactive element (left) and percent of apps with privacy policy links by interactive element (right). Our logistic regression analysis found that all elements except “Unrestricted Internet” were associated with an increased probability of having privacy policy links.

Main Findings

- Less than half (45.2%) of apps have links to privacy policies
- The percent of apps with policy links increased between September and December 2017
- Identified factors which affect the odds of apps having policy links



Percent of apps with different numbers of user ratings (left) and percent of apps with privacy policy links by number of ratings (right). Our logistic regression model found that apps with more ratings were more likely to have links.

Logistic Regression Analysis

We trained a logistic regression model to predict whether an app would have a privacy policy link (67.7% accuracy). Our model contained the following features:

Features	Last Update Date	Install Count Range	Editors' Choice Badge
Play Store Category	Price	Rating Count	Rating Value
ESRB Content Rating	ESRB Content Descriptor	Interactive Elements	Developer's Country

Among other things, the model predicts that:

- Apps with **more ratings** are more likely to have privacy policy links (odds **x2.648** when ratings increased by 100,000)
- Apps with **in-app purchases** are more likely to have privacy policy links (odds **x1.985**)
- Apps in the **“Books and Reference” category** are less likely to have privacy policy links (odds **÷2.105**)
- Apps with the **“Sexual Themes” content descriptor** are less likely to have privacy policy links (odds **÷1.534**)

Conclusion and Future Work

As evidenced by the low percent of apps with privacy policies, there is room for improvement in privacy standards. In extension of our previous research [2], we are working on a system which will compare the actual practices of apps with the practices described in their privacy policies. This system will allow regulators to discover potential compliance issues at a large scale and observe trends over time to improve app privacy.

References

- [1] Bryan Clark. 2017. Millions of apps could soon be purged from Google Play Store. (8 Feb. 2017). <https://thenextweb.com/google/2017/02/08/millions-apps-soon-purged-google-play-store/> Accessed: Dec. 13 2017.
- [2] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shormir Wilson, Norman Sadeh, Steven M. Bellovin, and Joel Reidenberg. 2017. Automated Analysis of Privacy Requirements for Mobile Apps. In 24th Network & Distributed System Security Symposium (NDSS 2017) (NDSS 2017). Internet Society, San Diego, CA.



This project is funded by the National Science Foundation under its Secure and Trustworthy Computing initiative (CNS-1330596, CNS-1330214, and SBE-1513957).