Mobile App Privacy Compliance: Automated Technology to Help Regulators, App Stores and Developers

Sebastian Zimmeck
Carnegie Mellon University
szimmeck@andrew.cmu.edu

Ziqi Wang
Carnegie Mellon University
ziqiw@andrew.cmu.edu

Lieyong Zou
Carnegie Mellon University
lieyongz@andrew.cmu.edu

Roger Iyengar
Carnegie Mellon University
ri@rogeriyengar.com

Bin Liu
Carnegie Mellon University
bliu1@cs.cmu.edu

Florian Schaub
University of Michigan
fschaub@umich.edu

Shomir Wilson
University of Cincinnati
shomir.wilson@uc.edu

Norman Sadeh
Carnegie Mellon University
sadeh@cs.cmu.edu

Steven M. Bellovin
Columbia University
smb@cs.columbia.edu

Joel Reidenberg
Fordham University
jreidenberg@law.fordham.edu

Abstract
Mobile apps have to satisfy various privacy requirements. Notably, app publishers are often obligated to provide a privacy policy and notify users of their apps' privacy practices. But how can a user tell whether an app behaves as its policy promises? In this study we are introducing a scalable system to analyze and predict Android apps' compliance with privacy requirements. We report on our collaboration with three regulatory agencies. We present analysis results for 17,991 apps. We expect to soon be able to support app store-wide analysis (i.e., over a million apps) and to track changes in non-compliant behavior over time. Beyond its use by regulators and activists our technology is also intended to assist app developers and app store owners in their internal assessments of privacy requirement compliance.

Author Keywords
privacy policy analysis; static analysis; mobile app privacy; data leakage; privacy compliance; privacy

ACM Classification Keywords
D.4.6 [Software]: Security and Protection; J.1 [Computer Applications]: Administrative Data Processing—law; K.4.1 [Computing Milieux]: Public Policy Issues—privacy; K.5.2 [Computing Milieux]: Governmental Issues—regulation

Introduction
"We do not ask for, track, or access any location-specific information [...]." This is what Snapchat's privacy policy

stated.[1] However, its Android app transmitted Wi-Fi- and cell-based location data from users' devices to analytics service providers. These discrepancies remained undetected before they eventually surfaced when a researcher examined. Snapchat's data deletion mechanism. His report was picked up by the Electronic Privacy Information Center and brought to the attention of the Federal Trade Commission (FTC), which launched a formal investigation requiring Snapchat to implement a comprehensive privacy program.[2]

The case of Snapchat illustrates that mobile apps can be non-compliant with privacy requirements [3]. However, any inconsistencies may have substantial consequences, particularly, as they can lead to enforcement actions by the FTC and other regulators. This is especially true if discrepancies continue to exist for many years, which was the case for Yelp's collection of childrens' information.[3] These findings not only demonstrate that regulators could benefit from a system that helps them identify potential privacy requirement inconsistencies, but also that it would be a useful tool for companies in the software development process. This would be valuable because researchers found that privacy violations often appear to be based on developers' difficulties in understanding privacy requirements rather than on malicious intentions.

Data Practice Analysis

There is no generally applicable federal statute demanding privacy policies for apps. However, California and Delaware enacted comprehensive online privacy legislation that effectively serves as a national minimum privacy threshold given that app publishers usually do not provide state-specific app versions or exclude California or Delaware residents. In this regard, the California Online Privacy Protection Act of 2003 (CalOPPA) requires online services that collect personally identifiable information (PII) to post a policy. The same is true according to Delaware's Online Privacy and Protection Act (DOPPA). In addition, the FTC's Fair Information Practice Principles (FTC FIPPs) call for consumers to be given notice of an entity's information practices before any PII is collected.

We concentrate our analysis on a subset of data types that are, depending on the context, legally protected: device IDs, location data, and contact information. In particular, it should be noted that in the current version of our system we interpret ad identifiers to be PII since they can be used to track users over time and across devices—our system can easily be adapted to support different interpretations or support different ways of prioritizing the reporting of potential compliance violations. We are also assuming that a user did not opt out of ad targeting (because otherwise no ad identifiers would be sent to opted out ad networks). We further interpret location data to particularly cover GPS, cell tower, and Wi-Fi locations.

We assume applicability of the discussed laws and perform our analysis based on the guidance provided by the FTC and the California Office of the Attorney General (Cal AG) in enforcement actions and recommendations for best practices. Specifically, we interpret the FTC actions as disallowing the omission of data practices in policies and assume that silence on a practice means that it does not occur—again, our system could easily be customized to support different interpretations or requirements coming from other jurisdictions. We assume that all apps in the US Play store are subject to CalOPPA and DOPPA, which we believe to be reasonable as we are not aware of any US app publisher excluding California or Delaware residents or providing state-specific app versions. From the laws and regulations we derive the privacy requirements against which we measure compliance.

[1] Complaint In the Matter of Snapchat, Inc. (Dec. 31, 2014).
[2] Decision and Order In the Matter of Snapchat, Inc. (Dec. 31, 2014).
[3] United States of America v. Yelp, Inc. (Sep. 17, 2014).

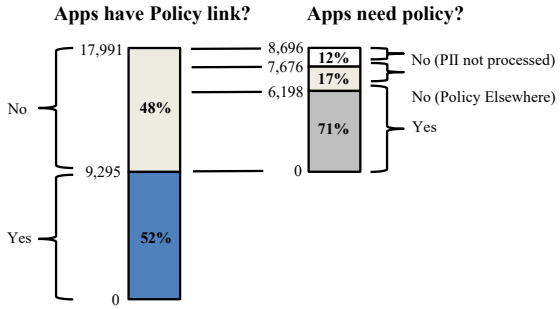

Figure 1: We analyze 17,991 free apps, of which 9,295 (52%) link to their privacy policy from the Play store (left). Out of the remaining apps, 6,198 (71%) appear to lack a policy while engaging in a data practice (i.e., PII is processed) requiring them to have one (right).

Preliminary Results

We found that 9,295 app Play Store pages (out of a total of 17,991) provided a link to their policy and 8,696 lacked such. As shown in Figure 1, our results suggest that 71% (6,198/8,696) apps without a policy link should have had a policy. We used the Play store privacy policy links as proxies for actual policies, which we find reasonable since regulators requested app publishers to post such links [2], and app store owners obligated themselves to provide the necessary functionality [1]. The apps in the full app set were offered by a total of 10,989 publishers, and their app store pages linked to 6,479 unique privacy policies.

We arrive at 71% after making two adjustments. First, if an app does not have a policy it is not necessarily noncompliant with the policy requirement. After all, apps that are not processing PII are not obligated to have a policy. Indeed, we found that 12% (1,020/8,696) of apps without a policy link are not processing PII and, thus, accounted for those apps. Second, despite the regulators' requests to post policy links in the Play store, some app publishers may still decide to post their policy elsewhere (e.g., inside their app). Thus, we randomly selected 40 apps from our full app set that did not have a policy link in the Play store but processed PII. We found that 83% (33/40) do not seem to have a policy posted anywhere. Accounting for an additional 17% (1,478/8,696) of apps having a policy elsewhere leads to 100% - 12% - 17% = 71% out of 8,696 apps potentially non-compliant with the policy requirement.

To evaluate our system we use a test set with 40 corresponding app/policy pairs. We manually evaluated each policy and app in the test set. As shown in Table 1, for example, for collection of identifiers our system was correct in 38 of the 40 app/policy pairs resulting in an accuracy of 38/40 = 0.95. The rightmost column of the table shows the results for 9,050 app/policy pairs.[4] For example, 50% of policies either did not mention collection of identifiers or explicitly stated that no identifiers are collected while the app actually seems to collect at least one identifier.

Our results suggest that there is potential non-compliance with privacy requirements on a broad level. Depending on the practice 2% to 63% of apps do not seem to have sufficient disclosures in their privacy policies of what happens to usersâĂŹ data. The relatively low percentages of 9% for collection of contact information and 2% for sharing of such could be a result that this type of information is often obtained directly from the user, for example, by letting the user enter his or her e-mail address into a form field. As our system does not detect those practices the real percentages might be higher. It should be also noted that the results strongly depend on the legal interpretation of privacy

[4] While we found that 9,295 apps have a policy, we only analyze 9,050 app/policy pairs to account for some policy links not actually leading to a privacy policy (e.g., due to a 404 error).

Practice	*Acc* *(n=40)*	*95% CI* *(n=40)*	*Prec*$_{pos}$ *(n=40)*	*Rec*$_{pos}$ *(n=40)*	*F-1*$_{pos}$ *(n=40)*	*F-1*$_{neg}$ *(n=40)*	*TP, FP, TN, FN* *(n=40)*	*Inconsistent* *(n=9,050)*
Collection Identifier	**0.95**	0.83–0.99	**0.75**	1	**0.86**	0.97	6, 2, 32, 0	**50%**
Collection Location	**0.83**	0.67–0.93	**0.54**	1	**0.7**	0.88	8, 7, 25, 0	**41%**
Collection Contact	**1**	0.91–1	-	-	-	1	0, 0, 40, 0	**9%**
Sharing Identifier	**0.85**	0.7–0.94	0.93	0.74	**0.82**	0.87	14, 1, 20, 5	**63%**
Sharing Location	**1**	0.91–1	1	1	**1**	1	3, 0, 37, 0	**17%**
Sharing Contact	**1**	0.91–1	1	1	**1**	1	1, 0, 39, 0	**2%**

Table 1: Identifying potential privacy requirement inconsistencies in the app/policy test set (n=40) and estimate for all 9,050 app/policy pairs.

policies, particularly, on whether not mentioning a practice means that it is permitted in the app or not.

Conclusion and next Steps

Our results suggest broad potential privacy requirement inconsistencies. Thus, we will extend our analysis and improve its accuracy. The privacy policy analysis can be further developed to capture nuances in policy wording. Similarly, the accuracy of the app analysis could be enhanced by taint flow analysis techniques. These improvements have to be balanced with our system's performance to enable the analysis of apps at app store scale.

Regulators are pushing for early enforcement of potentially non-compliant privacy practices. Automated approaches can relieve them from substantial parts of their workload allowing them to move towards systematic oversight. As many software publishers do not intend noncompliance but rather lose track of their obligations or are unaware of them, we also see potential for implementing privacy requirement analyses in software development tools and integrating them into the app vetting process in app stores.

Acknowledgments

This study is supported in part by the NSF under grants CNS-1330596, CNS-1330214, and SBE-1513957, as well as by DARPA and the Air Force Research Laboratory, under agreement number FA8750-15-2-0277. The US Government is authorized to reproduce and distribute reprints for Governmental purposes not withstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA, the Air Force Research Laboratory, the NSF, or the US Government.

REFERENCES

1. California Department of Justice. 2012. Attorney General Kamala D. Harris Secures Global Agreement to Strengthen Privacy Protections for Users of Mobile Applications. (Feb. 2012).
2. FTC. 2013. Mobile Privacy Disclosures. (Feb. 2013).
3. 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 '17) (NDSS '17)*. Internet Society, San Diego, CA.