On The Ridiculousness of Notice and Consent: Contradictions in App Privacy Policies

Ehimare Okoyomon
UC Berkeley
eokoyomon@berkeley.edu

Nikita Samarin
UC Berkeley
nsamarin@berkeley.edu

Primal Wijesekera
UC Berkeley / ICSI
primal@berkeley.edu

Amit Elazari Bar On
UC Berkeley
amit.elazari@berkeley.edu

Narseo Vallina-Rodriguez
IMDEA Networks Institute / ICSI
narseo@icsi.berkeley.edu

Irwin Reyes
ICSI
ioreyes@icsi.berkeley.edu

Álvaro Feal
IMDEA Networks Institute
alvaro.feal@imdea.org

Serge Egelman
ICSI / UC Berkeley
egelman@cs.berkeley.edu

Abstract

The dominant privacy framework of the information age – "notice and consent" – relies on service providers disclosing their data collection practices through privacy policies, and users consenting to their terms. Through analysis of 68,051 apps from the Google Play Store, their corresponding privacy policies, and observed data transmissions, we investigate the potential misrepresentations of apps in the Designed For Families (DFF) program, inconsistencies in disclosures regarding third-party data sharing, as well as contradictory disclosures about secure data transmissions. We find that of the 8,030 DFF apps (i.e., apps directed at children), 9.1% claim that their apps are not directed at children, while 30.6% claim to have no knowledge that the received data comes from children. In addition, we observe that 22,856 apps do not mention any third-party affiliates in their privacy policies, yet 7,147 still share user data, and only 22.2% of all apps explicitly name third parties. Furthermore, we find that 9,424 apps do not use TLS when transmitting personal identifiers, yet 28.4% of these apps claim to take measures to secure data transfer. Ultimately, these divergences between disclosures and actual app behaviors illustrate the ridiculousness of the notice and consent framework.

Introduction

Our work aims to demonstrate the inadequacy of privacy policies as a mechanism of notice and consent, focusing on Android smartphone applications ('apps'). Literature has shown questionable privacy behaviors and collection practices across the mobile app ecosystem [9]. This paper explores whether such specific questionable collection practices are represented in the privacy policies and disclosed to users. While past work has focused separately on app behavior analysis at practice [9, 2, 4, 5, 7, 12, 10, 8] or analysis of privacy policies [3, 14, 15, 6, 13], we aim to bridge this gap by considering these two problems in tandem. In other words, we compliment the dynamic analysis results, focusing on what is collected and with whom it is shared, with an analysis of whether users were adequately informed about such collection.

In this work we focus on three classes of discrepancies between collection at practice (de facto) and as per the online service's notice (de jure).

Children's Privacy. We examine mobile apps that participate in the Google Play Store's 'Designed for Families' (DFF) program and regulated under the Children Online Privacy Protection Act (COPPA), meaning their target audience includes children under the age of 13 [11]. We find that a substantial number of apps *targeted at children* include clauses in their privacy policy either claiming to not have knowledge of children in their audience, or outright prohibitions against the use of their apps by children.

Third-Party Data Sharing. The second aspect we are interested in analyzing is the disclosure of third-party services that receive and process user information. Regulations like GDPR (Article 13 1.e) and CCPA require developers to explicitly notify users about the recipients of information, either their names or categories. We explore how many app developers include information about their third-party affiliates in the privacy policy and how many of them explicitly name them.

Transit Encryption. Third, privacy policies often represent to users they implement reasonable security measures. At a minimum, one such measure should include TLS encryption. Protecting users' data using reasonable security measures is a regulatory requirement under COPPA, CCPA, and GDPR (Article 32). We explore how many apps potentially fail to adhere to their own represented policies, by transmitting data without using TLS.

Dataset and Methodology

In our work, we rely on the AppCensus dataset available at [1]. AppCensus is a tool that analyzes Android apps from Google Play Store in order to identify the personal information that apps access and share with other parties over the Internet. It leverages dynamic analysis techniques to automatically analyze an application's runtime and network behavior. AppCensus also fetches and stores privacy notices of each analyzed app, which we use to identify possible mismatches between the stated and actual app behavior. As of January 2019, it included information about 68,051 apps published on Google Play Store.

Policy Analysis

In our project, we focus on three types of misrepresentations that occur in privacy notices of mobile apps. Table 1 shows the total number of apps from the AppCensus dataset that we examine and the number of observations that we obtain for different types of analysis. For misrepresentations concerning children's privacy, we analyze 8,030 apps participating in Google's DFF program out of all 68,051 available apps. For third-party sharing practices and for TLS usage, we use the entire dataset of 68,051 apps.

Table 1: Number of observed apps for different types of analysis.

Description	Observed App #	Sample Size
Participate in DFF program	8,030	68,051
Claim not to target children	728	8,030
Claim no knowledge of children data	2,457	8,030
Mention third parties	45,195	68,051
Provide names of third parties	15,106	45,195
Undisclosed sharing (third parties not mentioned)	7,147	22,856
Transmit personal data	36,107	68,051
No TLS usage during transmission	9,424	36,107
Claim to secure data transmission	2,680	9,424

We analyze the text of privacy policies to identify potential misrepresentations. To verify compliance with COPPA, we first narrow our search to only include apps in Google's DFF category with any combination of the keywords "child", "kid", "COPPA", and "minor." Next, we manually read and process the policies for a subset of 200 DFF apps, focusing primarily on these keywords and frequently-used phrases and expressions.

We are further interested in exploring how many app developers disclose their information sharing practices. We look at all 68,051 available apps, aiming to collect the relevant clauses on information sharing with third-party services from their privacy policies and to determine whether the names (as opposed to categories) of those third-party recipients of information are disclosed.

First, we analyze the texts of privacy policies using regular expressions. In particular, we are interested to see whether any part of the text matches the phrase "third parties" or any variation thereof (e.g. "affiliate" or "partner" instead of "third party"). Focusing on matched privacy policies, we determine whether any third-party service providers are explicitly named. We use a list of 9,672 domains that receive data from mobile apps, including known analytics and advertising networks, which we obtain from the AppCensus dataset.

Finally, we want to ensure that app developers comply with their own policies whenever they promise to take reasonable steps to secure user data from unauthorized access. We first identify mobile apps that transmit personal information over the Internet without using TLS using the AppCensus dataset. We then analyze their privacy policies, identifying parts of the text that mention personal data. This is again done using regular expressions, matching "personal information", "personally identifiable information" and variants thereof. Finally, sentences containing information about personal data are scanned for specific key phrases (e.g. "security measures", "unauthorized disclosure", "reasonable steps to secure", "transmission", etc.), that provide security guarantees concerning data transmission.

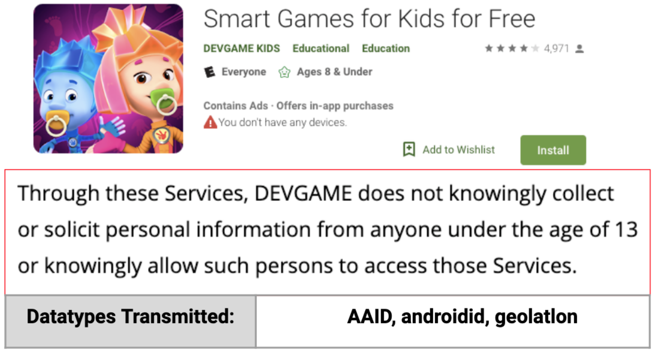
Results

We report our analysis along three aforementioned dimensions: Children's Privacy, Third-party Service Providers, and Secure Data Transmission.

Children's Privacy

For the Children's Data Privacy analysis, we looked at 8,030 apps in the Designed For Families program. Out of these apps, we found that there are 728 apps (9.1%) that claim they are not targeted at children and 2,457 (30.6%) that claim no knowledge of collecting any data from children under 13, with some overlap in apps that do both. In fact, only 4,649 (57.9%) mention any combination of the keywords "child", "kid", "coppa", and "minor".

For instance, “Smart Games for Kids for Free” made by the developer DEVGAME KIDS has a very obvious advertising directed at children as inferred by the application icon and name, in addition to its declaration in the Google Play Store being for Ages 8 & Under. Nevertheless, DEVGAME KIDS' privacy policy claims they do not knowingly allow such persons to access their Services. In addition to this, they claim to not knowingly collect or solicit personal information from children, but we have observed them transmitting the AAID, androidid, and geolatlon datatypes.


Smart Games for Kids for Free
DEVGAME KIDS
Educational
Education
★★★★☆ 4,971
Everyone
Ages 8 & Under
Contains Ads · Offers in-app purchases
You don't have any devices.
Add to Wishlist
Install

Through these Services, DEVGAME does not knowingly collect or solicit personal information from anyone under the age of 13 or knowingly allow such persons to access those Services.

Datatypes Transmitted:	AAID, androidid, geolatlon

Third-party Service Providers

We also identify apps that do not reveal the names of affiliated third parties in their privacy policies. We start by locating apps that mention third-party service providers. From there, we narrow this list only to include apps that explicitly name at least one third-party partner.

In our corpus, 45,195 (66.4%) mention third-party affiliates, which suggest that the remaining 22,856 apps should not transmit any personal data to outside domains. However, out of these 22,856 apps, 7,147 (31.3%) of them still share user identifiers with other service providers without giving notice to the users. In addition, we discover that only 15,106 apps (22.2% of 68,051) explicitly name their third-party affiliates.

Secure Data Transmission

Using the AppCensus dataset, we discover that 36,107 apps that are available on Google Play Store transmit personal data over the network. As of January 2019, 9,424 of these apps (26.1%) do not use TLS when transmitting personal identifiers. Out of those 9,424 apps, 2,680 apps (28.4%) claim to take measures to secure data transmission, but fail to employ TLS when transmitting PII.

Conclusion

This work accentuates the degree in which the privacy framework of notice and consent is flawed by analyzing Google Play Store apps and comparing their privacy policies with their behavior. Our analysis specifically focuses on highlighting the misrepresentation and lack of information that exists in of apps in the Designed for Families program, apps that interact with third parties, as well as apps that claim to utilize secure data transmission precautions, ultimately showing the level of carelessness and lack of priority when it comes to protecting user privacy.

Acknowledgments

This work was supported by the U.S. National Security Agency's Science of Security program (contract H98230-18-D-0006), the Department of Homeland Security (contract FA8750-18-2-0096), the National Science Foundation (grants CNS-1817248 and CNS-1564329), the European Union's Horizon 2020 Innovation Action program (grant Agreement No. 786741, SMOOTH Project), the Rose Foundation, the Data Transparency Lab, and the Center for Long-Term Cybersecurity at U.C. Berkeley.

REFERENCES

1. AppCensus AppSearch. 2019. https://search.appcensus.io/. (2019). Accessed: 2019-03-26.
2. Steven Arzt, Siegfried Rasthofer, Christian Fritz, Eric Bodden, Alexandre Bartel, Jacques Klein, Yves Le Traon, Damien Octeau, and Patrick McDaniel. 2014. Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. *Acm Sigplan Notices* 49, 6 (2014), 259–269.
3. Travis D Breaux and Florian Schaub. 2014. Scaling requirements extraction to the crowd: Experiments with privacy policies. In *Requirements Engineering Conference (RE), 2014 IEEE 22nd International*. IEEE, 163–172.
4. William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N Sheth. 2014. TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)* 32, 2 (2014), 5.
5. Yu Feng, Saswat Anand, Isil Dillig, and Alex Aiken. 2014. Apposcopy: Semantics-based detection of android malware through static analysis. In *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*. ACM, 576–587.
6. Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. 2018. Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. In *27th USENIX Security Symposium (USENIX Security 18)*. USENIX Association, Baltimore, MD, 531–548. https://www.usenix.org/conference/usenixsecurity18/presentation/harkous
7. Elleen Pan, Jingjing Ren, Martina Lindorfer, Christo Wilson, and David Choffnes. 2018. Panoptispy: Characterizing Audio and Video Exfiltration from Android Applications. *Proceedings on Privacy Enhancing Technologies* 2018, 4 (2018), 33–50.
8. Abbas Razaghpanah, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Christian Kreibich, Phillipa Gill, Mark Allman, and Vern Paxson. 2015. Haystack: In situ mobile traffic analysis in user space. *ArXiv e-prints* (2015).
9. Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. 2018. "Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale. *Proceedings on Privacy Enhancing Technologies* 2018, 3 (2018), 63–83.

10. Anastasia Shuba, Anh Le, Minas Gjoka, Janus Varmarken, Simon Langhoff, and Athina Markopoulou. 2015. Antmonitor: Network traffic monitoring and real-time prevention of privacy leaks in mobile devices. In *Proceedings of the 2015 Workshop on Wireless of the Students, by the Students, & for the Students*. ACM, 25–27.

11. U.S. Federal Trade Commission. 2015. Complying with COPPA: Frequently Asked Questions. (2015). `https://www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequently-asked-questions`

12. Narseo Vallina-Rodriguez, Jay Shah, Alessandro Finamore, Yan Grunenberger, Konstantina Papagiannaki, Hamed Haddadi, and Jon Crowcroft. 2012. Breaking for commercials: characterizing mobile advertising. In *Proceedings of the 2012 Internet Measurement Conference*. ACM, 343–356.

13. Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N Cameron Russell, and others. 2016a. The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 1330–1340.

14. Shomir Wilson, Florian Schaub, Rohan Ramanath, Norman Sadeh, Fei Liu, Noah A Smith, and Frederick Liu. 2016b. Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really Work?. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 133–143.

15. Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir 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*.