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

Ehimare Okoyomon[1], Nikita Samarin[1], Primal Wijesekera[1,2], Amit Elazari Bar On[1],
Narseo Vallina-Rodriguez[2,3], Irwin Reyes[2], Álvaro Feal[3,4], Serge Egelman[1,2]
[1]University of California, Berkeley, [2]International Computer Science Institute, [3]IMDEA Networks Institute
[4]Universidad Carlos III de Madrid

***Abstract*—The dominant privacy framework of the information age relies on notions of "notice and consent." That is, service providers will disclose, often through privacy policies, their data collection practices, and users can then consent to their terms. However, it is unlikely that most users comprehend these disclosures, which is due in no small part to ambiguous, deceptive, and misleading statements. By comparing actual collection and sharing practices to disclosures in privacy policies, we demonstrate the scope of the problem.**

Through analysis of 68,051 apps from the Google Play Store, their corresponding privacy policies, and observed data transmissions, we investigated 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 10.5% of 68,051 apps share personal identifiers with third-party service providers, yet do not declare any in their privacy policies, and only 22.2% of the apps explicitly name third parties. This ultimately makes it not only difficult, but in most cases impossible, for users to establish where their personal data is being processed. 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.

I. Introduction

Data protection and privacy regulations are largely informed by the Fair Information Practice Principles (FIPPs) – a set of practices governing the collection and usage of personal information by different entities [21]. Central to FIPPs, and privacy regulations more generally, are the principles of *transparency* and *choice*, which are often presented as "notice and consent," or informed consent. In the context of online privacy, service providers (such as websites or mobile applications) are often required to disclose their information collection practices to users and obtain their consent before collecting and sharing personal information. The most common mechanism of achieving this is by having the users consent to a privacy policy presented by the service provider.

Literature has demonstrated limitations of "notice and consent" [5], [8], [9], [24], [38] and recent regulations, such as the EU General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA), further require online services to provide users comprehensive information on data collection and sharing practices. Such information includes the type of personal information collected and shared, the purpose of the collection (in some cases), and the category of the recipient of the information [1], [14]. Such notice requirements play an important role in the mobile app ecosystem, where commonly used operating system permissions may inform users about potential data collection, but do not provide any insight as to who is the recipient, and for what purpose the information is collected.

Yet absent stringent enforcement actions, online services often draft privacy policies vaguely and obscurely, rendering the informed consent requirement ineffective. Moreover, scholarship has demonstrated privacy notices are often written at a college reading level, making them less comprehensible to average users [2], [17], [19], [35]. Even if privacy policies were more comprehensible, prior work suggested that users would still need to spend over 200 hours per year on average reading every privacy policy associated with every website they visit [25]. It comes as no surprise, therefore, that in reality, very few users choose to read privacy notices in the first place [26], [28], [31].

Our work aims to further demonstrate the inadequacy of privacy policies as a mechanism of notice and informed consent, focusing on Android smartphone applications ('apps'). Literature has shown questionable privacy behaviors and collection practices across the mobile app ecosystem [34]. 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 [4], [13], [15], [29], [33], [34], [36], [43] or analysis of privacy policies [7], [18], [44], [45], [48], 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 paper we focus on three classes of discrepancies between collection at practice (de facto) and as per the online service's notice (de jure):

- First, we examine mobile apps that participate in 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 [40]. 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.

- 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 explored how many app developers include information about their third-party affiliates in the privacy policy and how many of them explicitly name them.
- 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 explored how many apps potentially fail to adhere to their own represented policies, by transmitting data without using TLS.

II. Background and Related Work

This paper analyzes the transparency of mobile applications published in Google Play's Designed for Families Program. Particularly, we focus on identifying potential inconsistencies between DFF applications' runtime behavior and the information publicly disclosed in their privacy policies. We provide an overview of Google Play Store's DFF program and discuss below the most related work in the area of privacy violations in children apps, third-party libraries and automatic analysis of privacy policies.

Designed For Families (DFF): The DFF program is a platform for developers to present age appropriate content for the whole family, specifically users below the age of 13. By going through an optional review process, applications submitted to this program declare themselves to be appropriate for children and are able to list themselves under such family-friendly categories. Although the primary target audience of these apps may not be children, according to [40], "if your service targets children as one of its audiences - even if children are not the primary audience - then your service is 'directed to children'". As a result, DFF applications are all directed at children and thus required to be "compliant with COPPA (Children's Online Privacy Protection Rule), the EU General Data Protection Regulation 2016/679, and other relevant statutes, including any APIs that your app uses to provide the service." [16].

Privacy Violations in Children's Apps: There have been some efforts by the research community to investigate children's privacy in social media [23], smart toys [42], [46] and child-oriented mobile apps [34]. Furthermore, researchers have studied how well developers that target children as their audience comply with specific legislation such as COPPA [34] and the FTC has already established cases against numerous app developers gathering children's data [39] [41] [10].

The Mobile Advertising and Tracking Ecosystem: It is well known that mobile apps rely on third parties to provide services such as crash reports, usage statistics or targeted advertisement [32]. In order to study such ecosystem, previous work has relied on analysis of mobile traffic from several vantage points such as ISP logs, heavily instrumented phones or devices that use VPN capabilities to observe traffic [33], [36], [43]. Another way of studying the presence of such libraries is by performing dynamic analysis (i.e., studying runtime behavior of the app) or static analysis (i.e., studying application code to infer actual behavior). Both these approaches have long been used to analyze privacy in mobile apps [4], [13], [15], [29].

Automatic Analysis of Privacy Policies: The research community has previously studied publicly available privacy policies proving that they are usually written in a *"legal language"* that makes it difficult for average users to understand them [20], [30], [37]. Researchers have made positive efforts to extract relevant information and features from privacy policies using crowd-sourcing campaigns [7], [45], Natural Language Processing (NLP) [48] and deep-learning [18], [44] techniques to present this information in more accessible ways to users [6], [11], [18], [22], [27], [47]. Despite these positive efforts, privacy policies remain a confusing source of information for average users and these documents are even unavailable in many webpages and mobile apps [12], [18], [48].

Polisis: One such service that enables automatic analysis of privacy policies is Polisis [18]. Polisis enables users to submit websites for review, which it then crawls to find the associated privacy policies. It then analyzes the texts of these policies and produces a summary of data collection practices of the submitted websites. We initially considered using Polisis to parse the privacy policies and extract information revelant to our current work. However, we encountered an issue with rate-limiting as we created too many requests and were temporarily banned from using the service. Moreover, we discovered that the performance of Polisis on establishing the sections related to children's privacy was suboptimal. Out of a random subset of 100 policies that we categorized as claiming not being directed at children, Polisis was able to identify only 22% of policies. Therefore, due to the overhead in the data generation using Polisis as well as poor performance on tasks revelant to our work, we decided to perform our own policy analysis.

III. Dataset and Methodology

In this section, we describe the datasets and methods that we use to establish the misrepresentations occurring in privacy policies. We use a publicly available app analysis dataset [3] to evaluate the types of data that mobile apps access and share with other parties over the Internet. We then explain how we leverage this dataset to extract meaning and trends from the apps' behaviors and privacy declarations.

TABLE I
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
No TLS usage	9,424	68,051
Claim to secure data transmission	2,680	9,424

A. *AppCensus Dataset*

In our work, we rely on the AppCensus dataset available at [3]. 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 runs in a highly instrumented Android operating system that is able to detect the personal information linked to the test phone. Then, the tool runs each app in the test environment and inspects the network communications using a monitoring tool that records the source, destination and content of network flows. Thus, we have access to the network flows of every app and we can find whether it shares personal data over the network and with whom. 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.

AppCensus stores analysis results and other metadata in a large-scale dataset that contains detailed information about the runtime behavior and traffic patterns of Android apps. As of January 2019, it includes information about 68,051 apps published on Google Play Store. We compare the information provided in the privacy policies of these apps with their actual data sharing practices, as described in the next subsection.

B. *Policy Analysis*

In our project, we focus on three types of misrepresentations that occur in privacy notices of mobile apps. Table I 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.

We analyze the text of privacy policies to identify potential misrepresentations. To verify compliance with COPPA legislation, 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. This allows us to identify commonalities and phrases that could generally be categorized as:

1) Those that make no distinction between children under 13 and other users
2) Those that clearly indicate that the application either
   a) is not targeted to children; or
   b) does not knowingly collect personally identifiable information from children under 13
3) Other (needs further processing)

Enumerating these phrases into categories allows us to create complex strings representative of these groups and further categorize other policies, most commonly those of type 2 and 3, by searching for other policies with matching substrings. This approach enables us to classify over 400 more policies. From the remaining list, we are able to loosen the specificity of the search strings by simply searching for word combinations such as "not knowingly", "not targeted to/at", "do not address" and their variations to categorize policies containing these phrases in lists. We further process these initial lists to remove irrelevant policies if such phrases were not made in the context of children's privacy. Fortunately, this method produces quite a large throughput of roughly 700 more apps. The remaining policies can be read manually and largely represent policies in category 1. Furthermore, after obtaining and classifying the policy data, we look to draw conclusions directly from the classifications and text, as well as by cross-checking policy declarations with observed network transmissions collected from the respective applications.

We are further interested to explore 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"). While this approach is less sophisticated than techniques used by other automatic policy analyzers, we believe that it is sufficient to identify clauses that contain any information about third parties. Focusing on the sentences matched by our regular expression, we aim to identify which, if any, affiliates are explicitly named by the app developer. This requires solving the *named-entity recognition* (NER) problem and while state-of-the-art NER systems produce near human-level performance, we discovered that they are ill-suited for reliably recognizing the names of analytics and advertising network companies.

Instead, we use the app analysis dataset (68,051) to determine all domains that receive data from mobile apps. This produces a list of 9,672 domains, including known analytics and advertising networks, such as `crashlytics.com`, `vungle.com`, `flurry.com`, `mopub.com` among many others. By separating the domain names and performing a manual review, we end up with a list of 7,826 domains. The entries in this list are matched against the text in the privacy policies to determine the third-party service providers named by the app developer.

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. 'Reasonable security measures' is a broad concept and includes different techniques, such as data encryption, regular security patches, access control, among many others. We focus on secure data sharing, which we believe belongs to 'reasonable security measures', as we have access to data transmission information from the AppCensus dataset.

In order to achieve this, we first identify mobile apps that transmit personal information over the Internet without using TLS. 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.

IV. Results

We report our analysis along three aforementioned dimensions.

A. 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". Within this group there are even applications such as "Cinderella Classic Tale Lite" and "Dino Puzzle - free Jigsaw puzzle game for Kids" that make no commentary on children's use directly in their privacy policies but instead simply contain one or more of these keywords either in the name of the app or in a header on the website. Thus, it is interesting to note the sheer number of apps designed for children that do not even mention kids or children in their policies.

From these 2,457 apps in the DFF that claim no knowledge of collecting from children, we observed 68,594 network transmissions. Out of 2,457, 1,728 (70.3%) of them transfer data types such as android advertising ID (AAID), IMEI, geolocation and WiFi mac address. Since these apps are present in the DFF program, they are all liable under COPPA and thus are responsible for catering to an audience that includes minors. However, it is both confusing and troubling to see their policies contradict what they have acknowledged under DFF.

B. 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, 7147 (10.5% of 68,051) apps still share user identifiers with other service providers without giving notice to the users. We discover that only 15,106 apps (22.2% of 68,051) explicitly name third-party affiliates. For instance, the game development company Kwalee describes its analytics service providers in the following way:

> *We use third party providers Fabric (for crash analytics), AppsFlyer and Tenjin (user attribution and analytics), and SOOMLA (for analytics). Fabric, AppsFlyer and SOOMLA may collect analytics on our behalf and in accordance with our instructions, and their applicable privacy notices.*

C. Secure Data Transmission

As of January 2019, 9,424 of analyzed apps (13.8% of 68,051) that are available on Google Play Store do not use TLS when transmitting personal identifiers over the network. Although this fact is alarming in itself, we also investigate whether the developers of these apps make any deceptive claims in their privacy policies.

Out of those 9,424 apps, 2,680 apps claim to take measures to secure data transmission, but fail to employ TLS when transmitting PII. For instance, the game developer Sino Joy Group describes their procedures in the following way:

> *We also understand that it is important to keep your information safe and secure. . . . We do not believe that there is any transmission method over website or Internet that is completely flawless, even though our commercially reasonable security measures have*

> *been put in place against possible breaches of our sites' security and our user records and databases.*

V. Discussion

We looked at privacy policies for any contradictions from their own behavior and for important information missing in the policies. On a high level, we found that developers a) contradict themselves between what they mention in the policy and what they acknowledge in the Google Play Store, b) are not comprehensive in what they are claiming about their data sharing practices, and c) claim to be secure in communication when they are actually not in the real world.

A. Children's Privacy

The most troublesome finding from looking at the privacy policies of apps under DFF is that their own policies do not acknowledge their audience as targeting kids. We found 9.1% of apps under DFF contain phrases in their policies that indicate their developers do not mean for these products to be used by children.

> *"We do not specifically market to children under 13"*
> *"These Services do not address anyone under the age of 13"*

Although some phrasings are ambiguous about developers' intent to exclude children from their products, we believe confusing language still represents a form of misrepresentation. Our analysis reveals a set of policy misrepresentations as a result of unclear — and at times contradictory — language. Many privacy policies already pose difficulties for consumers to comprehend, but those challenges are further exacerbated when policies send mixed messages. Such an effect runs counter to what privacy policies are supposed to achieve in increasing transparency. For instance, the privacy policy from the developer DEVGAME KIDS claims:

> *Most of DEVGAME's Services are for kids audience and are directed to children under the age of 13. 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.*

A consumer, a concerned parent reading this policy would be left with conflicting impressions about whether DEVGAME intends for children to use their apps. Such a confusing and contradictory policy might indicate the level of importance (or lack thereof) this developer places on user privacy.

A particular set of contradictions occur in the 30.6% of apps under DFF whose policies claim that they do "not knowingly" collect information. This is concerning because either app developers need to be aware of what data their apps collect in order to conform to privacy laws such as COPPA or potentially this is a deceptive practice.

However, barring this, there also exists a variety of other examples of misleading and misrepresenting behaviors of these applications, as viewed from their policies and network transmissions. For example, the privacy policy for the app "Pony Crafting - Unicorn World" claims they do not knowingly collect information from children. However, they later mention that they can collect both personal and non-personal information, and that they do not actually know the source of their data:

> *...this information will be anonymous and we will not be able to tie this information to a specific user, this is a byproduct of an anonymous system not allowing us to distinguish between those over 13 and those under 13*

As a result, this makes one question the legitimacy of their claim of ignorance because they do not have any measures in place to be able to tell. In order to confirm if the transmission of children's information really occurs, we searched the database for network traffic observed from applications whose policies do not knowingly collect from children. In doing this we found 68594 distinct transmissions across the 1728 DFF apps that sent device identifiers across the network, including 1979 transmissions from 10 apps that share the same policy[1]. Since these results all come from automated runs of a UI Monkey that creates pseudo-random swipes and gestures, our testing employed no sophisticated methods at overcoming age-gating, parental consent, or any other barriers. Therefore, we believe this behavior can be akin to that of a child and thus the resulting transmissions would be expected to occur with real users. As a result, we believe the claim of "not knowingly" collecting information is misleading because the applications do very little to verify that collection does not come from children. The majority of the applications also detail that if data from minors is found or reported to them they will delete the information and block the associated account, but we strongly feel this is impractical from a usability standpoint as the majority of users will not actually know what information is being sent in the background and thus cannot make this report.

Furthermore, another transient detail we observed during analysis, as briefly mentioned above in the Methodology, is the presence of repeated sentences across policies. In fact, numerous times many of the *exact same* phrases are used, which becomes particularly obvious when they include the same spelling, grammar, or formatting mistakes such as in:

> *We market to*
> *We do not collect information from children under 13*
> *children under 13.*

Although it makes our analysis more efficient, only highly probably cause for this behavior is that companies and developers are not actually creating their own privacy policies but instead just copy and pasting sections from online. Regardless, this observations shows the level of priority these companies have to make policies usable for end users. As a result, this begs the questions of whether application developers are even aware of the statements in their policies (such as in the DEVGAME example above), as well as if they are even aware of

[1] http://www.vascogames.com/vasco-games-privacy-policy/ as accessed on 2018/11/25

the laws they legally must abide by. They could perhaps be just including the minimal information necessary, as determined by examining another application's policy, in order to escape under the radar. This hypothesis is further emphasized by our findings when we look at the application Cami's Dots that, instead of creating their own privacy policy, decided it was sufficient to provide a link pointing to github.com's privacy policy page[2]. Thus, a potential future work stemming from these findings could be to investigate the percentage of applications that are plagiarised and the degree to which this occurs.

B. Third Parties

Our analysis has demonstrated that mobile apps do not provide sufficient information to users about their third-party partners. Although over 75% of mobile apps make use of third-party services, only around 22% actually disclose the names of those services, while 10% do not mention any information about their affiliates at all. This is concerning, as the users are unable to learn about the information protection practices of the services that are provided access to their personal data. This raises questions about the feasibility of ensuring compliance with data protection legislation, as any potential contradictions within the policy of the app and that of the third party would go unnoticed.

Another common theme is the desire of app developers to relieve themselves of any responsibility for ways in which personal information about the users is handled once it leaves the application. In particular, the following clause was common in the privacy policies of mobile apps:

> *We are not responsible or liable for the privacy policies or practices of third parties.*

However, we believe that app developers should still be accountable for the actions performed by third parties, to which they send users' personal information. It is possible that the desire to avoid liability is what motivates app developers not to provide names of their partners in the first place.

C. Secure Transmission

As mentioned previously, 13.8% of 68,051 mobile apps that we analyzed do not use TLS when transmitting user identifiers and of these 28.4% claim to take reasonable security measures. The reasons for making such claim might include negligence (e.g. not knowing that TLS is not configured, or copy-pasting the policy from another app) or malicious intention (e.g. to create a semblance of security without using proper means of protection). Most of these apps also include a clause explaining how it is impossible to completely ensure the security of data transmission:

> *But remember that no method of transmission over the internet, or method of electronic storage is 100% secure and reliable, and we cannot guarantee its absolute security.*

As in the third party case, this clause also demonstrates the desire for app developers to avoid responsibility for negligent data protection standards. While it is true that it is impossible to be secure against all types of attacks on user data, we believe that using TLS for data transmission is one of the most basic steps that can be adopted by all app developers.

VI. Conclusion

This paper 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 consumer 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] "Assembly Bill No. 375, Chapter 55," *California Legislative Information*, 2018. [Online]. Available: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

[2] A. I. Anton, J. B. Earp, Q. He, W. Stufflebeam, D. Bolchini, and C. Jensen, "Financial privacy policies and the need for standardization," *IEEE Security & privacy*, vol. 2, no. 2, pp. 36–45, 2004.

[3] AppCensus AppSearch, https://search.appcensus.io/, accessed: 2019-03-26.

[4] S. Arzt, S. Rasthofer, C. Fritz, E. Bodden, A. Bartel, J. Klein, Y. Le Traon, D. Octeau, and P. McDaniel, "Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps," *Acm Sigplan Notices*, vol. 49, no. 6, pp. 259–269, 2014.

[5] S. Barocas and H. Nissenbaum, "On notice: The trouble with notice and consent," 2009.

[6] G. Boella, L. Di Caro, M. Graziadei, L. Cupi, C. E. Salaroglio, L. Humphreys, H. Konstantinov, K. Marko, L. Robaldo, C. Ruffini *et al.*, "Linking legal open data: breaking the accessibility and language barrier in european legislation and case law," in *Proceedings of the 15th International Conference on Artificial Intelligence and Law*. ACM, 2015, pp. 171–175.

[7] T. D. Breaux and F. Schaub, "Scaling requirements extraction to the crowd: Experiments with privacy policies," in *Requirements Engineering Conference (RE), 2014 IEEE 22nd International*. IEEE, 2014, pp. 163–172.

[8] F. H. Cate, "The failure of fair information practice principles," 2006.

[9] ——, "The limits of notice and choice," *IEEE Security & Privacy*, vol. 8, no. 2, pp. 59–62, 2010.

[10] U. F. T. Commission, "Two App Developers Settle FTC Charges They Violated Children's Online Privacy Protection Act," https://www.ftc.gov/news-events/press-releases/2015/12/two-app-developers-settle-ftc-charges-they-violated-childrens, 2016, accessed: September 26, 2017.

[2] https://help.github.com/articles/github-privacy-statement/ as accessed on 2018/11/25. Since 2018/12/18, Cami's Dots's policy moved to https://github.com/LTProjects/Cami-s-Dots/compare/master...jigglytep:patch-1

[11] M. Curtotti and E. McCreath, "A right to access implies a right to know: An open online platform for research on the readability of law," *J. Open Access L.*, vol. 1, p. 1, 2013.

[12] M. Degeling, C. Utz, C. Lentzsch, H. Hosseini, F. Schaub, and T. Holz, "We value your privacy... now take some cookies: Measuring the gdpr's impact on web privacy," *arXiv preprint arXiv:1808.05096*, 2018.

[13] W. Enck, P. Gilbert, S. Han, V. Tendulkar, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones," *ACM Transactions on Computer Systems (TOCS)*, vol. 32, no. 2, p. 5, 2014.

[14] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," *Official Journal of the European Union*, vol. L119, pp. 1–88, May 2016. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[15] Y. Feng, S. Anand, I. Dillig, and A. Aiken, "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, 2014, pp. 576–587.

[16] Google, "Developer policy center," [Online; accessed 2019-03-26]. [Online]. Available: https://play.google.com/about/families/designed-for-families/

[17] M. A. Graber, D. M. D'alessandro, and J. Johnson-West, "Reading level of privacy policies on internet health web sites.(brief report)," *Journal of Family Practice*, vol. 51, no. 7, pp. 642–646, 2002.

[18] H. Harkous, K. Fawaz, R. Lebret, F. Schaub, K. G. Shin, and K. Aberer, "Polisis: Automated analysis and presentation of privacy policies using deep learning," in *27th USENIX Security Symposium (USENIX Security 18)*. Baltimore, MD: USENIX Association, 2018, pp. 531–548. [Online]. Available: https://www.usenix.org/conference/usenixsecurity18/presentation/harkous

[19] M. Hochhauser, "Lost in the fine print: Readability of financial privacy notices," *Retrieved November*, vol. 27, p. 2009, 2001.

[20] C. Jensen and C. Potts, "Privacy policies as decision-making tools: an evaluation of online privacy notices," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 2004, pp. 471–478.

[21] M. K. Landesberg, T. M. Levin, C. G. Curtin, and O. Lev, "Privacy online: A report to congress," *NASA*, no. 19990008264, 1998.

[22] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang, "Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 501–510.

[23] M. Madden, A. Lenhart, S. Cortesi, U. Gasser, M. Duggan, A. Smith, and M. Beaton, "Teens, social media, and privacy," *Pew Research Center*, vol. 21, pp. 2–86, 2013.

[24] K. Martin, "Transaction costs, privacy, and trust: The laudable goals and ultimate failure of notice and choice to respect privacy online," 2013.

[25] A. M. McDonald and L. F. Cranor, "The cost of reading privacy policies," *ISJLP*, vol. 4, p. 543, 2008.

[26] G. R. Milne and M. J. Culnan, "Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices," *Journal of interactive marketing*, vol. 18, no. 3, pp. 15–29, 2004.

[27] G. R. Milne, M. J. Culnan, and H. Greene, "A longitudinal assessment of online privacy notice readability," *Journal of Public Policy & Marketing*, vol. 25, no. 2, pp. 238–249, 2006.

[28] J. A. Obar and A. Oeldorf-Hirsch, "The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services," *Information, Communication & Society*, pp. 1–20, 2018.

[29] E. Pan, J. Ren, M. Lindorfer, C. Wilson, and D. Choffnes, "Panoptispy: Characterizing audio and video exfiltration from android applications," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 4, pp. 33–50, 2018.

[30] H. J. Pandit, D. O'Sullivan, and D. Lewis, "Queryable provenance metadata for gdpr compliance," 2018.

[31] I. Pollach, "What's wrong with online privacy policies?" *Communications of the ACM*, vol. 50, no. 9, pp. 103–108, 2007.

[32] A. Razaghpanah, R. Nithyanand, N. Vallina-Rodriguez, S. Sundaresan, M. Allman, C. Kreibich, and P. Gill, "Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem," 2018.

[33] A. Razaghpanah, N. Vallina-Rodriguez, S. Sundaresan, C. Kreibich, P. Gill, M. Allman, and V. Paxson, "Haystack: In situ mobile traffic analysis in user space," *ArXiv e-prints*, 2015.

[34] I. Reyes, P. Wijesekera, J. Reardon, A. E. B. On, A. Razaghpanah, N. Vallina-Rodriguez, and S. Egelman, ""Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 3, pp. 63–83, 2018.

[35] X. Sheng and L. F. Cranor, "An evaluation of the effect of us financial privacy legislation through the analysis of privacy policies," *ISJLP*, vol. 2, p. 943, 2005.

[36] A. Shuba, A. Le, M. Gjoka, J. Varmarken, S. Langhoff, and A. Markopoulou, "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, 2015, pp. 25–27.

[37] R. Slavin, X. Wang, M. B. Hosseini, J. Hester, R. Krishnan, J. Bhatia, T. D. Breaux, and J. Niu, "Toward a framework for detecting privacy policy violations in android application code," in *Proceedings of the 38th International Conference on Software Engineering*. ACM, 2016, pp. 25–36.

[38] R. H. Sloan and R. Warner, "Beyond notice and choice: Privacy, norms, and consent," *J. High Tech. L.*, vol. 14, p. 370, 2014.

[39] U.S. Federal Trade Commission, "FTC Warns Children's App Maker BabyBus About Potential COPPA Violations," 2014. [Online]. Available: https://www.ftc.gov/news-events/press-releases/2014/12/ftc-warns-childrens-app-maker-babybus-about-potential-coppa

[40] ——, "Complying with COPPA: Frequently Asked Questions," 2015. [Online]. Available: https://www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequently-asked-questions

[41] ——, "Mobile Advertising Network InMobi Settles FTC Charges It Tracked Hundreds of Millions of Consumers' Locations Without Permission," 2016. [Online]. Available: https://www.ftc.gov/news-events/press-releases/2016/06/mobile-advertising-network-inmobi-settles-ftc-charges-it-tracked

[42] J. Valente and A. A. Cardenas, "Security & privacy in smart toys," in *Proceedings of the 2017 Workshop on Internet of Things Security and Privacy*. ACM, 2017, pp. 19–24.

[43] N. Vallina-Rodriguez, J. Shah, A. Finamore, Y. Grunenberger, K. Papagiannaki, H. Haddadi, and J. Crowcroft, "Breaking for commercials: characterizing mobile advertising," in *Proceedings of the 2012 Internet Measurement Conference*. ACM, 2012, pp. 343–356.

[44] S. Wilson, F. Schaub, A. A. Dara, F. Liu, S. Cherivirala, P. G. Leon, M. S. Andersen, S. Zimmeck, K. M. Sathyendra, N. C. Russell *et al.*, "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, 2016, pp. 1330–1340.

[45] S. Wilson, F. Schaub, R. Ramanath, N. Sadeh, F. Liu, N. A. Smith, and F. Liu, "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, 2016, pp. 133–143.

[46] B. Yankson, F. Iqbal, and P. C. Hung, "Privacy preservation framework for smart connected toys," in *Computing in Smart Toys*. Springer, 2017, pp. 149–164.

[47] L. Yu, T. Zhang, X. Luo, and L. Xue, "Autoppg: Towards automatic generation of privacy policy for android applications," in *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*. ACM, 2015, pp. 39–50.

[48] S. Zimmeck, Z. Wang, L. Zou, R. Iyengar, B. Liu, F. Schaub, S. Wilson, N. Sadeh, S. M. Bellovin, and J. Reidenberg, "Automated analysis of privacy requirements for mobile apps," in *24th Network & Distributed System Security Symposium (NDSS 2017), NDSS*, 2017.