Testimony of Serge Egelman, Ph.D.
Research Director, Usable Security & Privacy Group
International Computer Science Institute

CTO and Co-Founder
AppCensus, Inc.

United States Senate
Committee on Commerce, Science, and Transportation
Subcommittee on Consumer Protection, Product Safety, and Data Security
Hearing on “Protecting Kids Online: Internet Privacy and Manipulative Marketing”

May 18, 2021
## Contents

1 Introduction and Summary 3

2 Background on Mobile Tracking 4

3 Research Findings 6
   3.1 Collection of Contact and Location Information 6
   3.2 Insecure Transfer of Personal Information 6
   3.3 Targeted Advertising 7
      3.3.1 Ineffective Android Privacy Settings 7
      3.3.2 Ineffective SDK Privacy Settings 8

4 Recommendations for Fixing COPPA 10
   4.1 Moving from “Actual” to “Constructive” Knowledge 10
   4.2 Eliminating the Internal Operations Exemption 11
   4.3 Fixing the Safe Harbor Program 13
   4.4 Increasing Enforcement Efforts 14

5 Conclusion 15

References 16
1 Introduction and Summary

Chairman Blumenthal, Ranking Member Blackburn, and Distinguished Members of the Subcommittee, thank you for the opportunity to testify today about children’s online privacy and the mobile app ecosystem.

My name is Serge Egelman, and I direct the Usable Security and Privacy research group at the International Computer Science Institute, which is a research institute affiliated with the University of California, Berkeley.¹ I hold a PhD from Carnegie Mellon University’s School of Computer Science and a BS in computer engineering from the University of Virginia. I am also the CTO and co-founder of AppCensus, which is a startup that builds tools to analyze the privacy behaviors of mobile apps.² I also consult for state and federal regulators on issues pertaining to online consumer privacy and security.

For the past 17 years, I have been studying consumer privacy preferences, how they make online privacy decisions, and how the online ecosystem can be better designed to both protect consumers and help them make more informed decisions. For the past 10 years, I have studied privacy in the mobile app space, including examining what personal information mobile apps are collecting and sharing, and how that might contrast with consumer expectations, laws, and platform policies. Most relevant to the Subcommittee, two years ago my research group published a study of mobile apps’ compliance with the Children’s Online Privacy Protection Act (COPPA). We used our tools to test 5,855 Android apps that were directed to children and found that more than half appeared to be violating COPPA [3].

My goal through this testimony is to explain how online tracking works, my research on COPPA violations in the mobile app ecosystem, and how the law can be updated to keep pace with rapid technological change to better protect children online. Based on this research, I offer four specific recommendations for improving COPPA:

- **Moving from an “actual” to “constructive” knowledge standard**
- **Eliminating the internal operations exemption**
- **Fixing the Safe Harbor program**
- **Increasing enforcement**

¹ https://www.icsi.berkeley.edu/
² https://www.appcensus.io/
2 Background on Mobile Tracking

To monetize many online services, companies pay those services to show specific advertisements to specific users. They do this by inferring individual users’ preferences based on data automatically collected from them: the services they use, how they use them, from where they use them, and so forth. In short, online and offline activities are tracked, which allows companies to maintain detailed profiles of individual user behavior, which in turn is used to predict users’ interests, preferences, and even demographics. The collected information may be used to predict a consumer’s religion, health conditions, sexual orientation, or political affiliation; some of this information may be revealed by the phone’s GPS location alone, or even by just the name of the app that is being used.

In most cases, this data is used to target advertisements, but in some cases it is sold to data brokers, who use it to augment profiles of the same consumers that they collected from other sources, and then sell it to whoever is willing to pay for it. Obviously, this is even more concerning when the data comes from children, who are unlikely to understand that this is happening, much less consent to it, but who could potentially face enormous impacts due to future usage of this data. This data may be used for manipulative marketing campaigns, but also may feed biased and unaccountable algorithms that use it to make decisions about a child’s future, not to mention outright malicious uses of the data.

Contrary to popular belief, the reason why you receive oddly prescient ads is not because your devices are secretly recording your conversations, but because of this type of inference: your online and offline activities are tracked, and then sophisticated algorithms use that data to make predictions about you. Tracking is made possible by “persistent identifiers.” An identifier is any piece of information that allows an individual—or device—to be uniquely identified. “Persistent” identifiers are identifiers that tend to not change over time. For example, motor vehicles have persistent identifiers in the form of license plates: a license plate uniquely identifies a vehicle and vehicles tend to have the same license plates over time. Thus, if someone records all the license plates at a particular place over time, they can determine how many times in that period any individual vehicle was there. Similarly, if license plates are recorded at many different locations and that data is combined into a single dataset, one could use that to reconstruct the movements of individual vehicles in that dataset. As can be seen, combining a persistent identifier with information about where that identifier was observed allows a data recipient to reconstruct an individual’s activities. Using this knowledge, one could infer information about their routines, preferences, demographics, and even relations and social connections!
While this type of mass surveillance may seem appealing to some for the increased security they believe it may enable, a wealth of scholarship exists to show why this is a false tradeoff (e.g., [4, 5]).

This is precisely how mobile tracking occurs. Mobile phones have various identifiers associated with them, including some that cannot be easily changed (e.g., serial number, WiFi MAC address, IMEI, etc.). As mobile phones are very personal devices, a unique identifier for a mobile phone is consequently a unique identifier for that individual and can therefore be used to collect data about their activities, preferences, and demographics, simply based on data collection that associates it with the apps that were used, when, how, and where.

Why does this matter? By and large, this data is used for advertising purposes: these profiles are used to decide which ads to show which users, allowing advertisers to target individuals based on their inferred interests and preferences. However, the data is increasingly used for other purposes that are often completely opaque to consumers, particularly parents. For example, location data collected by apps is frequently resold to other businesses and used for everything from predicting social relations in the physical world, to predicting retail sales trends, for law enforcement surveillance, and even for political fundraising and advocacy. This data is being collected without consumers’ knowledge, and then is misused in ways that undermine individual rights. Worse, new uses for this type of data are invented all the time, which means that there’s no way of knowing exactly how collected data may be used in the future. Data collected from mobile apps and other services could end up being used for making major life decisions, such as whether offers of credit or employment are extended, or whether someone is admitted to a particular school, or even the type of medical care that they may receive. When this data comes from children, it is obviously even more concerning.
3 Research Findings

As part of prior research to study how mobile apps’ privacy practices comport with consumers’ expectations, my lab wrote bespoke instrumentation for the Android platform that allows us to run mobile apps and monitor exactly what personal data those apps access and to whom they transmit it [6, 7, 8, 2]. We wrote our tools for Google’s Android platform only because it is open source: having the source code for the operating system allowed us to modify it for this purpose; at the time, we didn’t look at Apple’s iOS simply because we didn’t have the source code to add the same level of instrumentation.

Starting in late 2016, we began downloading as many free apps in the “Designed for Families” (DFF) program as we could find, which ended up being just under 6,000 apps [3]. The DFF program is a section of the Play Store, Google’s centralized Android app market, which is exclusively for apps that are directed to children. Mobile app developers must participate in the program when they upload their app and disclose to Google that it is directed at children. As part of the program, they must affirm to Google that their app is in compliance with COPPA. Our goal was to evaluate whether that was the case in practice.

3.1 Collection of Contact and Location Information

In terms of the most serious privacy violations, we observed that roughly 300 of the apps that we tested (4.8%) were collecting children’s contact information (e.g., names, email addresses, and phone numbers) and/or precise location data, which included apps specifically targeted at children under 5. In most cases, this data was transmitted to third-party advertising companies, or third parties that otherwise support the advertising industry. I believe that this is a serious finding that should be put in perspective: roughly 1 in 20 of the apps that we examined were collecting information without the requisite verifiable parental consent, and for which the FTC has previously brought cases.

3.2 Insecure Transfer of Personal Information

The most common issue that we observed was the transmission of personal data using insecure means. Under COPPA, covered services are required to “establish and maintain reasonable procedures to protect the confidentiality, security, and integrity of personal information collected from children.”³ While neither the statute nor regulations define what are considered “reasonable procedures,” Transport Layer Security (TLS) and its predecessor have been industry standards for

more than three decades now; its use is required on U.S. government websites.\(^4\) Simply put, it is not considered “reasonable” to transmit personal information without the use of TLS to secure it. Nonetheless, we observed that 40% of the children's apps (2,344 apps) we tested failed to take this reasonable procedure.

What this means is that for users of these apps, their personal information is accessible to any eavesdroppers. This may include anyone sharing the same WiFi connection, as well as Internet service providers and other organizations. In an extreme case, this could enable someone to identify a specific child within a specific area, based on the insecure transmissions emanating from that child’s device.

### 3.3 Targeted Advertising

The remaining pervasive privacy issues that we discovered had to do with the collection of persistent identifiers. A persistent identifier is simply a label that is unique to an individual, such as a Social Security Number or the serial number of a personal device. While a persistent identifier might appear as an insignificant random number or combination of letters, as I explained, persistent identifiers are primarily what enable targeted advertising and other types of data collection. We identified multiple issues, including: (1) Google’s user privacy settings may fail to work due lack of policy enforcement and (2) many app developers fail to correctly configure third-party software components to limit data collection from children, resulting in children’s personal information being sent to third parties for targeted advertising and other purposes.

#### 3.3.1 Ineffective Android Privacy Settings

Prior to 2013, mobile apps for both Google’s Android and Apple’s iOS mobile operating systems collected a variety of different non-resettable identifiers that were used to track consumers. Unlike cookies in the web browser, which can be periodically cleared by the user, many of these identifiers cannot be reset, and so mobile device users had neither transparency into who was tracking them nor when they were being tracked, nor any control over it. In response, both Apple and Google created software-based “advertising identifiers” that could be reset through user-facing privacy controls. By policy, both platforms mandate that only these identifiers be used to track users, in lieu of other non-resettable identifiers. This is so that a consumer can opt out of tracking via the provided settings interface. However, as we discovered on Android, compliance with this policy is not enforced by Google: app developers and the third-party mobile SDKs embedded within their apps are able to collect other non-resettable identifiers alongside the advertising ID. When this happens, if a consumer resets their advertising ID or uses the privacy settings interface to opt out of tracking altogether, data recipients are simply on their honor to stop tracking that consumer.

---

\(^4\)https://https.cio.gov/
We observed that 39% of the children’s apps that we tested transmitted non-resettable identifiers alongside the user-resettable advertising ID. What this means is that for users of these 2,281 apps, Google’s ad privacy settings may simply be ignored.

### 3.3.2 Ineffective SDK Privacy Settings

Software engineering, like many other types of engineering, involves building products out of many pre-made components. For example, just as a car manufacturer does not make all the components in its cars (e.g., springs and shocks may come from other manufacturers, sheet metal is purchased from suppliers, etc.), a mobile app developer does not necessarily write all of the code found within their apps. Third-party software development kits (SDKs) allow developers to include pre-made software components, saving them time and effort. For example, rather than find advertisers, organize and/or create ad copy, and then determine which users to show which ads, app developers can simply outsource that work by incorporating a third-party ad SDK that has already implemented those things. There are third-party SDKs that help developers with displaying graphics, processing payments, streaming audio or video, and so forth. This type of “code reuse” is an accepted part of modern software engineering. However, it creates enormous risks, especially when app developers fail to verify that third-party components are functioning as expected (or if third-party components are misused).

Many of the potential COPPA violations that we observed were due to the data collection behaviors of third-party SDKs, and not necessarily due to code written by app developers; nonetheless, most apps embed these third-party SDKs, and therefore they impact a lot of apps. Many of these SDKs, because they are for use in a wide variety of mobile apps, offer app developers configuration options so that they can be customized to an app’s needs. Specifically, many of the SDKs that collect personal data with COPPA implications—those that may be used to collect personal information from children—offer developers configuration options to enable a COPPA-compliant data-collection mode. When the app developer uses one of these directives to signal that the user is a child, the SDK is instructed to either not use that child’s personal information for COPPA-prohibited purposes or to not send that data to its servers altogether. When developers of children’s apps fail to correctly configure these types of options, it likely results in children’s personal data being collected for targeted advertising and other prohibited purposes.

We observed that few developers were correctly configuring third-party advertising SDKs to disable the collection of personal information for profiling and/or ad targeting purposes. For example, we observed that 1,280 of the children’s apps we tested (21.9%) transmitted users’ personal information to Facebook’s servers. Of these, only 75 (5.9%) correctly signaled to Facebook that the user is a child and that the data should be handled pursuant to COPPA. However, Facebook is not an isolated example: of the third-party SDKs that we observed collecting personal
information and that offered options for child-directed treatment, none were consistently configured correctly by app developers.

Other third-party SDKs simply provide terms of service that prohibit their use in child-directed apps. However, we observed that developers of children’s apps use these SDKs anyway. By reading the terms of service and privacy policies of these data recipients, my research team identified several data recipients who (1) describe using data received from their SDKs for practices that would be prohibited by COPPA, if that data were to come from children; and (2) prohibit inclusion of their SDKs in child-directed apps and disclaim any knowledge of receiving data from children. Despite this, we identified 1,100 children’s apps transmitting personal information to these companies (18.8% of the children’s apps we tested).
4 Recommendations for Fixing COPPA

Based on my research, which exposed evidence of rampant non-compliance with COPPA’s existing requirements, I have several recommendations for strengthening COPPA, which I detail in this section:

- Moving from an “actual” to “constructive” knowledge standard
- Eliminating the internal operations exemption
- Fixing the Safe Harbor program
- Increasing enforcement

4.1 Moving from “Actual” to “Constructive” Knowledge

Many of the potential violations that we observed amounted to sharing of persistent identifiers—without verifiable parental consent—with companies whose privacy policies state that those identifiers will be used for user profiling and/or behavioral advertising, activities that are prohibited by COPPA (when that data comes from children). These persistent identifiers are generally collected and transmitted by third-party SDKs, and so it is plausible that many app developers simply do not know when this data is being transmitted. However, the third-party data recipients know, and in most cases, the information that they are currently receiving allows them to trivially determine that the transmitting app was directed at children.

The privacy policies of many of the companies that receive personal information from children’s apps state they are directed at general audiences and have “no actual knowledge” of receiving personal information from children, thereby absolving them of any responsibility under COPPA. This, however, ignores the fact that each transmission from an SDK usually includes the name of the app that transmitted the data. The claim that a third-party data recipient does not have actual knowledge relies on not knowing whether a particular app is targeted at children. Yet, when one looks at the marketing materials of the companies receiving this data, and their business models, it is apparent that this is precisely the type of knowledge that they claim to possess!

Many online advertising business models rely on knowing the demographics of specific apps so that they can target ads based on those demographics. That is, their internal data allow them to already know or trivially find out which apps are child-directed. For data recipients who genuinely do not maintain that data, they can simply query the Google Play Store to determine whether or not a given app is in the Designed for Families program (and therefore targeted at children) based on its public metadata. I can personally write and test the code to do this.
in under an hour. There are also many commercial offerings that offer companies programmatic access to this type of data. But despite the ease with which data recipients could automatically determine whether or not they are receiving data from a child-directed app, they choose not to. Instead, most developers of third-party SDKs place the burden on app developers, rather than using the information that is likely already in their possession—or trivially available to them—to automatically configure their services for COPPA compliance.

As I have observed in the course of my research, many app developers configure these settings incorrectly (or are simply unaware that such settings exist), which results in children being tracked and profiled. If third-party data recipients are held to a “constructive knowledge” standard, under which they would be required to use the information at their disposal to identify whether the data they receive originates from child-directed services, this would not only result in greater compliance and reduced harm to children, but it would also result in drastic cost savings, especially amongst smaller software development companies and individual entrepreneurs. One ad network using their existing data—or data reasonably available to them—to automatically apply child-directed treatment to the data they receive would negate the need for app developers to individually spend time and effort to correctly configure that company’s SDK to do so. More to the point, a constructive knowledge standard would shift the burden of compliance away from millions of small app developers—who would still need to report whether or not their apps and services are child-directed—to the significantly fewer number of data recipients, who are much better positioned to apply privacy protections to the data that they collect (and are much more likely to do so correctly). In sum, my research and experience suggest that moving to a constructive knowledge standard would result in fewer incidents of children being inadvertently tracked and profiled, as well as economic savings to businesses by lessening their compliance costs.

### 4.2 Eliminating the Internal Operations Exemption

Currently, persistent identifiers can be collected from children without parental consent if they are used for the site or service’s “internal operations,” which are currently defined by regulations as using the data to:

1. Maintain or analyze the functioning of the Web site or online service;
2. Perform network communications;
3. Authenticate users of, or personalize the content on, the Web site or online service;
4. Serve contextual advertising on the Web site or online service or cap the

---

5 16 C.F.R. §312.2
frequency of advertising;
5. Protect the security or integrity of the user, Web site, or online service;
6. Ensure legal or regulatory compliance; or
7. Fulfill a request of a child as permitted by §312.5(c)(3) and (4);

From a technical standpoint, the collection of persistent identifiers that allow a user’s activities to be tracked between apps is unnecessary for any of these purposes. The primary issue is that each of these use cases could be facilitated by an identifier that is unique to a session, an app installation, or developer, which in turn could not be used to track the user across other apps and services. For example, serving a contextual ad simply requires knowing the type of app or website that a user is using or visiting, which is information that is already collected; by definition, contextual ads are based on those things alone and not the user’s identity, and therefore do not require the collection of persistent identifiers. Similarly, conversion tracking, measurement, fraud detection, and advertising attribution also do not need persistent identifiers that can identify users across apps. If they are not performing COPPA-prohibited profiling and behavioral advertising, an advertising company only needs to know how many people clicked on a specific ad, not who those individuals are. When user-specific identifiers are needed, ephemeral app-specific or session-specific identifiers can be used. This functionality is already supported on both Android and iOS, and therefore eliminating the internal operations exemption should not create an undue compliance burden.

Furthermore, claims that persistent identifiers are needed for these purposes are disingenuous because many app developers are already prevented by platform policies from using identifiers for many of these purposes. Indeed, on iOS, if a user opts out of online tracking, apps are outright prevented from accessing identifiers that could be used to track that user’s behaviors across apps. Further, Apple already requires that no persistent identifiers can be collected from children’s apps. Google provides best practices for developers that explain how ephemeral identifiers can be used for many of these use cases to preserve user privacy. Thus, it is patently false to claim that persistent identifiers are necessary for these purposes.

The FTC has previously advocated for companies to take a “data minimization” approach to online privacy. I recommend that the Subcommittee heed this advice with regard to children’s privacy: because long-term persistent identifiers are unnecessary for these purposes, the internal operations exemption should be eliminated from COPPA.

---

7 https://developer.android.com/training/articles/user-data-ids
4.3 Fixing the Safe Harbor Program

The FTC is charged with certifying Safe Harbor self-regulation programs under COPPA. As of this date, the FTC’s website indicates that seven such programs are currently certified. In the course of my group’s research [3], we identified 237 Android apps that gave outward appearances of having been certified as COPPA-compliant by these programs. Yet, when we examined their behaviors, we observed that 24 (10%) collected location data and/or contact information without verifiable parental consent, while 77 (32%) transmitted personal information without taking “reasonable” security precautions (e.g., using TLS encryption). We concluded that apps certified by these programs were just as likely to comply with COPPA as apps not certified by them. Indeed, this finding is consistent with prior research on industry self-regulation, which found that websites receiving trust certifications “are more than twice as likely to be untrustworthy as uncertified sites” [1]. This begs the question, if an organization is already complying with the law, why would they spend additional money to protect themselves from enforcement of that law?

Given the poor incentive structures and lack of transparency into how apps are being certified or even determining which apps are certified, current Safe Harbor programs do not appear to be effective. I have three suggestions for improvements that can be made:

1. **Apps and services should be certified only after independent forensic evaluations of their privacy behaviors.**

2. **The FTC should develop, in consultation with privacy experts, standards for forensic evaluations of mobile apps’ privacy behaviors.**

3. **Certification organizations should publish lists of the apps that they have certified (including versions).**

Based on my examination of the public documents that describe COPPA Safe Harbor certification processes, it appears as though current certification processes rely primarily on self-reports from app developers, rather than forensic examinations of their apps (that would yield the type of data that is necessary to assess compliance). Given that many app developers are unaware of the privacy issues associated with their apps, it would hardly be a surprise that those behaviors do not get disclosed to the certification organizations, resulting in COPPA-violative apps inadvertently being certified.

Relatedly, one of the hardest parts of my analysis was simply finding the apps that had been certified by each organization, as many did not publish information about

---

9 https://www.ftc.gov/safe-harbor-program
how they certified each app nor what specific apps or versions were even certified. Instead, we relied on press releases from those companies, as well as images and text on their websites and references in the privacy policies of individual apps. Upon publication of these findings, many Safe Harbor organizations claimed that the apps that we examined were not actually certified by their organizations (despite their names and logos appearing on each other’s websites). Given that a team of multiple PhDs and a lawyer could not disambiguate what has and has not been certified by each program, it is hard to expect the average parent to be able to. Thus, by mandating that this information be public and in an accessible manner, not only would it empower parents to make better decisions, but it would strengthen the free market through increased transparency, thereby promoting competition.

4.4 Increasing Enforcement Efforts

Finally, all of the above changes are moot without increased enforcement efforts. In under a year of work, my research lab identified the transmission of personal information for tracking and advertising purposes from literally thousands of child-directed mobile apps. At the same time, the FTC, the primary entity empowered with enforcing COPPA, historically has pursued only 1-2 COPPA enforcement actions each year. This is not for want of known violations. To be clear, the FTC employs very capable attorneys and technologists who do excellent work. The problem is that there simply are not enough of them to investigate all of the violations brought to their attention. As the primary agency tasked with enforcing COPPA, it is my opinion that the FTC does not have enough resources to bring enough cases for the threat of enforcement to serve as a deterrent; similar resourcing problems appear to prevent state attorneys general from filling this enforcement vacuum. Simply put, if the FTC continues to not receive funding commensurate with its enforcement responsibilities, COPPA will remain another unfunded mandate.

I strongly believe that the enforcement problems can be addressed in two complementary ways. First, the FTC needs a significant increase to its privacy enforcement budget. However, unless this budget is increased by orders of magnitude, it is still unlikely to be enough for them to be able to investigate all of the potential violations brought to their attention. That is why I believe that as a second recommendation, Congress should look to the free market and create a private right of action. With a private right of action, market forces will drive compliance, while at the same time, they will also drive competition among industry self-regulation programs. These industry self-regulation programs can then be better regulated by the FTC to ensure that they are accurate and transparent.
5 Conclusion

My research has shown that despite COPPA, mobile apps directed at children frequently collect children’s personal information and share it with third-party advertisers and data brokers. I believe that many of the problems that I’ve outline in this testimony can be addressed through changes to COPPA. I believe that these proposed changes will result in greater levels of compliance amongst online services, increased transparency for parents, better protections for their children, and increased competition in the marketplace.

Thank you for giving me the opportunity to testify today. Please do not hesitate to follow up with me regarding any questions that you may have.
References


