Issue Brief: Shadow Profiling and User Control

By: Ginny Kozemcak, Will Monge, Nathan Good, Gauri Gupta, Joel Reardon, Dan Kinney and William Boag

Executive Summary

A majority of Americans feel they have little control over how their personal data is collected and shared online. There is widespread concern about the future of our privacy as users lack proper legal protections and control over their data. And, because much of our data flows through our mobile devices, the mobile app ecosystem is especially subject to data abuse. The International Digital Accountability Council (IDAC) is concerned about a trend identified throughout several of our investigations: the circumvention of user privacy preferences on mobile devices through the proliferation and misuse of identifiers or numbers that are uniquely assigned to an individual's mobile device in order to track users over time. While this practice can be put to good use, for example, to address fraud or for enterprise mobile management, our investigations found that many mobile apps collect multiple identifiers without notice or choice for users in order to monetize the personal information collected for third party profiling and advertising. Some apps also share personal information with data brokers.

We refer to the practice of organizing and linking information gathered through unique identifiers as "Shadow Profiling". The app may employ **ID bridging**, which links multiple identifiers for the same user on different devices and across multiple apps, which permits tracking over time across platforms. ID bridging circumvents user privacy preferences by using the link to maintain the association with a user even after the user or platform has reset a past identifier.

These practices appear to be "unfair" and "deceptive" under Section 5 of the FTC Act and to violate existing law. For example, the Children's Online Privacy Protection Act (COPPA) requires verifiable parental consent before collecting or sharing any personal information from children, including persistent identifiers such as unique device identifiers and the newly-enacted California Privacy Rights Act, which applies to data collected as of January 2022, provides a number of new privacy protections, including, among other things, opt-in sharing of "sensitive personal information," limits on profiling of such data, and a new consumer right to opt-out from sharing personal information for "cross contextual behavioral advertising."

Both the Apple and Android platforms have also adopted measures to limit persistent tracking in their operating systems by moving toward the use of resettable advertising identifiers in lieu of permanent identifiers. In particular, Apple's recent change to the privacy rules for apps in the Apple Store, which require apps to provide a pop-up notice asking for opt-in consent to track the user across websites and apps from other companies, is likely to significantly impact shadow profiling. But the size and scale of the app ecosystem also requires due diligence by investors and developers, new third-party regulatory and certification initiatives and cooperation and support for independent investigators and researchers.

This issue brief outlines IDAC's previous research and possible use cases for identifier transmission, including those purposes *valuable to the user* such as security, anti-fraud, telephony, or enterprise mobile management -- as well as more problematic and potentially privacy-violating ones, such as user tracking, the creation of behavioral profiles, and harmful targeted advertising.

Part I: Defining Identifier Transmissions, Shadow Profiling, ID Bridging

What are identifiers and how are they used on our devices?

A mobile phone contains a variety of data elements that can be used to uniquely identify it. These are typically unique values tied to the hardware of the device or are assigned by the manufacturer. These types of identifiers are permanent, and the user cannot easily change them. We refer to these as **permanent identifiers**.

In addition, platforms have introduced unique identifiers for customers, which can be changed in order to limit tracking over time. These non-permanent identifiers are akin to a pseudonym. For example, imagine that every time a customer ordered a coffee at a local coffee shop, they had to provide their social security number so the shop could associate people with their coffee order. This kind of data collection would be excessive for this type of transaction and it gives the coffee shop more information than the transaction requires. Now imagine the coffee shop has a reward program the customer would like to join; instead of providing their social security number, the customer gives a pseudonym. The coffee shop can still customize their experience, keeping track of coffee orders and noting preferences, but without collecting and storing personally permanently identifiable information like names and social security numbers. A user can change their pseudonym, which may result in the loss of this accumulated history and preferences but is offset by the user's desire to reset the relationship. The pseudonym is the user's **resettable identifier**.

However, apps are often not respecting user decisions about resetting identifiers. In collaboration with our partners at AppCensus and Good Research, IDAC has investigated hundreds of apps, using technical tools that analyze apps' data flows. We found many instances where apps **simultaneously transmit multiple identifiers**, collecting and sending a device's permanent and resettable identifiers to either itself and/or a third party.

For example, IDAC's investigation into close to 500 global education technology (ed tech) apps spanning 22 countries, found 218 apps sending multiple identifiers. In an investigation of children's apps, we found the same practices in popular apps including Princess Salon, Number Coloring and Cats & Cosplay, which together had been downloaded more than 20 million times. Research by the International Computer Science Institute and Norwegian Consumer Council have also found that this practice occurs on many of the most popular apps across the Android and iOS app stores.

There are certain situations in which the practice may be necessary or have benefits for the user, such as security and anti-fraud measures. In those instances, transmitting multiple identifiers should be allowable because the developer has a narrowly tailored reason and the user's consent to do so.

However, when there is an unnecessary collection of identifiers sent to third party advertisers or data brokers, the app risks engaging in **shadow profiling**. Data brokers and third-party advertisers may amass significant amounts of personal data without users' knowledge for advertising or sale in a market with minimal oversight.

ID bridging is a commonly used tool to build a shadow profile. Multiple Identifiers are not only collected and shared but also *bridged* or linked together by the developer or third-party data processor. The practice of bridging is intended to identify a given user over time and to defeat users' efforts to limit tracking. If one identifier changes, an app developer or a data broker can simply use the other IDs to continue to identify that device. In this way, ID bridging will enable a party to track a user or a device regardless of ID changes made by the user or operating system. Using the example above, if the coffee shop collects the customer's social security number along with their pseudonym every time they order coffee, then, despite their ability or choice to change aliases, the shop will still be able to identify and track users because they also collected a permanent social security number. From a more technical perspective, these two different unique identifiers are linked so that if one of the identifiers is changed by either the user or the operating system, the other linked and unchanged value can bind the old and new values.

In the following sections, we examine each of these practices more closely and explain why ID bridging in particular poses real privacy harms to users.

Part II: Simultaneous Transmission of Multiple Identifiers

A developer may simultaneously transmit multiple identifiers in order to best serve its users. In Android operating systems, the permanent identifier is known as the Android ID and the resettable identifier is known as the Android Advertising ID, or AAID. In iOS, Apple's resettable identifier is known as the Identifier for Vendors, or IDFV, and it can only be accessed after a developer has received user consent. This consent requirement became mandatory with iOS 14.5 through Apple's new App Tracking Transparency (ATT) policy. Developers are not allowed to circumvent this measure or use any information to create an alternative stable identifier, or "fingerprint," of the user for tracking purposes.

Collecting multiple identifiers can, in theory, help protect against fraud,[1] provide data for internal analytics, and aid in recovering user preferences in the event of app crashes, all of which improve the functionality of the app.

Additionally, in some instances, an app may separately collect pieces of information for different purposes and never link these identifiers together. For example, a company's engineering team may collect the permanent identifier for diagnostics or debugging errors,[2] while its marketing department collects the resettable identifier for advertising and analytics purposes. In this case, there is no ID bridging because the company never linked these two identifiers.

On the other hand, anti-fraud, telephony, and enterprise device management are all use cases that may benefit users, but nonetheless pose potential privacy risks. For example, to protect against fraud, financial technology and online banking apps often use tools that evaluate a financial transaction in real time to assess the risk of fraud, sending information collected to third-party transaction handlers. While these transmissions may be intended to serve as an anti-fraud measure, they often collect information

[1] Many security advisors believe that device identifiers are an important tool in security measures, but are no longer sufficient in and of themselves.

[2] Google's developer guidance encourages the use of the Android ID (or SSAID) for tracking signed out users across apps that are created by the same developer and share the same signing key, as these apps would also share the same Android ID. On the other hand, the Advertising ID is recommended for tracking signed-out user behavior across unrelated apps on the same device in accordance with Google Play Developer Content Policy rules.

extraneous to the transaction purchase, including contextual information linked to persistent hardware identifiers, in order to assess the risk that the particular transaction is fraudulent.

Apps that deal with telephony -- carrier apps that help users configure and manage their phone line and voicemail settings -- might need to use identifiers such as the International Mobile Equipment Identity (IMEI) or a phone's SIM card ID. The IMEI may also be necessary to collect when organizations need to remotely manage corporate mobile devices assigned to their employees as part of a Mobile Device Management system. In order to remotely add new devices without having an administrator physically interact with it, a privileged app might need access to hardware identifiers, such as IMEI.

These use cases relayed above are limited and well-known by platforms and the platforms have attempted to reduce the availability of these identifiers in ways that do not impact beneficial use cases. Since the release of the Android 10 mobile operating system, an app can only access the IMEI identifier in limited circumstances with special permissions, such as carrier permission, device administrator permission, or the "read" permission. And Apple has made special exceptions to its policies, relaxing user permission requirements for anti-fraud measures.

Ultimately, even internal data collection that is intended for the benefit of the user may be problematic because an external party would be unable to distinguish between the purposes for which data is being collected and used, making policing the system a difficult task.

Part III: Shadow Profiling and ID Bridging

Shadow profiling encompasses a range of techniques that companies and data brokers use to collect as much identifying information about users as possible, which may include cookie syncing and browser fingerprinting. Third-party cookies in particular are one of the key online marketing tools used to track users. Advertisers routinely use cookies to collect and store large amounts of personal data obtained from users to create targeted ads. With Google's announcement to phase out third-party cookies by 2022, marketing companies are now developing alternatives to third-party cookie tracking that rely on first-party cookies and Google's Privacy Sandbox initiative.

Our past research shows that when shadow profiling involves the use of permanent and resettable identifiers, it is likely that unnecessary data collection lacking a clear, beneficial purpose for the user occurs. Developers can avoid this by incorporating privacy-by-design principles, such as ensuring that privacy is the app's default setting and limiting data collection only to what is necessary.

When IDAC conducted investigations of education technology and COVID-19 related apps, we observed hundreds of apps collecting permanent and resettable identifiers without any discernible, tailored purpose. This was especially concerning considering that many individuals and families have had to rely on these apps for education and healthcare needs.

In these prior investigations, we observed apps sending identifiers to outside third parties, which in some cases, likely resulted in at least some ID bridging. As explained earlier, ID bridging occurs when identifiers are both transmitted and *linked*. On a more technical level, it occurs when a resettable identifier is linked to permanent identifiers such as device identifiers, IMEI, other hardware identifiers, various contextual identifiers such as the name of the device's currently used Wi-Fi network, or MAC address of the Wi-Fi access point in use. In some cases, a list of the device's apps, user's email, phone number, and other pieces of information are linked to the resettable identifier.

Apps may send permanent and resettable identifiers to themselves, for reasons we discussed in the previous section, or to third parties. IDAC is especially concerned about transmissions to third parties because *even if the developer is not linking identifiers, the receiving party may do so*. In these cases, bridging may occur by entities that the user is not even aware is party to the data transaction.[3] These third parties often include mobile advertising companies and mobile analytics companies, and some companies even perform a combination or all these functions. Mobile ad networks enable publishers to display their advertisements in websites and apps, while analytics firms will provide companies with data on traffic, user engagement, conversion, and ad revenue.

IDAC's investigations have shown apps sending permanent and resettable identifiers to ad networks and analytics firms; if multiple apps collect and bridge a device's identifiers, then a third party can track the user across apps. For instance, imagine User An extensively uses Game App, Social Media App, and Food Delivery app on their phone. All three apps transmit permanent and resettable identifiers to the Super Ads & Analytics Company. Super Ads & Analytics can now identify User A and compile a significant amount of data from all three apps -- tracking User A's social media likes, restaurant order history, and gaming behavior.

If User A owns a device that uses an Android operating system, they can go into their device settings, tap "Reset Advertising ID" to create a new AAID, and in theory, be able to create a "clean slate" without historical data about their online activities. However, if User A's Advertising ID is linked to permanent identifiers, that option is effectively voided because, according to the Norwegian Consumer Council, a third party like Super Ads & Analytics can "simply append the new Advertising ID to the other identifier, and resume tracking the user." If User A owns a device using an iOS operating system, Game App, Social Media App, and Food Delivery app would all be required to obtain her consent in order to collect both the resettable and permanent identifiers. Nevertheless, developers and advertisers may already be finding ways to circumvent Apple's new framework.

We explain both the harms of user tracking and Google and Apple's policies in greater detail in the following sections, but the primary concern here is that the lines between simultaneous transmissions of multiple identifiers, shadow profiling, and ID bridging are at times blurry, if not indistinguishable. In a best case, identifiers can aid security and anti-fraud measures -- but they can also allow for unnecessary data collection or, at worst, enable invasive tracking unbeknownst to users. And without access to companies' back-end data management, the platforms and watchdog organizations are unable to fully monitor and ensure compliance with best practices and system rules.

Part IV: The Harmful Impacts of User Tracking and Targeted Advertising

[3] However, not every third-party transmission is necessarily invasive: some third parties act as an extension of the first party company. For example, a third-party may be used as a database or backend as a service for the app developer, without sharing the information outside of their organization. While such entities may appear to be a third party, they may simply function as an extension of the first party to provide the app developer with better infrastructure and support services. Similarly, web hosting services such as AWS and Google web services may be used by app developers for renting web servers, which does not necessarily mean the third party actually has access to the data transmitted. Not all third parties track users, and some third parties act under strict contractual obligations with the app developer to provide a particular service.

Tracking users is financially profitable for many companies, so in order to understand why shadow profiling and ID bridging are so prevalent, these practices must be placed in the context of the mobile advertising industry. Ad tech's underlying business model "incentivizes companies to amass as much information as they can: what their users do on the platforms themselves and what they do elsewhere on the internet." Information such as web browsing history, geographic location, and shopping habits is gathered and analyzed in order to maximize revenue streams. Companies like mParticle, for instance, "collect customer data to drive better customer interactions." These customer interactions are based on "historical context" and can lead to "hyper-personalization in real time." Historical context is, essentially, the user's history of activity, which is compiled into personal profiles. Similarly, the company Amplitude gives developers access to "user properties," or the user's activity on an app including their preferences and device details.

Many companies admit these hyper-personalized user profiles are created to increase users' time with an app and the likelihood of clicking on an advertisement -- also referred to as increasing an advertisement's return-on-investment (ROI). The mobile advertising industry relies on a system called real-time bidding, in which "tens or even hundreds of advertisers automatically bid to display their ads based on determining the value of the individual consumer." The higher the value of the customer, the greater potential ROI. For instance, AppsFlyer allows developers to "easily determine the lifetime value of each user" in order to "retarget high-value users with relevant ads, push messages or emails to improve their value." Similarly, Branch can "increase mobile revenue with enterprise-grade links built to acquire, engage, and measure across all devices, channels, and platforms."

While the mobile advertising industry is eager to utilize its data-rich field, consumer and privacy advocates are increasingly concerned about the social harms caused by the ubiquity and inescapability of these practices. Targeted ads are not only used to sell products, they are in used to spread disinformation and clickbait and may lead to price discrimination through personalized sales offers

Most importantly, targeted advertising can harm marginalized communities and undermine legal protections against discrimination in housing, employment and credit. The Equal Employment Opportunity Commission declared that targeted ads for employment and housing violate civil rights laws, and Amnesty International argues that it undermines fundamental rights to privacy, equal access to information, and free speech.

Trust in the digital ecosystem may be undermined, as well. If the industry's underlying financial incentive relies on amassing as much data as possible, that will in many instances conflict with privacy-by-design principles such as data minimization and privacy-by-default settings. If an individual relies on an app for their education, healthcare, or financial needs, but they experience invasive tracking as a result, it erodes their trust in those digital services.

Part V: Platform Policies, Laws, and Regulations

Currently, Google and Apple have differing policies on accessing users' permanent and resettable identifiers.

Google's guidelines on Usage of Android Advertising ID allow the transmission of multiple identifiers so long as the identifiers are not bridged once they arrive in a developer or third party's database. Google does allow identifiers to be bridged for analytics purposes, but *only* with the explicit consent of the user. For instance, an app may collect and bridge the Android ID and AAID for crash analytics, but it must first

obtain the explicit consent of the user, ideally as a first sign-in notification. Google's policy also states that the advertising identifier may *not* be connected to persistent device identifiers for advertising purposes. However, the AAID may be connected to *personally identifiable information* for advertising purposes, again, only with the explicit consent of the user.

Apple, on the other hand, does not allow alternative identifiers to track users for any purpose, except with the explicit consent of the user. It also instructs developers to refrain from using any information or signal to create a fingerprint of the user for tracking purposes.

While Google and Apple are taking important steps to address tracking in their platform policies, privacy advocates are increasingly looking to legal measures to better protect users. According to the European Union's GDPR, unique identifiers relating to an individual are classified as personal data. For instance, under Article 4, the GDPR defines personal data as "any piece of information that relates to an identifiable person," including IP addresses, cookie identifiers, and "online identifiers provided by devices," especially when "combined with unique identifiers." Since identifiers are subject to the GDPR, data protection rights such as the right to access, erase, rectify, move or object to the processing of their personal data all apply. Organizations such as NOYB -- European Center for Digital Rights -- are filing GDPR complaints to ensure that apps and platforms use identifiers in compliance with the GDPR.

In the United States, there is no comprehensive federal privacy law. Instead, a patchwork of industry-specific privacy laws has led to gaps that leave users unprotected. IDAC has argued that a national privacy law with enforceable codes of conduct would better protect users, provide guidance to developers, and foster international cooperation in establishing privacy standards. Law enforcement and consumer protection agencies such as the Federal Trade Commission and state attorneys general should have access to more resources to enforce current laws.

State privacy laws such as the California Privacy Rights Act (CPRA) , which will apply to data collected in January 2022 will also play an important role in deterring invasive tracking. The CPRA applies to companies that conduct online transactions with California residents or have other connections to California, with a few exceptions. The CPRA also gives California consumers rights over their personal data, which is defined as "information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked directly or indirectly with a particular consumer or household." Mobile device identifiers fall under this category, and thus users' data rights apply as well, including the right to request that a company *not* sell personal data.

Conclusion

Shadow profiling and tracking undermines user choice that is critical to ensuring consumer privacy and erodes trust in the digital ecosystem. It evades the only meaningful control that Android mobile device users have to stop tracking for behavioral advertising: resetting their identifier with the purpose of assuming a fresh, new identity. Shadow profiling poses a serious risk to the overall integrity of the mobile app ecosystem. In an increasingly mobile-centric world, platforms, developers and investors should take immediate steps to stop shadow profiling; enforcement bodies should use existing law to put an end to the practice and policymakers should ensure privacy laws deter these practices.