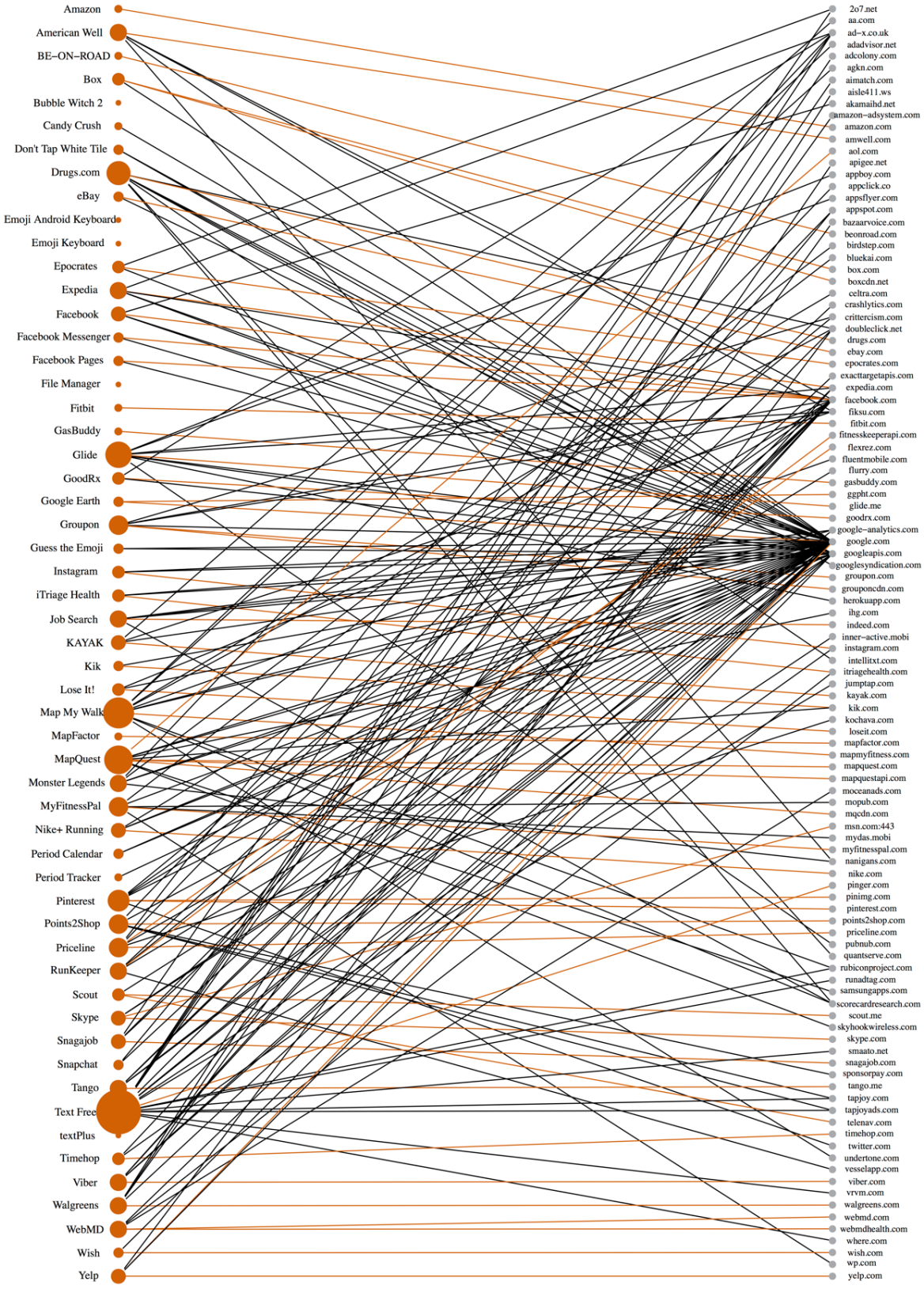
Zang J, Dummit K, Graves J, Lisker P, Sweeney L. Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps. Technology Science. 2015103001. October 30, 2015. http://techscience.org/a/2015103001

Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps

Jinyan Zang, Krysta Dummit, James Graves, Paul Lisker, and Latanya Sweeney

Highlights

- We tested 110 popular, free Android and iOS apps to look for apps that shared personal, behavioral, and location data with third parties
- 73% of Android apps shared personal information such as email address with third parties, and 47% of iOS apps shared geo-coordinates and other location data with third parties
- 93% of Android apps tested connected to a mysterious domain, safemovedm.com, likely due to a background process of the Android phone
- We show that a significant proportion of apps share data from user inputs such as personal information or search terms with third parties without Android or iOS requiring a notification to the user



Sharing of sensitive data by Android apps (left) to domains (right)

Abstract

What types of user data are mobile apps sending to third parties? We chose 110 of the most popular free mobile apps as of June-July 2014 from the Google Play Store and Apple App Store, across 9 categories likely to handle potentially sensitive data about users including job information, medical data, and location. For each app, we used a man-in-the-middle proxy to record HTTP and HTTPS traffic that occurred while using the app and looked for transmissions that include personally identifiable information (PII), behavior data such as search terms, and location data, including geo-coordinates. An app that collects these data types may not need to notify the user in current permissions systems.

Results summary: We found that the average Android app sends potentially sensitive data to 3.1 third-party domains, and the average iOS app connects to 2.6 third-party domains. Android apps are more likely than iOS apps to share with a third party personally identifying information such as name (73% of Android apps vs. 16% of iOS apps) and email address (73% vs. 16%). For location data, including geo-coordinates, more iOS apps (47%) than Android apps (33%) share that data with a third party. In terms of potentially sensitive behavioral data, we found that 3 out of the 30 Medical and Health & Fitness category apps in the sample share medically-related search terms and user inputs with a third party. Finally, the third-party domains that receive sensitive data from the most apps are Google.com (36% of apps), Googleapis.com (18%), Apple.com (17%), and Facebook.com (14%). 93% of Android apps tested connected to a mysterious domain, safemovedm.com, likely due to a background process of the Android phone. Our results show that many mobile apps share potentially sensitive user data with third parties, and that they do not need visible permission requests to access the data. Future mobile operating systems and app stores should consider designs that more prominently describe to users potentially sensitive user data sharing by apps.

Introduction

Since the introduction of the Apple App Store in 2007 and the Google Play Store in 2008, smartphones (and more recently, tablets and other devices running mobile operating systems) have become a dominant means of personal computing. Smartphones run programs called applications or “apps,” and the Apple App Store and Google Play Store make millions of apps available for all kinds of uses. Google reports one billion monthly active users (MAUs) on its Android platform [1], and estimates put iOS MAUs at 500 million-600 million [2]. More than 1.5 million different apps are available to users on both the App Store and Play Store [59], with the average consumer using 26 apps per month [93].

Given the popularity of apps on smartphones, consumers worry about how much personal information apps share. In a survey of more than 2,000 Americans, the Pew Research Center found that 54% of users decided to not install an app after learning about how much personal information they would need to share to use it [4]. Pew indicated that 30% of users reported uninstalling an app already on their phone because they learned that it collected

personal information that they did not want to share [4]. Similar rates of avoiding apps or uninstalling apps due to privacy concerns are seen for both iOS and Android users [4]. Consumers are sensitive about the collection of geolocation data, with 30% of smartphone owners turning off the location tracking feature of their phone owing to concerns about who might access that information [4]. In a different survey of more than 1,100 Americans, 70% of respondents said that they would "definitely not allow" a cellphone provider to use their location to tailor ads [5]. In another survey of more than 3,100 Americans, 60% reported being "very upset" if an app shares their location with an advertiser [6].

Governments are starting to address data collection and sharing in apps. For example, California's Online Privacy Act, last amended in 2013, requires developers to have privacy policies that state whether third parties can collect personally identifiable information on users [7]. In 2013, the Federal Trade Commission (FTC) pursued Goldenshore Technologies, LLC for violating the Federal Trade Commission Act, because the company's "Brightest Flashlight Free" app had a privacy policy that did not reflect the app's use of personal data, including location data, and because the app presented consumers with a false choice about sharing location data [8]. Internationally, in 2013, the European Union's Article 29 Data Protection Working Party ruled that the 1995 Data Protection Directive and the ePrivacy Directive apply to all mobile apps regardless of the developer's country of origin, and that users must first give consent before an app can install or access any information from their device and send it to third parties [9].

Why do apps trigger concerns from consumers and governments? First, an app may share a unique IDs related to a device such as a System ID, SIM card ID, IMEI, MEID, MAC address, UDID, etc. The ID can be used to track an individual [10, 11]. Second, an app can request user permission to access device functions and potentially personal or sensitive data, with the most popular requests being access to network communications, storage, phone calls, location, hardware controls, system tools, contact lists, and photos & videos [12, 13]. Some apps practice over-privileging, where the app requests permissions to access more data and device functions than it needs for advertising and data collection [14, 16, 17, 18, 19]. Third, any data collected by the app may be sent to a third party, such as an advertiser [10, 46, 58]. Fourth, a user may have a hard time understanding permission screens and other privacy tools in a device's operating system [15, 21].

In 2010, a *Wall Street Journal* (WSJ) study raised concerns about the amount of data sharing by apps. The WSJ conducted a survey of 101 popular Android and iOS apps [22]. By using network analysis to examine the data transmitted by different apps, the WSJ found that 56 apps sent the device's unique ID to third parties without a user's awareness or consent [22]. Forty-seven apps sent the device's location [22]. Five sent age, gender, and other personal details to third parties [22]. One major beneficiary of this data sharing was Google, with its AdMob, AdSense, Analytics and DoubleClick products receiving data from 38 of the 101 apps tested [22]. Publication of the *Wall Street Journal* report prompted multiple lawsuits against

Apple, Pandora, The Weather Channel, Dictionary.com, and 5 other app developers in the US and Canada [23, 24, 25].

The marketplace for apps has changed significantly since 2010. Certain regulators, such as California's Attorney General, now require privacy policies for all mobile apps in the belief that policies will specify personal data collection and sharing [7]. In 2012, Apple's App Stores started displaying an app's privacy policy before allowing a user to download the app [26]. Apple started phasing out UDID as a unique tracking identifier in 2011 [27] and MAC addresses and cookies in 2013 [28, 29] in favor of its own IDFA (ID For Advertisers) system, which allows a user to opt out of sharing their IDFA for tracking purposes [30]. Google launched App Ops for Android, which allows a user to toggle permissions by app after installation, in 2013, but removed the feature later the same year [31]. Google re-launched these features in May 2015 as part of Android M [32]. While operating system designers, Apple and Google, are moving in the direction of providing a user more controls over data sharing, mobile advertising has blossomed in recent years, growing from 5% of total digital advertising in 2010 to 37% in 2014, or $19 billion [22, 33]. One fast-growing area of mobile advertising is location-based ad targeting that requires sharing of geo-location data. Forecasts indicate that location-based ads will account for 52% of mobile ad spending by 2017 [34].

Given these changes in the app marketplace, our study examines how frequently apps share geo-location information. In addition, what other kinds of personal data are apps sharing today, and with what parties?

Background

There are three main approaches to surveying data sharing by mobile apps: permissions analysis, static code analysis and dynamic analysis.

Permissions analysis examines the permission requests from an app either before installation or during use as disclosed to the user, usually on the app's download page in the Google Play Store or Apple App Store [12, 13, 35, 36, 37]. The benefit of this approach is that it allows efficient review of thousands of apps at once. The shortcoming is that the review is only at a high level, without knowing whether the app actually collects the requested data and who receives it [43]. One study of over 22,000 Android apps found that free apps and "look-alike" apps with names similar to popular ones request more permissions [13]. There is a correlation between number of downloads and the number of permission requests. The greater the number of downloads, the more likely the app requests more permissions [13]. Barrera and Oorschot's review of 1,100 Android apps found an exponential decay in the number of apps requesting large numbers of permissions. A few apps ask for very many permissions [35].

Static code analysis studies the code of an app after decompiling to look for the permissions it requests as part of its design [14, 20, 38, 39, 40, 41, 42, 44]. This approach provides more

insight into the design of the app and remains fairly easy to automate, but its accuracy depends on the decompiler used. Also, results may be too high because they include "dead code" that never actually executes in use [43]. In one review of 114,000 apps, on average, each app had at least one ad library, which often requests permissions to access the Wi-Fi network, camera, contact list, microphone, and browser history [38]. Static analysis can also detect over-privileging in apps [14]. Beyond just looking at permissions, Egele et al. found that more than half of the 1,400 iOS apps in their sample shared a unique device ID [39]. Static analysis can also uncover potential malware and vulnerabilities, such as ad libraries directly fetching and running unexpected code from the Internet [40, 41, 44].

Dynamic analysis can capture what is actually happening when an app is used, but it requires human intervention, which makes it more difficult to scale [10, 22, 43, 45, 46, 47]. Taintdroid for Android tracks private information flows from the app to its destination [45]. In one study, it found 97 out of the 145 apps tested sent potentially private information such as phone information, device IDs, or geo-coordinates to primary or third-party servers [58]. Researchers at the University of Washington expanded on Taintdroid's functions to look for leakages of other data types such as AndroidID and to check for commonly used native functions such as MD5 hashing that may obscure the extent of data sharing [10]. Another dynamic analysis method uses a virtual private network (VPN) to monitor traffic from a device, employing tools such as Meddle or AntMonitor, which have found apps on iOS and Android that share personally identifiable information, such as name, email, and password, as plaintext [46, 47]. The 2010 WSJ study used a third method by monitoring a man-in-the-middle Wi-Fi network that a device used to connect to the Internet [22].

For this study, we focused on examining actual transmissions of personal data by apps during routine use. As our method, we selected dynamic analysis monitoring with a man-in-the-middle Wi-Fi network, as used in the WSJ report.

Other research at the FTC and Privacy Rights Clearinghouse also uses this approach [22, 48, 49]. In 2014, a researcher from the FTC's Mobile Lab conducted a runtime analysis of 15 health and fitness apps on mobile phones [48]. In total, from the apps tested, 18 third parties received device-specific identifiers, 14 received consumer-specific identifiers, and 22 received other health information [48]. Overall, the study found that 12 of the apps surveyed transmitted information to 76 different hosts, with many of the third parties receiving information from multiple apps [48]. This result supported the findings of a 2013 Privacy Rights Clearinghouse study that surveyed 43 mobile health and fitness apps and found that the biggest risks to the privacy of the personal information of users of mobile health and fitness apps resulted from apps using unencrypted connections to third-party advertisers and analytics services [49]. Our study expands on this work by studying 110 apps in health and other categories with potentially sensitive data.

Methods

Our goal was to select and use popular free apps from the Google Play Store for Android and the Apple App Store for iOS and to record the amount and kind of sensitive data transmitted from the user's device. Afterwards, we analyzed the recordings looking for different kinds of information sharing—specifically, personally identifiable information (PII), behavioral, and location information shared with a primary or third-party domain.

Selecting the apps

We began by selecting a number of popular free apps from the Google Play Store and from the Apple App Store. We focused on the Play Store and App Store, since they are the two largest mobile app stores, with four times the number of apps of the closest competitor, the Amazon Appstore [59]. Before March 2015, a developer could submit and publish any app in the Google Play Store with no human review, and as a result the Play Store is a largely non-curated facility [60]. An app appearing in the Apple App Store must pass a human review and a registration process [60]. The process requires the app to have a privacy policy and terms of use statement describing requests for and uses of personal information [61, 62]. Thus, the Apple App Store is a more curated facility. By choosing apps from these two stores, we thought our results might also reveal whether curating makes a difference in the transmission of personal information by the most popular apps.

Using the apps

To test an app, we simulated typical use for 10 to 20 minutes, sufficient to establish personal accounts with passwords, populate requests with personally identifiable information (PII), and use the basic functionalities of the app such as conducting a search, looking at a page of results, or playing one level of a game. Thus, the time spent on each app varied and depended on the nature of the app. We set all permissions to the most permissible—i.e., we allowed all requests for sharing geolocation and agreed to any other permission requests. However, we generally did not permit push notifications, which allow an app to send data in the background when not in use, such as when a different app was being tested. We wanted to avoid contaminating the data capture during each app's testing with push notifications that would cause background activity from unrelated apps to bleed through. We also deleted all apps on the tested smartphone not essential to the operating system. We tested our iOS apps on an iPhone 5 and the Android apps on a Samsung Galaxy S3.

Recording app communications

We monitored and recorded all communications between the phone and the Internet using the described man-in-the-middle approach with the free software mitmproxy [25]. The mobile phone connected to the Internet through a computer running mitmproxy. Thus, mitmproxy passed along and recorded the Internet traffic on HTTP and HTTPS going to and from the phone. Encrypted communications on HTTPS were visible as clear text in this set-up, because the mitmproxy computer records the keys necessary to decipher encrypted communications. Mitmproxy recorded two pieces of information for each flow or instance of

HTTP or HTTPS traffic to and from the phone: (1) the full site address and (2) clear text, if available, metadata about an image, javascript, or other files transmitted.

For each app, we assumed the flows that occur during app testing are likely due to that app's activity. As mentioned, we minimized background processes such as push notification for other apps as much as possible to reduce contamination. However, we could not shut off all background processes, such as those related to the phone's own operating system. Thus if Android or iOS sent traffic to the domains of Google, Apple, or others during testing, these connections might have been recorded as belonging to the specific app that was open for testing.

Analyzing the recorded app communication data

We used Python scripts to help analyze captured data. These scripts searched for transmissions in clear text for different kinds of personal data that we put into an app, such as PII and behavioral data, as well as data from the phone, such as geolocation via longitude and latitude values. Table 1 lists the kinds of personal data types that our scripts tracked and defines the categories we used. A complete list of the terms can be found in the Appendix.

When our scripts found a potential occurrence in clear text of a match to one of our inputs, we visually inspected the occurrence to determine whether the match was accurate. For example, if we input a birthday field into an app as June 1, 1980 and the script found a potential match to "06011980" in one of recorded communications for the app, we visually inspected to make sure that the match looked like part of a transmission related to a birthday rather than being part of a very large integer. One limitation of our approach is that we only had HTTP and HTTPS data, and we only looked for clear text matches based on our list of terms. Thus, if the app uses a different protocol to transmit the data or hashes data like birthday date into a less obvious string, our approach would not identify that transmission of the potentially sensitive data.

Data category	Data type	Variation
PII	Address	street address, hometown
	Birthday	birthday (month, day), birth year
	Email	
	Gender	
	Name	first name, last name
	Password	
	Phone Info	
	Phone Number	
	ZIP code	
Behavior	Employment	job searches
	Friend	name, email, phone number
	Medical Info	diseases, medications, height, weight, diet, exercise
	Post	texts, chats, likes
	Search	clothing, groceries, locations
	Username	

Data category	Data type	Variation
Location	Location	current GPS location, city

Table 1. Kinds of potentially sensitive data shared. A complete list of terms tracked for each data type and related variations can be found in the Appendix.

For analysis, we merged domains that are the same at the top two levels. For example, we combined "trafficservicecdn.telenav.com" and "logshedcn.telenav.com," which are subdomains of "telenav.com". In the case of websites that have country-specific suffixes, such as "ad-x.co.uk", we merged three levels of naming. We researched each domain in order to categorize it as either a primary domain belonging to the app-maker or as a third-party domain.

We used the statistical package R to render graphs, using bipartite graphs to show how apps connected to domains where they sent potentially sensitive data. See our data citation below to access an archived copy of raw communications captured, analyses, and scripts used.

Results

We tested 110 free apps, 55 each from the Google Play Store and the Apple App Store. We tested and recorded these apps in two waves. Wave 1 was done on June 24-26, 2014 and Wave 2 on July 15-22, 2014. During Wave 1, we chose the five most popular free apps from the Google Play Store in each of the following categories: Business, Games, Health & Fitness, and Travel & Local. In the App Store, we tested similar categories: Business, Games, Health & Fitness, and Navigation. In July 2014, we expanded our testing with Wave 2 and tested the five most popular free apps in the Play Store categories Communication, Medical, and Shopping and in the App Store categories Lifestyle, Medical, and Photo & Video. In addition, we made deeper dives—testing ten apps rather than five—in the categories Health & Fitness, Social, and Travel & Local for the Play Store and in the Health & Fitness, Navigation, and Social categories for the App Store. We chose the targeted categories in Wave 1 and 2 due to their likely handling of potentially sensitive data including job information, medical data, and location. Wave 2 did not re-test apps previously tested in Wave 1. Table 2 and 3 show the list of the apps in Android and iOS that we tested along with their wave for testing. When there was a problem testing an app, we replaced that app with the next most popular app not already tested. A complete list of all apps, including those we were unable to test is in the Appendix.

Category	App	Wave
Business	Box	2
	Facebook Pages	2
	File Manager	2
	Job Search	1
	Snagajob	1
Communication	Facebook Messenger	2
	Glide	2

Category	App	Wave
	Kik	2
	Skype	2
	Viber	2
Games	Bubble Witch 2	1
	Candy Crush	1
	Don't Tap White Tile	1
	Guess the Emoji	1
	Monster Legends	1
Health & Fitness	Fitbit	1
	iTriage Health	1
	Lose It!	2
	Map My Walk	2
	MyFitnessPal	1
	Nike+ Running	2
	Period Calendar	2
	Period Tracker	1
	RunKeeper	2
	WebMD	1
Medical	American Well	2
	Drugs.com	2
	Epocrates	2
	GoodRx	2
	Points2Shop	2
Shopping	Amazon	2
	eBay	2
	Groupon	2
	Walgreens	2
	Wish	2
Social	Emoji Android Keyboard	2
	Emoji Keyboard	2
	Facebook	2
	Instagram	2
	Pinterest	2
	Snapchat	2
	Tango	2
	Text Free	2
	textPlus	2
	Timehop	2
Travel & Local	BE-ON-ROAD	2
	Expedia	2
	GasBuddy	1
	Google Earth	1
	KAYAK	2
	MapFactor	1
	MapQuest	2
	Priceline	2
	Scout	2
	Yelp	1

Table 2. List of tested Android apps. These were the most popular apps on Google Play for Android accessed during Wave 1 (June 2014, highlighted in orange) and during Wave 2 (July 2014) in the eight categories of Business, Communication, Games, Health & Fitness, Medical, Shopping, Social, and Travel & Local. Apps appear alphabetically per category. A more thorough list of apps, including those that could not be tested, appears in the Appendix.

Category	App	Wave
Business	Adobe Reader	1
	ADP Mobile Solutions	1
	Job Search - Indeed.com	1
	Job Search - Snagajob	1
	SmartScan Express	2
Games	Fish Out of Water!	1
	Fruit Ninja	1
	Guess the Emoji	1
	Piano Tiles	1
	TwoDots	1
Health & Fitness	Fitbit	1
	Lose It!	2
	Map My Run	1
	MyFitnessPal	1
	Nike+ Running	2
	Pacer - Pedometer plus	2
	Period Tracker Lite	2
	RunKeeper	1
	The Bump Pregnancy	2
	WebMD	1
Lifestyle	Amazon	2
	eBay	2
	Groupon	2
	Walgreens	2
	Wish	2
Medical	American Well	2
	GoodRx	2
	Leafly Marijuana	2
	Ovia Fertility	2
	Urgent Care	2
Navigation	Geocaching Intro	2
	Google Maps	1
	GPS by Telenav	2
	INRIX XD	2
	Local Scope	2
	MapQuest	1
	Moovit	2
	Phone Tracker	2
	Scout GPS	1

Category	App	Wave
	Track Kit Pro	1
Photo & Video	Flipagram	2
	Instagram	2
	InstaSize	2
	Snapchat	2
	YouTube	2
Social	Emoji Keyboard 2	2
	Facebook	2
	Facebook Messenger	2
	Hangouts	2
	Kik	2
	Pinterest	2
	Skype for iPhone	2
	Tango	2
	Timehop	2
	Viber	2

Table 3. List of tested iOS apps. These were the most popular apps on the App Store for iOS accessed during Wave 1 (June 2014, highlighted in orange) and during Wave 2 (July 2014) in the eight categories of Business, Games, Health & Fitness, Lifestyle, Medical, Navigation, Photo & Video, and Social. Apps appear alphabetically per category. A more thorough list of apps, including those that could not be tested, appears in the Appendix.

Android results

Out of the 55 apps that we tested for Android, Text Free, Glide, and Map My Walk sent potentially sensitive data to the most primary and third-party domains (Figure 1). The top three domains that received potentially sensitive data from the largest number of apps are google.com, googleapis.com, and facebook.com, though that appears to be less the case for location data versus PII or behavior data (Figures 1, 2, 3, and 4). Facebook.com was also the primary domain for three of the apps tested: Facebook Messenger, Facebook Pages, and Instagram.

For PII data, Text Free, Glide, and Map My Walk again rose to the top as sending data to the most domains (Figure 2). For behavior data such as a search term input into the app, Pinterest and Drugs.com are the top two apps (Figure 3). In the case of location data, such as the user's current coordinates, Text Free and MapQuest sent the data to the most domains (Figure 4).

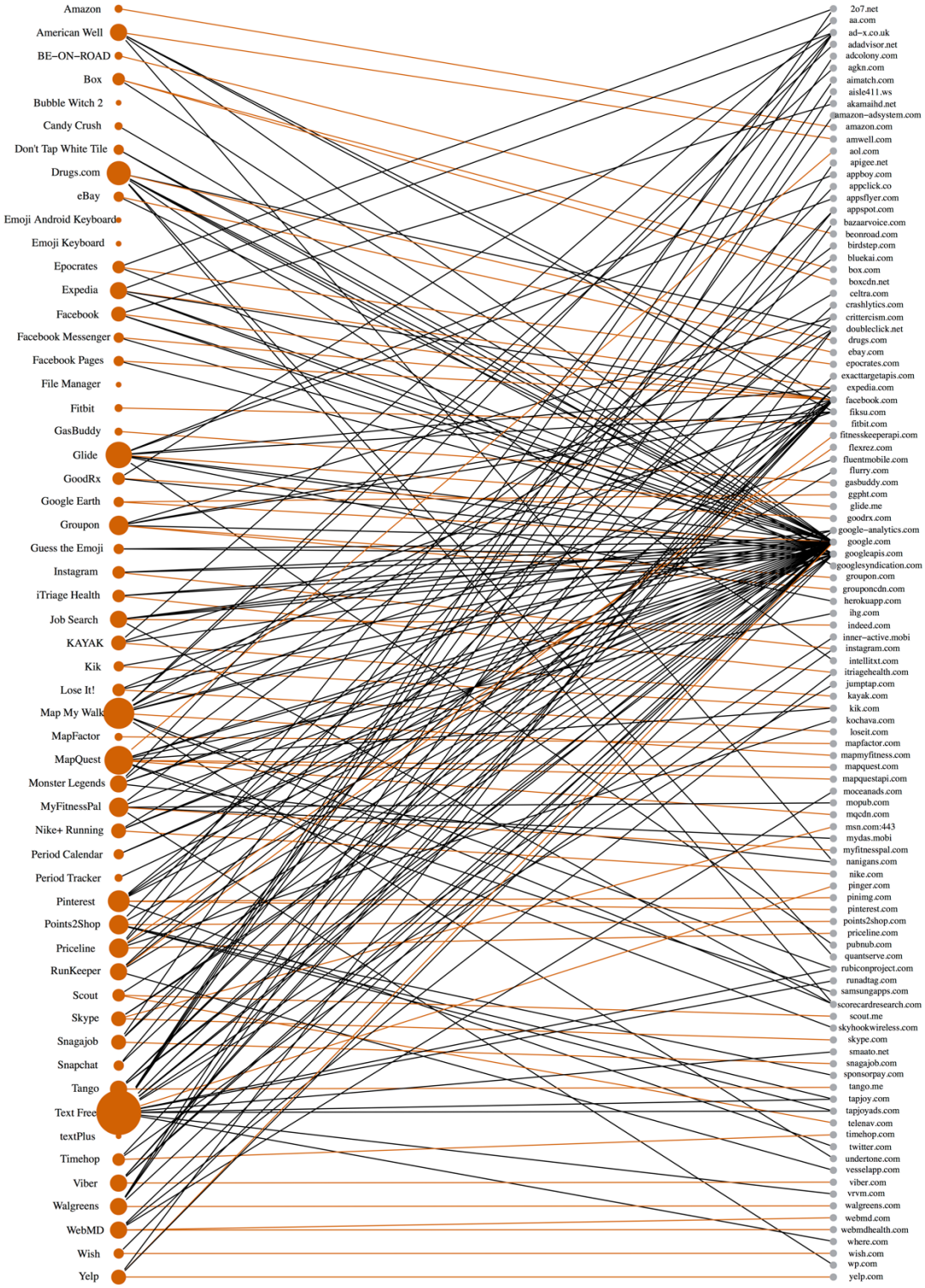


Figure 1. Sensitive data sharing by Android apps. Apps (left) connected to various domains (right. The color of the line indicates whether the domain is that of the primary maker (orange) of the app or of a third party (black). Apps with bigger circles shared sensitive data with more domains, both primary and third-party.

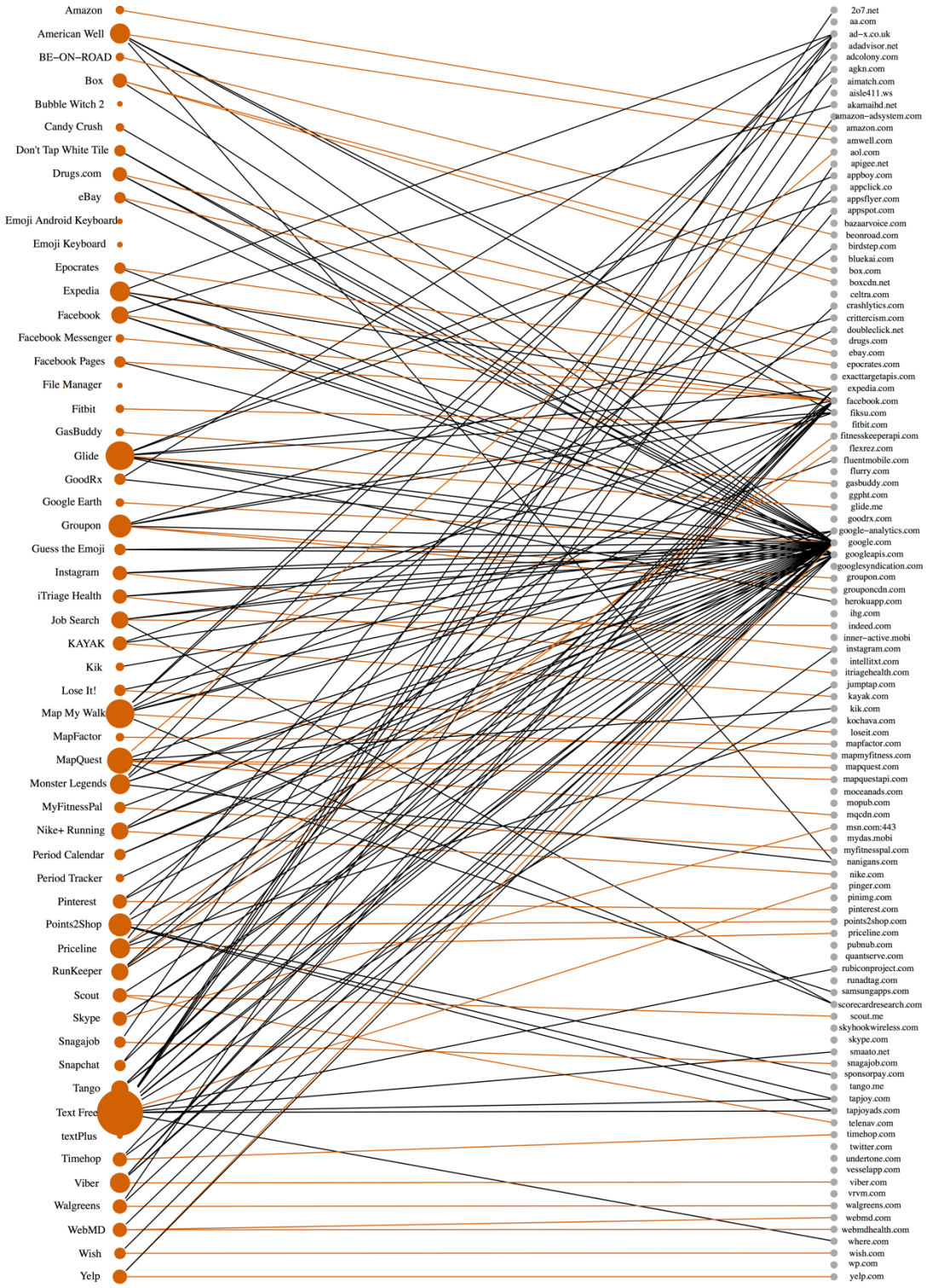


Figure 2. PII data sharing by Android apps. Apps (left) connected to various domains (right). The color of the line indicates whether the domain is that of the primary maker (orange) of the app or of a third party (black). Apps with bigger circles shared PII data with more domains, both primary and third-party.

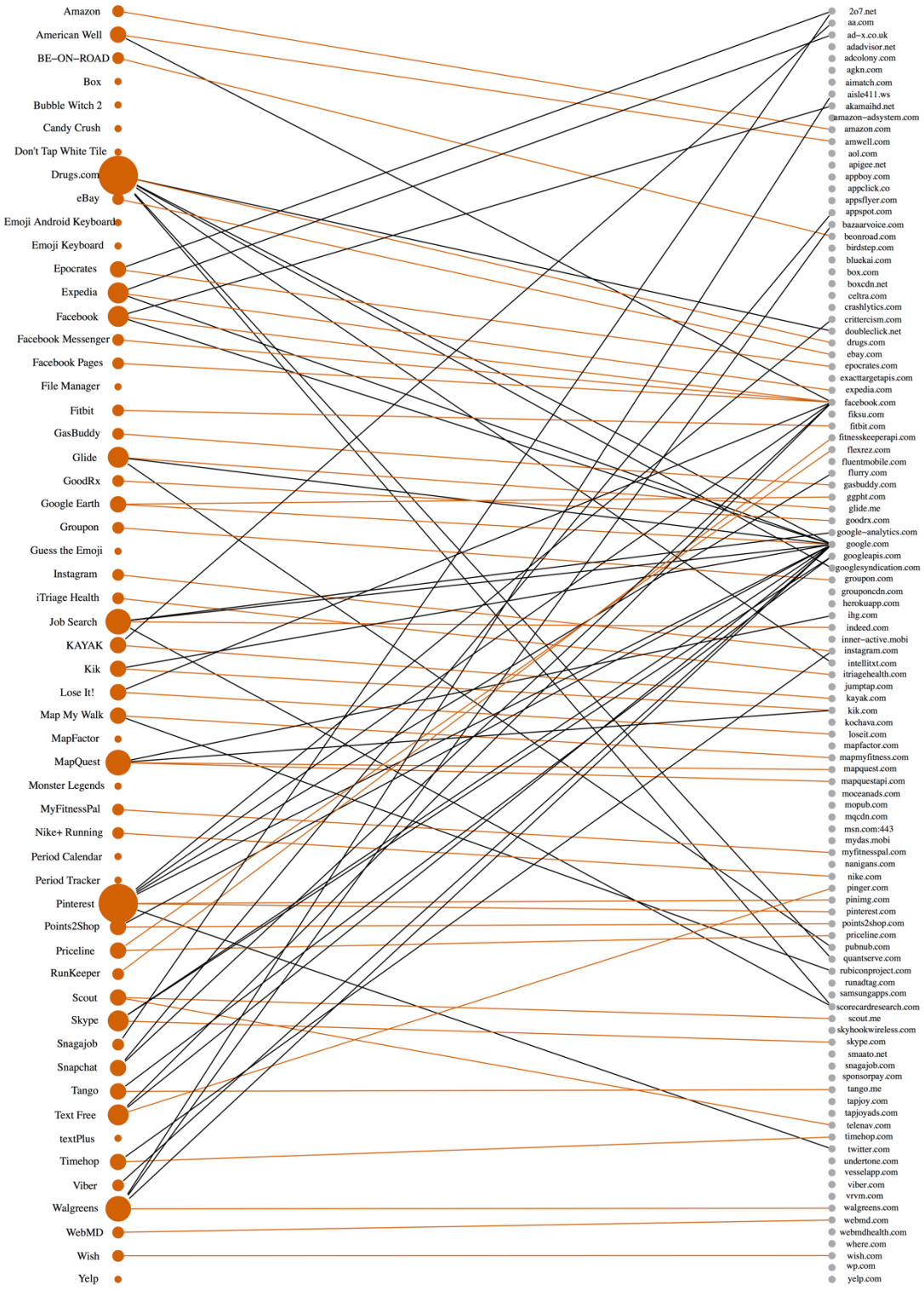


Figure 3. Behavior data sharing by Android apps. Apps (left) connected to various domains (right). The color of the line indicates whether the domain is that of the primary maker (orange) of the app or of a third party (black). Apps with bigger circles shared behavior data with more domains, both primary and third-party.

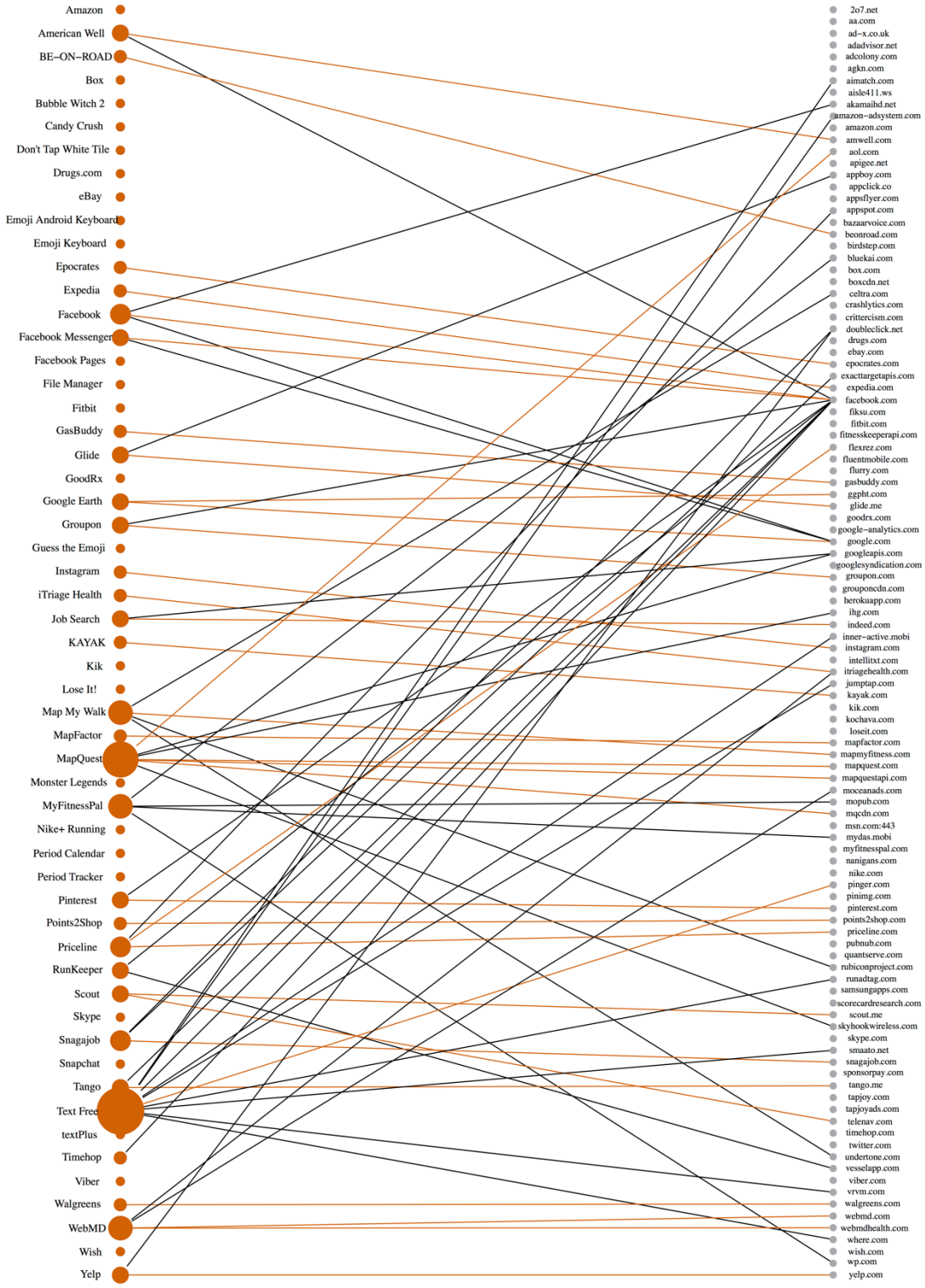


Figure 4. Location data sharing by Android apps. Apps (left) connected to various domains (right). The color of the line indicates whether the domain is that of the primary maker (orange) of the app or of a third party (black). Apps with bigger circles shared location data with more domains, both primary and third-party.

In general, there were only one or two primary domains per app that received sensitive data, but the average number of third-party domains was 3.1 (Table 4). Health & Fitness and Communication apps sent sensitive data, mostly PII data, to more third-party domains than apps in other categories. Text Free, an app listed under the Social category of the Play Store, sent sensitive data to 11 third-party domains, more than any other app, with 9 domains receiving PII data, 2 receiving behavior data, and 6 receiving location data. The apps in the sample generally sent PII data to more third-party domains than behavior or location data. Glide, Map My Walk, and Text Free are each sending PII data to 7 or more third-party domains. Many apps have no observable traffic to any third-party domains that contain behavior or location data, hence the many empty cells in Table 4 for those two columns.

Category	App	Domains receiving any sensitive data		Domains receiving PII data		Domains receiving Behavior data		Domains receiving Location data	
		Primary	Third party	Primary	Third party	Primary	Third party	Primary	Third party
Business	Box	2	1	2	1				
	Facebook Pages	1	1	1	1	1			
	File Manager								
	Job Search	1	4	1	3	1	3	1	1
	Snagajob		3		1		1		2
Communication	Facebook Messenger	1	1	1		1		1	1
	Glide	1	8	1	7	1	2	1	1
	Kik	1	1		1	1	1		
	Skype	1	2	1	2		2		
	Viber		4		4		1		
Games	Bubble Witch 2								
	Candy Crush		1		1				
	Don't Tap White Tile		2		2				
	Guess the Emoji		2		2				
	Monster Legends		5		5				
Health & Fitness	Fitbit	1		1		1			
	iTriage Health	1	2	1	2	1		1	
	Lose It!	1	2	1	1	1	1		
	Map My Walk	1	9	1	7	1	1	1	2
	MyFitnessPal	1	4	1	1	1			3
	Nike+ Running	1	3	1	3	1			
	Period Calendar		2		2				
	Period Tracker		1		1				
	RunKeeper	1	3	1	3	1			1
	WebMD		3		1				2
Medical	American Well	1	4	1	4	1	1	1	1
	Drugs.com	1	7	1	2	1	6		
	Epocrates	1	2	1	1	1	1	1	
	GoodRx	1	2		2	1			
	Points2Shop	1	2	1	2	1	1	1	
Shopping	Amazon	1		1		1			
	eBay	1	1	1	1	1			
	Groupon	2	4	2	4	1		1	1
	Walgreens		4		2		3		
	Wish		1		1				

Category	App	Domains receiving any sensitive data		Domains receiving PII data		Domains receiving Behavior data		Domains receiving Location data	
		Primary	Third party	Primary	Third party	Primary	Third party	Primary	Third party
Social	Emoji Android Keyboard								
	Emoji Keyboard								
	Facebook	1	3	1	3	1	2	1	2
	Instagram	1	2	1	2	1		1	
	Pinterest	2	4	1	2	2	4	1	1
	Snapchat		2		2		2		
	Tango		4		4		1		1
	Text Free	1	11	1	9	1	2	1	6
	textPlus								
	Timehop		2		2		1		1
Travel & Local	BE-ON-ROAD	1		1		1		1	
	Expedia	1	4	1	4	1	2	1	
	GasBuddy	1		1		1		1	
	Google Earth	2		1		2		2	
	KAYAK	1	3	1	2	1	1	1	
	MapFactor	1		1				1	
	MapQuest	4	5	4	3	2	2	4	2
	Priceline	2	4	2	3	2		2	1
	Scout		1		1				
	Yelp	1	2	1	1				1

Table 4. Distribution of domains receiving any sensitive data for Android apps tested. Empty cells indicate no observed data of that type was sent to a primary or third-party domain by the app.

iOS results

Out of the 55 apps that we tested for iOS, Local Scope sent potentially sensitive data to the most primary and third-party domains (Figure 5). The top three domains that received potentially sensitive data from the most apps are apple.com, yahooapis.com, and exacttargetapis.com, especially for location data versus PII or behavior data (Figures 5, 6, 7, and 8).

For PII data, Pinterest, Map My Run, MapQuest, Piano Tiles, and Timehop rose to the top as sending data to the most domains (Figure 6). For behavior data such as search terms input into the app, Local Scope is the app sending data to the most domains (Figure 7). For location data such as the user's current GPS coordinates, Local Scope again sent that data to the most domains (Figure 8).

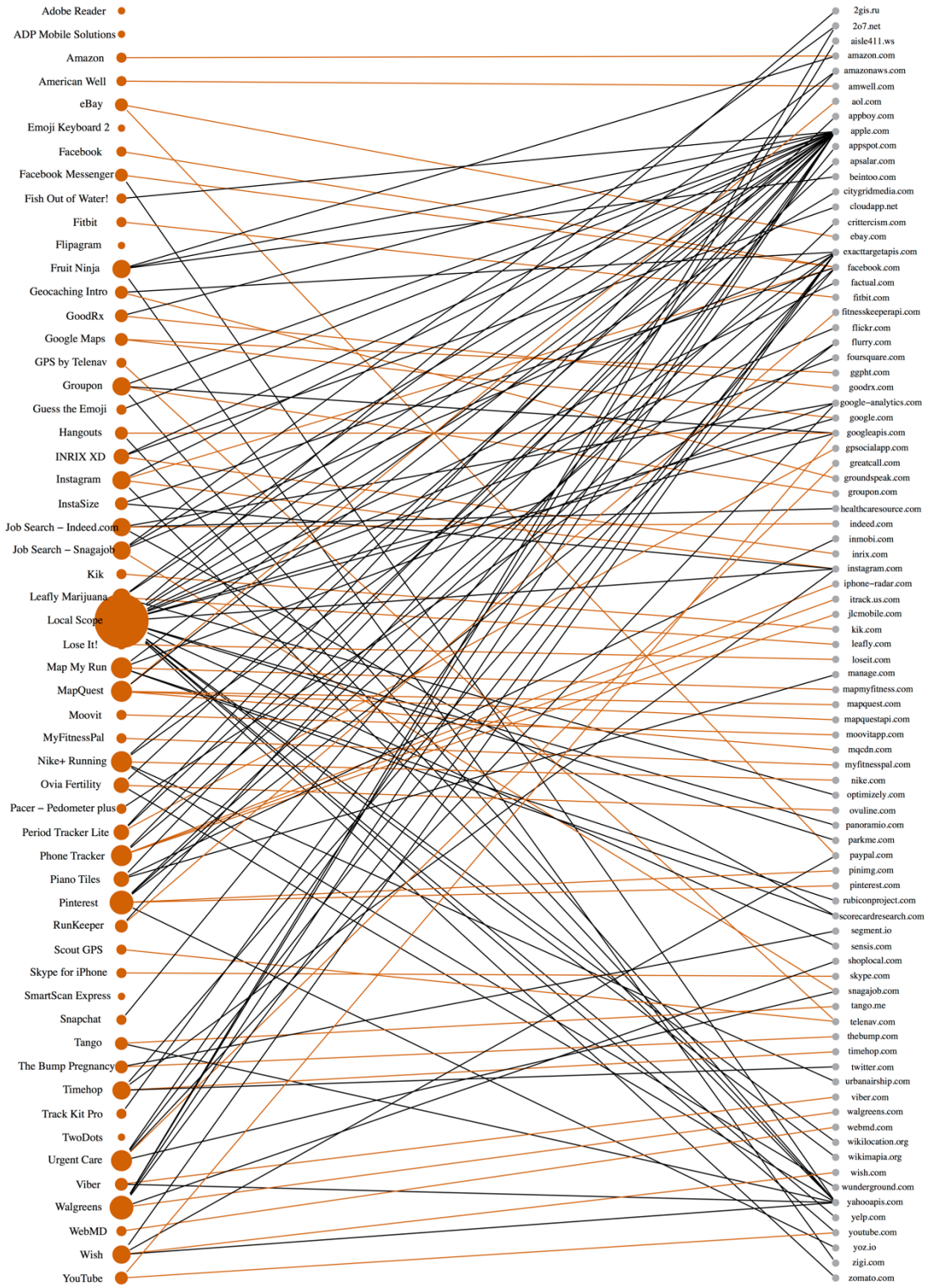


Figure 5. Sensitive data sharing by iOS apps. The color of the line indicates whether the domain is that of the primary maker (orange) of the app or of a third party (black). Apps with bigger circles shared sensitive data with more domains, both primary and third-party.

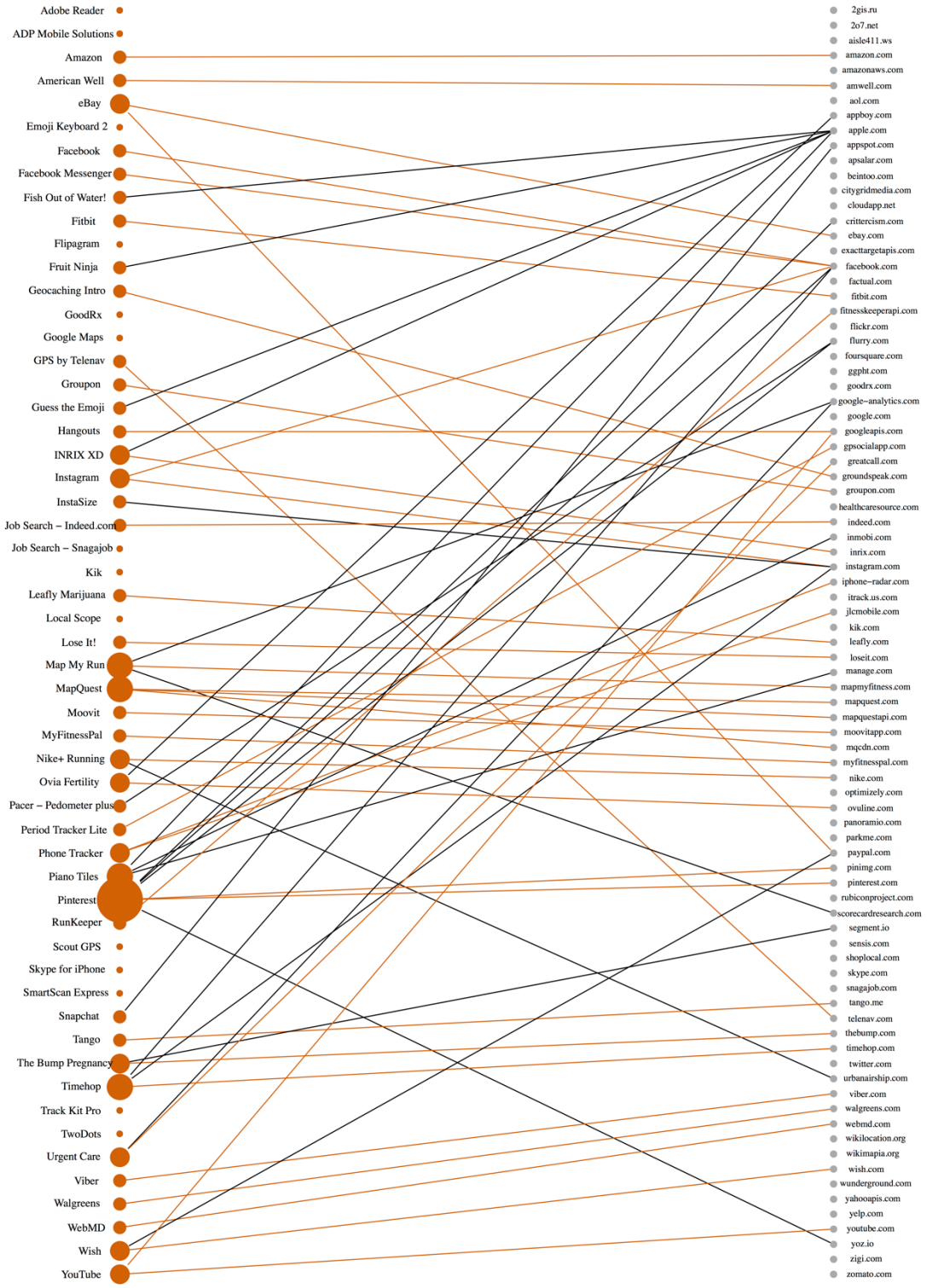


Figure 6. PII data sharing by iOS apps. The color of the line indicates whether the domain is that of the primary maker (orange) of the app or of a third party (black). Apps with bigger circles shared PII data with more domains, both primary and third-party.

Figure 7. Behavior data sharing by iOS apps. The color of the line indicates whether the domain is that of the primary maker (orange) of the app or of a third party (black). Apps with bigger circles shared behavior data with more domains, both primary and third-party.

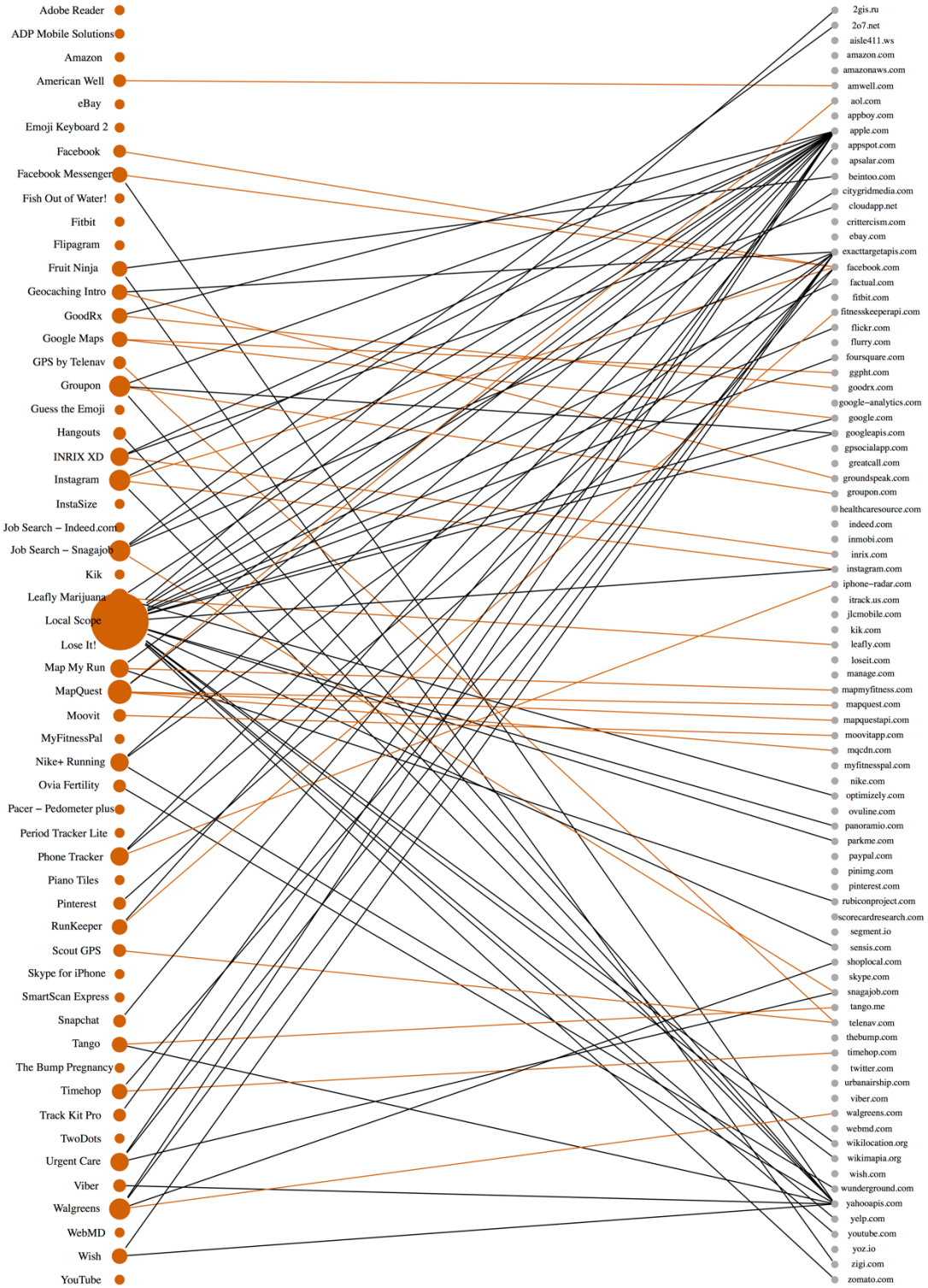


Figure 8. Location data sharing by iOS apps. The color of the line indicates whether the domain is that of the primary maker (orange) of the app or of a third party (black). Apps with bigger circles shared location data with more domains, both primary and third-party.

Much like Android apps, iOS apps usually send sensitive data just one or two primary domains, but on average to 2.6 third-party domains (Table 5). Every category had a mix of apps that sent sensitive data to third-party domains and apps that did not. Local Scope, an app listed under the Navigation category of the App Store, sent sensitive data to 17 third-party domains, more than any other app, with 15 domains receiving behavior data, and 17 receiving location data. Piano Tiles and Pinterest both sent PII data to at least 3 third-party domains. Job Search – Indeed.com and Local Scope sent behavior data to at least 3 third-party domains. Job Search – Snagajob, Nike+ Running, Groupon, Walgreens, Urgent Care, Local Scope, and Phone Tracker sent location data to at least 3 third-party domains.

Category	App	Domains receiving any sensitive data		Domains receiving PII data		Domains receiving Behavior data		Domains receiving Location data	
		Primary	Third party	Primary	Third party	Primary	Third party	Primary	Third party
Business	Adobe Reader								
	ADP Mobile Solutions								
	Job Search - Indeed.com	1	3	1		1	3		
	Job Search - Snagajob	1	3			1	2	1	3
	SmartScan Express								
Games	Fish Out of Water!		1		1		1		
	Fruit Ninja		4		1		2		2
	Guess the Emoji		1		1		1		
	Piano Tiles		3		3		1		
	TwoDots								
Health & Fitness	Fitbit	1		1		1			
	Lose It!	1		1		1			
	Map My Run	1	4	1	2	1		1	2
	MyFitnessPal	1		1		1			
	Nike+ Running	1	4	1	1	1			3
	Pacer - Pedometer plus		1		1				
	Period Tracker Lite	1	2	1			2		
	RunKeeper	1	1	1		1		1	1
	The Bump Pregnancy	1	1	1	1	1	1		
	WebMD	1		1		1			
Lifestyle	Amazon	1		1		1			
	eBay	2		2		1			
	Groupon	1	3	1		1		1	3
	Walgreens	1	5	1		1	2	1	3
	Wish	1	3	1	1	1			2
Medical	American Well	1		1		1		1	
	GoodRx	1	1			1	1	1	1
	Leafly Marijuana	1	3	1		1	1	1	2
	Ovia Fertility	1	2	1	1	1			1
	Urgent Care	1	4	1	1	1	1		3
Navigation	Geocaching Intro	1	1	1		1		1	1
	Google Maps	2				1		2	
	GPS by Telenav	1		1		1		1	
	INRIX XD	1	2	1	1	1	2	1	2
	Local Scope		17				15		17

Category	App	Domains receiving any sensitive data		Domains receiving PII data		Domains receiving Behavior data		Domains receiving Location data	
		Primary	Third party	Primary	Third party	Primary	Third party	Primary	Third party
	MapQuest	4	1	3		4		4	1
	Moovit	1		1		1		1	
	Phone Tracker	3	2	2		1		1	2
	Scout GPS	1						1	
	Track Kit Pro		1						1
Photo & Video	Flipagram								
	Instagram	2	2	2		2		2	2
	InstaSize		2		1		2		
	Snapchat		1		1		1		1
	YouTube	2		2		2			
Social	Emoji Keyboard 2								
	Facebook	1		1		1		1	
	Facebook Messenger	1	1	1		1		1	1
	Hangouts	1	1	1		1			1
	Kik	1				1			
	Pinterest	2	4	2	4	2	1		1
	Skype for iPhone	1				1			
	Tango	1	1	1				1	1
	Timehop	1	3	1	2	1	2	1	1
	Viber	1	1	1					1

Table 5. Distribution of domains receiving any sensitive data for iOS apps tested. Empty cells indicate no observed data of that type was sent to a primary or third-party domain by the app.

Potentially sensitive data types shared with third-party domains

For Android apps, the most common data type shared with a third-party domain is a user's email address, which is PII data, with 73% of the Android apps transmitting that data (Table 6 and 8). Other commonly shared data types in Android include name (49% of apps), address (25% of apps), and phone information such as IMEI number (24% of apps) for the PII data category, username (25% of apps) for the behavior data category, and location data such as the user's current GPS coordinates (33% of apps) (Table 8).

Less commonly shared data types may still be potentially sensitive data. For example, the Drugs.com app shared medical info input by the user in testing—including words such as "herpes" or "interferon"—with 5 third-party domains: doubleclick.net, googlesyndication.com, intellitxt.com, quantserve.com, and scorecardresearch.com. None of the 5 domains directly received any PII from the app, though google.com and googleapis.com did receive names and email addresses while the app ran. For a different type of potentially sensitive behavior data, the Business category apps, Job Search and Snagajob, shared employment-related search terms such as "driver," "cashier," and "burger" with third-party domains google.com, google-analytics.com, scorecardresearch.com, and 2o7.net during testing. One of the domains in the Job Search app, google.com, also received PII data

including the user's email address. The third-party domains that received passwords from apps include crashlytics.com for RunKeeper, appspot.com for Snapchat, and instagram.com for Timehop.

Finally, some apps sent to the same third-party domain potentially sensitive combinations of data such as name and current GPS location. Facebook.com connected with 7 apps, American Well, Groupon, Pinterest, RunKeeper, Tango, Text Free, and Timehop, to access this data combination. Appboy.com received this data on the Glide app.

		PII									Behavior						Location
Category	App	Address	Birthday	Email	Gender	Name	Password	Phone info	Phone Number	ZIP code	Employ-ment	Friend	Medical info	Post	Search	Username	Location
Business	Box			1													
	Facebook Pages			1													
	File Manager																
	Job Search	3		1							3					1	1
	Snagajob	1									1						2
Communication	Facebook Messenger																1
	Glide	1		4	1	3		2	1			1		1		1	1
	Kik			1		1										1	
	Skype			2		1						2				1	
	Viber		1	2	1	3										1	
Games	Bubble Witch 2																
	Candy Crush			1													
	Don't Tap White Tile			1		1											
	Guess the Emoji			2	1	2											
	Monster Legends			1	1	1		3									
Health & Fitness	Fitbit																
	iTriage Health			2													
	Lose It!			1								1					
	Map My Walk			1	3	2		1							1		3
	MyFitnessPal			1													4
	Nike+ Running	2		2		3											
	Period Calendar			1		1											
	Period Tracker			1													
	RunKeeper	1		2		2	1										2
	WebMD			1													2
Medical	American Well			1		2		2				1					1
	Drugs.com	1		2		1							5			1	
	Epocrates			1										1			
	GoodRx	1		1		1		1									
	Points2Shop			1				4								1	
Shopping	Amazon																
	eBay			1		1											
	Groupon	3	1	3	1	3		1	1								1
	Walgreens			1		1		1				1			2		
	Wish			1		1											
Social	Emoji Android Keyboard																
	Emoji Keyboard																
	Facebook	1		2		1									1	1	2
	Instagram	1		2		1											
	Pinterest	1	1	1	1	2						3		1		2	1
	Snapchat		1	1		1	1		1			1				1	
	Tango	1		2		3						1					1
	Text Free			1	7	2		5		4		1				1	9
	textPlus																
	Timehop	1	1		1	1	1		1							1	1
Travel & Local	BE-ON-ROAD																
	Expedia	3		2	1	2		1							2	1	
	GasBuddy																
	Google Earth																
	KAYAK			1				1						1			
	MapFactor																
	MapQuest			2				1							1	1	3
	Priceline			1				2									1
	Scout			1													
	Yelp				1	1											1

Table 6. Categories of sensitive data (columns) shared to third-party domains by Android apps (rows). Cells shaded orange indicate that at least one third-party domain received data of that category while the selected app ran. The values inside orange cells show specifically how many third-party domains received the data.

For iOS apps, the most common data type shared with a third-party domain was a user's current location and GPS coordinates, with 47% of the apps transmitting that data (Table 7 and 8). Other commonly shared data types in iOS include name (18% of apps) and email

address (16% of apps) in the PII data category (Table 8). 4 out of the 5 Game apps tested transferred name and email data to the domain apple.com, specifically to Apple's Game Center site at service.gc.apple.com. Pinterest, a Social category app, sent names to 4 third-party domains, yoz.io, facebook.com, crittercism.com, and flurry.com. The third-party domains that received passwords from apps include instagram.com for Timehop for InstaSize and appspot.com for SnapChat.

A few different apps shared potentially sensitive behavior data from user inputs and searches with third-party domains. For example, Period Tracker Lite shared an input into a symptom field of "insomnia" with apsalar.com. In the Business category, the two Job Search apps from Indeed.com and Snagajob shared employment-related inputs such as "Nurse" and "Car mechanic" with 4 third-party domains, 207.net, healthcareresource.com, google-analytics.com, and scorecardresearch.com.

Finally, compared to Android, fewer of the tested iOS apps sent the same third-party domain potentially sensitive combinations of data such as name and current GPS location. Facebook.com connected with 2 apps, Pinterest and Timehop, to access this data combination.

		PII								Behavior						Location
Category	App	Address	Birthday	Email	Gender	Name	Password	Phone Number	Zipcode	Employment	Friend	Medical Info	Post	Search	Username	Location
Business	Adobe Reader															
	ADP Mobile Solutions															
	Job Search - Indeed.com									3				2		
	Job Search - Snagajob									1				2		3
	SmartScan Express															
Games	Fish Out of Water!			1		1									1	
	Fruit Ninja			1		1									1	2
	Guess the Emoji			1		1									1	
	Piano Tiles			1	2	1									1	
	TwoDots															
Health & Fitness	Fitbit															
	Lose It!															
	Map My Run					2										2
	MyFitnessPal															
	Nike+ Running				1											3
	Pacer - Pedometer plus				1											
	Period Tracker Lite											2				
	RunKeeper															1
	The Bump Pregnancy			1											1	
	WebMD															
Lifestyle	Amazon															
	eBay															
	Groupon															3
	Walgreens												2			3
	Wish					1										2
Medical	American Well															
	GoodRx													1		1
	Leafly Marijuana													1		2
	Ovia Fertility			1		1										1
	Urgent Care				1							1				3
Navigation	Geocaching Intro															1
	Google Maps															
	GPS by Telenav															
	INRIX XD	1												2		2
	Local Scope													15		17
	MapQuest															1
	Moovit															
	Phone Tracker															2
	Scout GPS															
	Track Kit Pro															1
Photo & Video	Flipagram															
	Instagram															2
	InstaSize					1	1				1		1		2	
	Snapchat		1	1			1	1			1		1		1	1
	YouTube															
Social	Emoji Keyboard 2															
	Facebook															
	Facebook Messenger															1
	Hangouts															1
	Kik															
	Pinterest		1	2	1	4					1					1
	Skype for iPhone															
	Tango															1
	Timehop		1	1		1	1								2	1
	Viber															1

Table 7. Categories of sensitive data (columns) shared to third-party domains by iOS apps (rows). Cells shaded orange indicate that at least one third-party domain received data of that category while the selected app ran. The values inside orange cells show specifically how many third-party domains received the data.

Compared to Android apps, fewer iOS apps shared PII and behavior data with third-party domains. In some data types, the contrast is significant, with 73% of Android apps transmitting email addresses versus 16% of iOS apps. In addition, 49% of Android apps transferred either first or last name, compared to 18% of iOS apps. On the other hand, more iOS apps (47%) than Android apps (33%) transmitted current location data, including GPS coordinates, to a third-party domain. In terms of all 110 apps tested across both operating systems, the top three data types most commonly shared were email (45% of apps), location (40% of apps), and name (34% of apps).

Data category	Data type	All apps		Android		iOS	
		# of apps	%	# of apps	%	# of apps	%
PII	Address	15	14%	14	25%	1	2%
	Birthday	8	7%	5	9%	3	5%
	Email	49	45%	40	73%	9	16%
	Gender	16	15%	11	20%	5	9%
	Name	37	34%	27	49%	10	18%
	Password	6	5%	3	5%	3	5%
	Phone Info (Android only)	13	24%	13	24%	N/A	
	Phone Number	5	5%	4	7%	1	2%
	ZIP code	1	1%	1	2%	0	0%
Behavior	Employment	4	4%	2	4%	2	4%
	Friend	12	11%	9	16%	3	5%
	Medical Info	3	3%	1	2%	2	4%
	Post	7	6%	4	7%	3	5%
	Search	11	10%	5	9%	6	11%
	Username	22	20%	14	25%	8	15%
Location	Location	44	40%	18	33%	26	47%

Table 8. Summary of the number of apps in Android and iOS sharing data with third-party domains by data type. We looked for phone info for Android apps only.

Third-party domains that received potentially sensitive data from the most apps

Table 9 shows the 13 third-party domains that received potentially sensitive data from at least 4 of the Android or iOS apps that we tested. The top 6 third-party domains provided API functions for the app that allowed the app to access code libraries and datasets provided by Google, Apple, Facebook, ExactTarget, and Yahoo. Apps mostly shared PII and behavior data with Google.com and Googleapis.com in Android, while apple.com received location data in iOS. Since we were not able to disable all background processes ran by the Android or iOS operating systems during, some of the observed data transmissions to Google or Apple domains may have been due to unrelated background processes. No single analytics or advertising third-party domain dominated in receiving potentially sensitive data across a large number of the apps in the sample. The most popular analytics domain, google-analytics.com, and the most popular advertising domain, scorecardresearch.com, received data from only 5% of the apps tested. We found 94 distinct third-party domains that received at least one instance of potentially sensitive data from one of the 110 apps tested.

Domain	Function	All apps (out of 110)	Android apps (out of 55)				iOS apps (out of 55)			
		Any data	Any data	PII	Behavior	Location	Any data	PII	Behavior	Location
google.com	API	36%	39	38	12	2	1	0	1	1
googleapis.com	API	18%	18	16	2	2	2	0	1	2
apple.com	API	17%	0	0	0	0	19	5	7	15

Domain	Function	All apps (out of 110)	Android apps (out of 55)				iOS apps (out of 55)			
		Any data	Any data	PII	Behavior	Location	Any data	PII	Behavior	Location
facebook.com	API	14%	12	11	5	7	3	2	2	2
exacttargetapis.com	API	7%	1	0	0	1	7	0	0	7
yahooapis.com	API	7%	0	0	0	0	8	0	0	8
google-analytics.com	Analytics	5%	3	3	1	0	3	2	2	0
ad-x.co.uk	Analytics	5%	5	5	1	0	0	0	0	0
scorecardresearch.com	Advertising	5%	3	2	2	0	2	1	1	0
2o7.net	Analytics	4%	2	1	2	0	2	0	2	1
doubleclick.net	Advertising	4%	4	1	1	2	0	0	0	0
fiksu.com	Advertising	4%	4	4	0	0	0	0	0	0
instagram.com	API	4%	1	1	1	0	3	2	3	1

Table 9. Top 13 third-party domains that received any sensitive data from the apps tested. The top 13 domains received sensitive data from at least 4 apps in the sample. The table categorizes each domain by its primary function for its API, analytics, or advertising-related capabilities.

One third-party domain not included the tables and figures presented is safemovedm.com, which was connected to by 51 or 93% of the Android apps tested. The purpose of this domain connection is unclear at this time; however, its ubiquity is curious. When we used the phone without running any app, connections to this domain continued. It may be a background connection being made by the Android operating system; thus we excluded it from the tables and figures in order to avoid mis-attributing this connection to the apps we tested. The relative emptiness of the information flows sent to safemovedm.com indicate the possibility of communication via other ports outside of HTTP not captured by mitmproxy. These other ports—which may be monitored by sniffers such as Wireshark—may be of future interest in a subsequent mobile app security study.

```
---START REQUEST-----------------
('_flow_', 79)
('_scheme_', 'http')
('_host_', '54.230.38.136')
('_port_', 80)
('_path_', '/webping-s.html?unused=1405960213165')
('_method_', 'GET')
_assemble_
GET /webping-s.html?unused=1405960213165 HTTP/1.1
Cache-Control: no-cache
User-Agent: Dalvik/1.6.0 (Linux; U; Android 4.4.2; SPH-L710
Build/KOT49H)
Host: hla2.safemovedm.com
Accept-Encoding: gzip


----END REQUEST--------------------
```

Figure 9. Flow of information from mitmproxy for connections to safemovedm.com. Since mitmproxy only examines HTTP and HTTPS traffic, it may be possible that other tools such as Wireshark might be used in future studies to monitor FTP and other types of traffic to safemovedm.com.

Discussion

We found that many mobile apps transmitted potentially sensitive user data to third-party domains, especially a user's current location, email, and name. In general, iOS apps were less likely to share sensitive data of nearly every type with third-party domains than were Android apps, except for location data (Table 8). One reason might be the App Store human curation process that checks to see if apps only ask for personal information for app-related purposes [61, 62]. Collecting location data, including GPS coordinates, requires an app to request the permission of the user, which would occur before installation on the app download page for Android and as a pop-up notification during use for iOS [64, 64]. Thus, receiving location data requires user approval of a more prominent notification for iOS, and we saw more iOS apps (47%) sending location data to third parties than Android apps (33%). Our results for each operating system were in line with other studies [22, 36, 58]. In contrast, we found significantly less sharing of behavioral data, such as search terms from Medical and Health & Fitness apps, compared to previous research on data-sharing on healthcare websites. A 2015 study of more than 80,000 healthcare webpages found that on 70% of the pages, third parties can learn about the specific "conditions, treatments, and diseases" viewed [65, 66]. In our study, only 3 apps out of 30 Medical and Health & Fitness apps sent medical info, including search terms, to a third party (Table 6, 7).

The average Android app sent sensitive data to 3.1 third-party domains, and the average iOS app connected to 2.6 third-party domains. The top domains that received sensitive data from the most apps belonged to Google and Apple (Table 9). Other studies have found a similar

dominance by Google [10, 46, 58]. One factor may be the mobile ad networks and services operated by Google with AdMob, DoubleClick, and Google Analytics [68], and by Apple with iAds [69]. It is also possible that system processes that we were unable to turn off on Android and iOS sent data to the two companies' domains in the background while we tested our apps. Besides Google and Apple, no other third-party domain in our study received data from more than 14% of the apps tested. By contrast, the reach of third-party advertisers on websites is very extensive, with the top 12 ad networks all reaching more than 50% of American Internet users [67].

Implications for technology design and policy

The results of this study point out that the current permissions systems on iOS and Android are limited in how comprehensively they inform users about the degree of data sharing that occurs. Apps on Android and iOS today do not need to have permission request notifications (Figure 10, 11) for user inputs like PII and behavioral data. Three options are under current development by researchers, regulators, and companies for users who may want to more comprehensively protect their privacy while using mobile apps. These are: (1) send false data in response to app requests, (2) allow users to opt out of data collection, and (3) design app stores to prominently inform users about third parties who may receive their data.

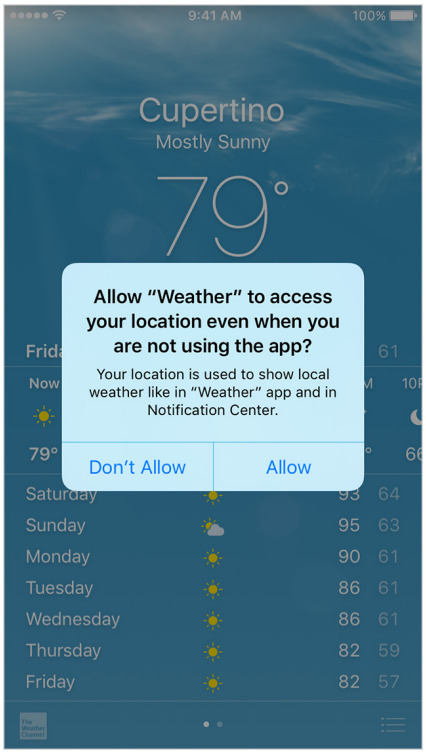


Figure 10. iOS permission request notification for location data [94]. iOS does not require apps to have notifications like this for PII or behavior data.

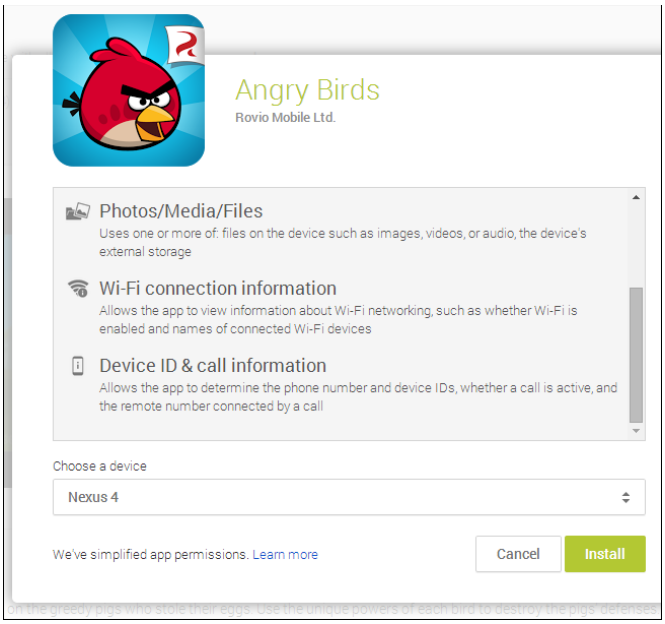


Figure 11. Android permission request notification for location data [95]. Android does not require apps to have notifications like this for PII or behavior data.

Researchers have designed tools that can protect user privacy by sending false data to satisfy permission requests from apps. MockDroid, TISSA, and AppFence are three examples that send fake information back to the app if it makes certain API calls [70, 71, 80]. It may be possible to modify these tools to send fake user data inputs as well when the recipient is a third-party domain, though that may also impact the experience of the app for targeted advertising and other functions that depend on accurate user data.

Another option is to provide mobile users an opt-out option to limit tracking and sending of user data to third parties. In recent years, web browsers, with the support of the Federal Trade Commission and the White House, implemented "Do Not Track" settings that signal to sites that the user is choosing to limit data collection of their search and browsing patterns [72]. However, this signal is voluntary with no set standard on how websites should respond [72]. Since 2014, California started requiring each website to describe in its privacy policy how the site will respond to a "Do Not Track" signal from a user's browser, though sites may still choose to ignore the signal [72]. In 2013, the FTC stated in a non-binding report that mobile apps should include a do-not-track feature to safeguard personal information [73]. No federal or state legislation mandates compliance with a do-not-track signal.

Despite the lack of formal regulation, Google and Apple in recent years implemented tracking prevention settings to a degree on their mobile operating systems. In September 2012, Apple launched a "Limit Ad Tracking" feature as part of iOS 6 that blocks ad networks from collecting their IDFA, a unique device ID [74]. By April 2014, Apple stated that it may remove or deny apps that don't respect the "Limit Ad Tracking" setting [75]. Following Apple's lead,

Google implemented a similar "Opt out of interest-based ads" setting in Android KitKat in October 2013 [76]. However, as Google notes, this setting will not stop interest-based ads not served by Google or not part of the Google Display Network [77]. Also, even if a user opts out of interest-based ads, an app may still track user activity for "contextual advertising, frequency capping, conversion tracking, reporting, security and fraud detection" [78]. Interestingly, as gatekeepers of the operating system, companies such as Apple and Google moved faster than the regulators in providing and enforcing an opt-out option to tracking on mobile apps to consumers.

Finally, app stores can show the degree of third-party data sharing more prominently on their app download page to inform users before they install the app. Many apps may describe the degree of data collection and sharing with third parties in their privacy policies, which research has found to be confusing, dense, misunderstood, and often ignored by consumers [79, 81]. App stores may choose to feature this information more noticeably, using notices similar to what exists for children's apps today. Apps meant to be used for children, such as learning or game apps, are under more scrutiny by regulators as a result of laws such as the federal Children's Online Privacy Protection Act (COPPA), enforced by the FTC to control the amount of geolocation data, photos, videos, audio recordings, and persistent identifiers collected and shared by apps without parental consent [82]. In California, S.B. 568 gives minors the right to an "Eraser Button" that will remove any content or information they submitted to websites or apps [83]. Beyond regulations, civil society groups such as Moms With Apps have signed on more than 300 app developers to practice best practices by disclosing the data collection and sharing activities of their apps [84]. Mom With Apps even built its own app store, which allows parents to filter apps by requirements such as "Works without internet," "No in-app purchases," "No links to social networks," and "No advertising" [85]. In September 2013, Apple launched Kids App Store, which includes apps for children that comply with COPPA restrictions and limit advertising [86]. Google followed in April 2015 with the launch of its "Designed For Families" program for Android apps [87]. In conclusion, app stores could possibly adapt the current designs for children's apps more broadly to apply to all apps by clearly describing the degree of third-party data sharing by an app before it is downloaded.

Future work

To expand on the results of this study, future research can focus on improving the accuracy of the Internet traffic captured for each app, testing more apps under different conditions, and reviewing whether privacy policies reflect the data collection and sharing activities recorded for each app.

We can improve the app testing process by looking at non-TCP traffic, leakage through simple hashing like MD5, and less contaminated transmissions without background system processes. The Privacy Rights Clearinghouse study of 43 health apps used a man-in-the-middle proxy like ours to monitor all HTTP and HTTPS traffic, and it also had WireShark and

tcpdump to monitor all packet-level traffic that is not on TCP [49]. Therefore, a future study could also incorporate WireShark and other tools to examine non-TCP traffic for data leakage (we note that one study using a VPN method to capture all device traffic found that over 90% of traffic volume from apps is on TCP [46]). We might also look for potential leakage of simply encrypted versions of sensitive user data sent not as clear text but by using common hashes such as MD5 that may be vulnerable to attack [88]. One 2011 study found that multiple apps sent AndroidIDs simply hashed with MD5 as plaintext to different ad networks including Google's DoubleClick and AdMob [10]. There was no "salting" or extending the identifier with new data to make it more difficult to decrypt by those who may have intercepted the plaintext traffic [10]. Finally, future work may be able to use modified tools such as Taintdroid to monitor both the operating system and the app so as to better distinguish between leakage that occurred as a result of app activity versus a background system process [45].

Future research might also expand the scope of our testing with more apps under different conditions. We tested apps in only 9 categories in the Play Store and App Store. As of September 2015, the Play Store had 27 categories, and the App Store had 23 categories [89, 90], so there are many more categories to examine. We could also test paid apps to see whether their data sharing patterns differ from those of free apps. Beyond just testing more apps, we can re-test the apps in our sample to track changes in their data sharing over time. Finally, we could test apps under the condition of turning on Android's "opt out of interest-based ads" and iOS' "Limit Ad Tracking" settings to see if we observe a difference in app activity.

Finally, we could review the privacy policies and terms of use for each tested app to see if the policy matches the practice. The FTC has fined companies, such as Path and Goldenshore Technologies, for deception for collecting and sharing data in their mobile apps despite the claims in their privacy policies [8, 91, 92]. Future work may provide other examples of apps whose practices contradict their privacy policies.

References

1. Reisinger D. Android by the Numbers: 1B Monthly Active Users. CNET. June 25, 2014. http://www.cnet.com/news/android-by-the-numbers-1-billion-monthly-active-users/

2. Bajarin B. iOS, Android, and the Dividing of Business Models. Tech Opinions. June 30, 2014. http://techpinions.com/ios-android-and-the-dividing-of-business-models/32237

3. Savage C. Between the Lines of the Cellphone Privacy Ruling. The New York Times, June 25, 2014. http://www.nytimes.com/interactive/2014/06/25/us/annotated-supreme-court-cellphone-privacy-decision.html

4. Boyles J, Smith A, Madden M. Privacy and Data Management on Mobile Devices. Pew Research Center. September 5, 2012. http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_MobilePrivacyManagement.pdf

5. Urban J, Hoofnagle C, Li S. Mobiles Phones and Privacy. BCLT Research Paper Series. July 10, 2012. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2103405

6. Felt A, Egelman S, Wagner D. I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices. October 19, 2012. http://dl.acm.org/citation.cfm?id=2381943

7. Office of the Attorney General. Attorney General Kamala D. Harris Issues Guide on Privacy Policies and Do Not Track Disclosures. State of California Department of Justice. May 21, 2014. https://oag.ca.gov/news/press-releases/attorney-general-kamala-d-harris-issues-guide-privacy-policies-and-do-not-track

8. Federal Trade Commission. In the Matter of Goldenshore Technologies, LLC. and Erik M. Geidl, Decision and Order. March 31, 2014. https://www.ftc.gov/system/files/documents/cases/140409goldenshoresdo.pdf

9. Article 29 Data Protection Working Party. Opinion 02/2013 on apps on smart devices. European Commission. February 27, 2013. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp202_en.pdf

10. Han S, Jung J, Wetheral D, A Study of Third-Party Tracking by Mobile Apps in the Wild. University of Washington Technical Report. March 1, 2012. ftp://ftp.cs.washington.edu/tr/2012/03/UW-CSE-12-03-01.PDF

11. Yildirim E. Mobile Privacy: Is There an App for That? On smart mobile devices, apps, and data protection. Master's Thesis. University of Amsterdam Institute for Information Law. May 2012. http://dare.uva.nl/cgi/arno/show.cgi?fid=461811

12. Frank M, Dong B, Felt A, Song D. Mining Permission Request Patterns from Android and Facebook Applications. 2012 IEEE 12th International Conference on Data Mining. December 10, 2012. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6413840

13. Chia P, Yamamoto Y, Asokan N. Is this app safe?: a large scale study on application permissions and risk signals. Proceedings of the 21st international conference on World Wide Web. April 16, 2012. http://dl.acm.org/citation.cfm?id=2187879

14. Felt A, Chin E, Hanna S, Song D, Wagner D. Android Permissions Demystified. In Proceedings of the 18th ACM Conference on Computer and Communications Security. October 17, 2011. http://dl.acm.org/citation.cfm?id=2046779

15. Felt A, Ha E, Egelman S, Haney A, Chin E, Wagner D. Android Permissions: User Attention, Comprehension, and Behavior. Proceedings of the 8th Symposium on Usable Privacy and Security, July 11, 2012. http://dl.acm.org/citation.cfm?id=2335360

16. Enck W, Ongtang M, McDaniel P. On Lightweight Mobile Phone Application Certification. Proceedings of the 16th ACM Conference on Computer and Communications Security. November 11, 2009. http://dl.acm.org/citation.cfm?id=1653691

17. Davi L, Dmitrienko A, Sadeghi A, Winandy M. Privilege Escalation Attacks on Android. Proceedings of the 13th International Conference on Information Security. October 25, 2010. http://dl.acm.org/citation.cfm?id=1949356

18. Bugiel S, Davi L, Dmitrienko A, Fischer T, Sadeghi A, and Shastry B. Towards Taming Privilege-Escalation Attacks on Android. Proceedings of the 19th Network and Distributed System Security Symposium. February 7, 2012. http://www.internetsociety.org/towards-taming-privilege-escalation-attacks-android

19. Portokalidis G, Homburg P, Anagnostakis K, and Bos H. Paranoid Android: Versatile Protection for Smartphones. Proceedings of the 26th Annual Computer Security Applications Conference. December 6, 2010. http://dl.acm.org/citation.cfm?id=1920313

20. Enck W, Octeau D, McDaniel P, Chaudhuri S. A Study of Android Application Security. Proceedings of the 20th USENIX Conference on Security, 2011. August 8, 2011. http://dl.acm.org/citation.cfm?id=2028067.2028088

21. Kelley P, Consolvo S, Cranor L, Jung J, Sadeh N, Wetherall D. A conundrum of permissions: installing applications on an android smartphone. Proceedings of the 16th international conference on Financial Cryptography and Data Security. March 2, 2012. http://dl.acm.org/citation.cfm?id=2426027

22. Thurm S, Kane Y. Your Apps Are Watching You. The Wall Street Journal. December 18, 2010. http://www.wsj.com/articles/SB10001424052748704368004576027751867039730 For methodology, see http://www.wsj.com/articles/SB10001424052748704034804576025951767626460

23. Kane Y. Apple Sued Over Mobile App Privacy. The Wall Street Journal. December 28, 2010. http://blogs.wsj.com/digits/2010/12/28/apple-sued-over-mobile-app-privacy/

24. Mui Y. Apple, app makers hit with privacy lawsuits. The Washington Post. December 28, 2010. http://www.washingtonpost.com/wp-dyn/content/article/2010/12/28/AR2010122803648.html

25. Canada Province of Quebec District of Montreal Superior Court. G Abilia vs. Apple, Inc. and Apple Canada Inc., and Gogii, Inc., Pandora Media, Inc., and Backflip Studios, Inc., and The Weather Channel, Inc., and Dictionary.com, LLC., and Outfit7 LTD., and Room Candy, Inc., and Sunstorm Interactive, Inc. Motion To Authorize the Bringing of A Class Action & To Ascribe the Status of Representative. December 30, 2012.

26. Panzarino M. CA Attorney General says Apple has agreed to make apps disclose privacy policy BEFORE download. The Next Web. February 22, 2012. http://thenextweb.com/mobile/2012/02/22/ca-attorney-general-says-apple-has-agreed-to-disclose-app-privacy-policy-before-download/

27. Schonfeld E. Apple Sneaks A Big Change Into iOS 5: Phasing Out Developer Access To The UDID. TechCrunch. August 19, 2011. http://techcrunch.com/2011/08/19/apple-ios-5-phasing-out-udid/

28. Perez S. iOS 7 Eliminates MAC Address As Tracking Option, Signaling Final Push Towards Apple's Own Ad Identifier Technology. TechCrunch. June 14, 2013. http://techcrunch.com/2013/06/14/ios-7-eliminates-mac-address-as-tracking-option-signaling-final-push-towards-apples-own-ad-identifier-technology/

29. Perez S. Apple Rejecting Apps Using Cookie-Tracking Methods, Signaling Push To Its Own Ad Identifier Technology Is Now Underway. TechCrunch. February 25, 2013. http://techcrunch.com/2013/02/25/apple-rejecting-apps-using-cookie-tracking-methods-signaling-push-to-its-own-ad-identifier-technology-is-now-underway/

30. Perez S. Apple Developers Must Now Agree To Ad Identifier Rules Or Risk App Store Rejection. TechCrunch. April 11, 2014. http://techcrunch.com/2014/04/11/apple-developers-must-now-agree-to-ad-identifier-rules-or-risk-app-store-rejection/

31. Rosenblatt S. Why Android won't be getting App Ops anytime soon. CNET. December 19, 2013. http://www.cnet.com/news/why-android-wont-be-getting-app-ops-anytime-soon/

32. Gibbs S. Why it took us so long to match Apple on privacy – a Google exec explains. The Guardian. June 9, 2015. http://www.theguardian.com/technology/2015/jun/09/google-privacy-apple-android-lockheimer-security-app-ops

33. Olmstead K, Lu K. Digital News — Revenue: Fact Sheet. Pew Research Center. April 29, 2015. http://www.journalism.org/2015/04/29/digital-news-revenue-fact-sheet/

34. O'Malley G. Location-Based Mobile Ads Forecast To Hit $10.8B In 2017. MediaPost. January 16, 2014. http://www.mediapost.com/publications/article/217520/forecast-location-mobile-ads-to-hit-108b-in-201.html

35. Barrera D, Kayacik H, Oorschot P, Somayaji A. A methodology for empirical analysis of permission-based security models and its application to android. Proceedings of the 17th ACM conference on Computer and communications security. October 4, 2010. http://dl.acm.org/citation.cfm?id=1866317

36. Felt A, Greenwood K, Wagner D. The effectiveness of application permissions. Proceedings of the 2nd USENIX conference on Web application development. June 15, 2011. http://dl.acm.org/citation.cfm?id=2002175

37. Vidas T, Christin N, Cranor L. Curbing Android Permission Creep. Proceedings of the 2011 Web 2.0 Security and Privacy Workshop. May 2011. http://www.ieee-security.org/TC/W2SP/2011/papers/curbingPermissionCreep.pdf

38. Book T, Pridgen A, Wallach D. Longitudinal Analysis of Android Ad Library Permissions. Mobile Security Technologies. April 18, 2013. http://arxiv.org/abs/1303.0857

39. Egele M, Kruegel C, Kirda E, Vigna G. PiOS: Detecting Privacy Leaks in iOS Applications. NDSS Symposium 2011. February 9, 2011. http://www.internetsociety.org/doc/pios-detecting-privacy-leaks-ios-applications-paper

40. Chin E, Felt A, Greenwood K, Wagner D. Analyzing inter-application communication in Android. Proceedings of the 9th international conference on Mobile systems, applications, and services. June 28, 2011. http://dl.acm.org/citation.cfm?id=2000018

41. Felt A, Wang H, Moshchuk A, Hanna S, Chin E. Permission Re-Delegation: Attacks and Defenses. 20th Usenix Security Symposium. 2011. https://wkr.io/assets/refs/felt2011permission.pdf

42. Rosen S, Qian Z, Mao Z. AppProfiler: A Flexible Method of Exposing Privacy-Related Behavior in Android Applications to End Users. Proceedings of the third ACM conference on Data and application security and privacy. February 18, 2013. http://dl.acm.org/citation.cfm?id=2435380

43. Lin J. Understanding and Capturing People's Mobile App Privacy Preferences. Dissertation. 2013. http://dl.acm.org/citation.cfm?id=2604485

44. Grace M, Zhou W, Jiang X, Sadeghi A. Unsafe exposure analysis of mobile in-app advertisements. Proceedings 5th ACM conference on Security and Privacy in Wireless and Mobile Networks. April 16, 2012. http://doi.acm.org/10.1145/2185448.2185464

45. Enck W, Gilbert P, Chun B, Cox L, Jung J, McDaniel P, Sheth A. TaintDroid: An Information-flow Tracking System for Realtime Privacy Monitoring on Smartphones. Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation. October 4, 2010. http://dl.acm.org/citation.cfm?id=1924971

46. Ashwin R, Kakhki A, Razaghpanahe A, Tang A, Wang S, Sherry J, Gill P, Krishnamurthy A, Legout A, Mislove A, Choffnes D. Using the Middle to Meddle with Mobile. Technical Report. Northeastern University. December 2013. http://meddle.mobi/papers/meddle-main.pdf

47. Le A, Varmarken J, Langhoff S, Shuba A, Gjoka M, Markopoulos A. AntMonitor : A System for Monitoring from Mobile Devices. ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big Internet Data (C2BID). August 17, 2015. http://dl.acm.org/citation.cfm?id=2787396

48. Ho J. A Snapshot of Data Sharing by Select Health and Fitness Apps. Spring Privacy Series: Consumer Generated and Controlled Health Data. Federal Trade Commission. Washington, DC. May 7, 2014. http://www.ftc.gov/news-events/events-calendar/2014/05/spring-privacy-series-consumer-generated-controlled-health-data

49. Njie C. Technical Analysis of the Data Practices and Privacy Risks of 43 Popular Mobile Health and Fitness Applications. Privacy Rights Clearinghouse. August 2013. https://www.privacyrights.org/mobile-medical-apps-privacy-technologist-research-report.pdf See also https://www.privacyrights.org/mobile-health-and-fitness-apps-what-are-privacy-risks

50. Lin J, Amini S, Hong J, Sadeh N, Lindqvist J, Zhang J. Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy Through Crowdsourcing. UbiComp '12. September 5, 2012. http://dl.acm.org/citation.cfm?id=2370290

51. Rainie L, Kiesler S, Kang R, Madden M. Anonymity, Privacy, and Security Online. Pew Research Center. September 5, 2013. http://www.pewinternet.org/2013/09/05/anonymity-privacy-and-security-online/

52. Futuresight. Mobile Privacy: Consumer research insights and considerations for policy makers. GSMA. February 2014. http://www.gsma.com/publicpolicy/wp-content/uploads/2014/10/GSMA-Mobile-Privacy-Booklet_WEBv2.pdf

53. Futuresight. User perspectives on mobile privacy. GSMA. September 2011. http://www.gsma.com/publicpolicy/wp-content/uploads/2012/03/futuresightuserperspectivesonuserprivacy.pdf

54. Singer N. Sharing Data, but Not Happily. The New York Times. June 4, 2015. http://www.nytimes.com/2015/06/05/technology/consumers-conflicted-over-data-mining-policies-report-finds.html

55. Madden M, Rainie L. Americans' Views About Data Collection and Security. Pew Research Center. May 20, 2015. http://www.pewinternet.org/2015/05/20/americans-views-about-data-collection-and-security/

56. Gruteser M, Grunwald D. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. Proceedings of the 1st international conference on Mobile systems applications and services MobiSys 03. March 4, 2003. http://dl.acm.org/citation.cfm?doid=1066116.1189037

57. Krumm J. A survey of computational location privacy. Personal and Ubiquitous Computing. August 1, 2009. http://dl.acm.org/citation.cfm?id=1569363

58. Wang Y, Chen Y, Ye F, Yang J, Liu H. Towards Understanding the Advertiser's Perspective of Smartphone User Privacy. 2015 IEEE 35th International Conference on Distributed Computing Systems. June 29, 2015. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7164915

59. Statista. Number of apps available in leading app stores as of July 2015. July 2015. http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

60. Cunningham A. Google Play apps and updates are now subject to a review process. Ars Technica. March 17, 2015. http://arstechnica.com/gadgets/2015/03/google-play-apps-and-updates-are-now-subject-to-a-review-process/

61. Baek D. Checklist: How to avoid the 10 most common App Store rejection reasons when submitting your app. Nodes. March 6, 2015. http://www.nodesagency.com/how-to-avoid-the-most-common-app-store-rejection-reasons-when-submitting-your-app/

62. Apple. App Store Review Guidelines. Accessed September 22, 2015. https://developer.apple.com/app-store/review/guidelines/

63. Fawaz K, Feng H, Shin K. Anatomization and Protection of Mobile Apps ' Location Privacy Threats. Proceedings of the 24th USENIX Security Symposium. August 12, 2015. https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/fawaz

64. Arthur C. Android's permissions gap: why has it fallen so far behind Apple's iOS? The Guardian. December 20, 2013. http://www.theguardian.com/technology/2013/dec/20/android-apps-permission-app-ops

65. Libert T. Privacy Implications of Health Information Seeking on the Web. Communications of the ACM. February 23, 2015. http://dl.acm.org/citation.cfm?id=2658983

66. Brandeisky K. Your Embarrassing Online Searches About Health Problems Aren't Private. February 25, 2015. http://time.com/money/3721200/health-privacy-online/

67. comScore. comScore Ranks the Top 50 U.S. Digital Media Properties for May 2015. June 19, 2015. http://www.comscore.com/Insights/Market-Rankings/comScore-Ranks-the-Top-50-US-Digital-Media-Properties-for-May-2015

68. Google. Monetize and promote with Google Ads. Google Developers. August 12, 2015. https://developers.google.com/ads/?hl=en

69. Apple. iAd for Developers. Apple Developer. Accessed September 22, 2015. https://developer.apple.com/iad/

70. Beresford A, Rice A, Skehin N. MockDroid: trading privacy for application functionality on smartphones. Proceedings of the 12th Workshop on Mobile Computing Systems and Applications. March 1, 2011. http://dl.acm.org/citation.cfm?id=2184500

71. Zhou Y, Zhang X, Jiang X, Freeh V. Taming information-stealing smartphone applications (on Android). June 22, 2011. http://dl.acm.org/citation.cfm?id=2022245.2022255

72. Office of the Attorney General. Making Your Privacy Practices Public: Recommendations on Developing a Meaningful Privacy Policy. California Department of Justice. May 2014. https://oag.ca.gov/sites/all/files/agweb/pdfs/cybersecurity/making_your_privacy_practices_public.pdf

73. Wyatt E. F.T.C. Suggests Privacy Guidelines for Mobile Apps. The New York Times. February 1, 2013. http://www.nytimes.com/2013/02/02/technology/ftc-suggests-do-not-track-feature-for-mobile-software-and-apps.html

74. Dilger D. Apple adds new "Limit Ad Tracking" feature to iOS 6. AppleInsider. September 13, 2012. http://appleinsider.com/articles/12/09/13/apple_adds_new_limit_ad_tracking_feature_to_ios_6

75. O'Grady J. Why the iOS 'Limit Ad Tracking' setting is more important than ever. ZDNet. April 14, 2014. http://www.zdnet.com/article/why-the-ios-limit-ad-tracking-setting-is-more-important-than-ever/

76. Sterling G. Google Replacing "Android ID" With "Advertising ID" Similar To Apple's IDFA. Marketing Land. October 31, 2013. http://marketingland.com/google-replacing-android-id-with-advertising-id-similar-to-apples-idfa-63636

77. Google. Opt out – Ads Help. Accessed September 22, 2015. https://support.google.com/ads/answer/2662922?hl=en

78. Google. Google Play Developer Program Policies. Google Play. Accessed September 22, 2015. https://play.google.com/about/developer-content-policy.html

79. McDonald A, Cranor L. The Cost of Reading Privacy Policies. A Journal of Law and Policy for the Information Society. October 6, 2008. http://lorrie.cranor.org/pubs/readingPolicyCost-authorDraft.pdf

80. Hornyack P, Han S, Jung J, Schechter S, Wetherall D. These Aren't the Droids You're Looking for: Retrofitting Android to Protect Data from Imperious Applications. October 17, 2011. http://doi.acm.org/10.1145/2046707.2046780

81. Wetherall D, Choffnes D, Greenstein B, Han S, Hornyack P, Jung J, Scheschter S, Wang X. Privacy revelations for web and mobile apps. Proceedings of the 13th USENIX conference on Hot topics in operating systems. May 9, 2011. http://portal.acm.org/citation.cfm?id=1991596.1991625

82. Cohen K, Yeung C. Kids' Apps Disclosures Revisited. Federal Trade Commission. September 3, 2015. https://www.ftc.gov/news-events/blogs/business-blog/2015/09/kids-apps-disclosures-revisited

83. Abramson S. California Senate Unanimously Passes Online Privacy Bill That Would Give Minors an "Eraser Button". Inside Privacy. May 6, 2013. http://www.insideprivacy.com/childrens-privacy/california-senate-unanimously-passes-online-privacy-bill-that-would-give-minors-an-eraser-button/

84. Dredge S. Moms With Apps aims to direct parents to responsible apps for kids. The Guardian. December 31, 2014. http://www.theguardian.com/technology/2014/dec/31/moms-with-apps-parents-kids

85. Know What's Inside. Discover Kids Apps. Accessed September 22, 2015. https://knowwhatsinside.com/discover

86. Perez S. Introducing Apple's New "Kids" App Store. TechCrunch. September 22, 2013. http://techcrunch.com/2013/09/22/introducing-apples-new-kids-app-store/

87. Perez S. Google Play's New Program "Designed For Families" Will Highlight Pre-Approved, Kid-Safe Apps. TechCrunch. April 14, 2015. http://techcrunch.com/2015/04/14/google-plays-new-program-designed-for-families-will-highlight-pre-approved-kid-safe-apps/

88. Black J, Cochran M, Highland T. A Study of the MD5 Attacks: Insights and Improvements. Proceedings of the 13th international conference on Fast Software Encryption. March 15, 2006. http://dl.acm.org/citation.cfm?id=2123073

89. Google. Applications – Android Apps on Google Play. Accessed September 22, 2015. https://play.google.com/store/apps/category/APPLICATION?hl=en

90. Apple. App Store Downloads on iTunes. Accessed September 22, 2015. https://itunes.apple.com/us/genre/ios/id36?mt=8

91. Federal Trade Commission. Path Social Networking App Settles FTC Charges it Deceived Consumers and Improperly Collected Personal Information from Users' Mobile Address Books. February 1, 2013. https://www.ftc.gov/news-events/press-releases/2013/02/path-social-networking-app-settles-ftc-charges-it-deceived

92. Federal Trade Commission. Android Flashlight App Developer Settles FTC Charges It Deceived Consumers. December 5, 2013. https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived

93. Lovejoy B. How many apps do you use a month? Study shows the average is 26. 9to5Mac. July 2, 2014. http://9to5mac.com/2014/07/02/how-many-apps-do-you-use-a-month-study-shows-the-average-is-26-poll/

94. Apple. About privacy and Location Services for iOS 8 and iOS 9. September 16, 2015. https://support.apple.com/en-is/HT203033

95. How-to-Geek. Android's App Permissions Were Just Simplified — Now They're Much Less Secure. June 11, 2014. http://www.howtogeek.com/190863/androids-app-permissions-were-just-simplified-now-theyre-much-less-secure/

Appendix

Sensitive data that was searched for on Android.

Designation	Data Type	Canary	Term searched
PII	Address	ADDRESS2/CITY2	Chicago
PII	Address	ADDRESS1	Michigan
PII	Birthday	BIRTHDATE1	1990
PII	Birthday	BIRTHDATE3	2/14/90
PII	Birthday	BIRTHDATE4	2-14-90
PII	Birthday	BIRTHDATE5	Feb 14
PII	Birthday	BIRTHDATE6	birth
PII	Birthday	BIRTHDATE7	February
PII	Email	EMAIL1	baileylogan202@gmail.com
PII	Email	EMAIL2	baileylogan202%40 gmail.com
PII	Email	EMAIL3	baileylogan202\\\\u0040gmail.com
PII	Email	WRONG_EMAIL	baileylogan@gmail.com
Behavior	Employment	SEARCH24	burger
Behavior	Employment	SEARCH25	driver
Behavior	Employment	SEARCH26	truck

Designation	Data Type	Canary	Term searched
Behavior	Employment	SEARCH27	cashier
Behavior	Friend	FRIEND NAME1	Addison
Behavior	Friend	FRIEND NAME2	Hardwick
Behavior	Friend	FRIEND EMAIL1	addisonhardwick823@gmail.com
Behavior	Friend	FRIEND EMAIL2	baileylogan202%40 gmail.com
Behavior	Friend	FRIEND EMAIL3	baileylogan202\\\\u0040gmail.com
Behavior	Friend	PHONE NUMBER1	617-678-9364
Behavior	Friend	PHONE NUMBER2	6176789364
Behavior	Friend	PHONE NUMBER3	\(617\)678-9364
PII	Gender	GENDER1	Female
PII	Gender	GENDER2	gender
Location	Location	LAT_DC	38\.9
Location	Location	LAT_DC2	38\.8
Location	Location	LON_DC	-77\.0
Location	Location	LON_DC2	-76\.9
Location	Location	CITY/CITY1	Washington
Behavior	Medical Info	SEARCH1	antidepressants
Behavior	Medical Info	MEDICATION1	interferon
Behavior	Medical Info	SEARCH10	asparagus
Behavior	Medical Info	MEDICATION2	valtrex
Behavior	Medical Info	MEDICATION3	sofosbuvir
Behavior	Medical Info	MEDICATION4	zoloft
Behavior	Medical Info	CONDITION1	herpes
Behavior	Medical Info	CONDITION2	hepatitis
Behavior	Medical Info	MEDINFO1	blood pressure
Behavior	Medical Info	MEDINFO2	cervical
Behavior	Medical Info	CONDITION3	Appendicitis
Behavior	Medical Info	HEIGHT	height
Behavior	Medical Info	WEIGHT	weight
Behavior	Medical Info	MEDINFO3	bagel
PII	Name	NAME1	Bailey
PII	Name	NAME2	Logan
PII	Password	PASSWORD	202P3nNs
PII	Phone Info	MEID	[redacted]
PII	Phone Info	MAC ADDRESS	[redacted]
PII	Phone Number	PHONE	^\D?(\d{3})\D?\D?(\d{3})\D?(\d{4})$
PII	Phone Number	PHONE NUMBER4	312-498-2841
PII	Phone Number	PHONE NUMBER5	3124982841
PII	Phone Number	PHONE NUMBER6	\(312\)498-2841
Behavior	Post	INPUT1	Resurrection
Behavior	Post	INPUT2	magnolia
Behavior	Post	INPUT4	student
Behavior	Post	INPUT5	omelette
Behavior	Post	INPUT6	Awwwww
Behavior	Post	SEARCH22	floral
Behavior	Post	INPUT7	ohmygodbeckylookatthat
Behavior	Post	INPUT8	gorgonzola

Designation	Data Type	Canary	Term searched
Behavior	Post	INPUT9	basketball
Behavior	Post	INPUT10	spectacular
Behavior	Post	INPUT11	anointed
Behavior	Post	INPUT12	transylvania
Behavior	Post	INPUT13	calories
Behavior	Post	INPUT14	aphrodite
Behavior	Post	INPUT15	malaysia
Behavior	Post	INPUT16	beedubs
Behavior	Post	INPUT17	masquerading
Behavior	Post	INPUT18	lithograph
Behavior	Post	INPUT19	Ouagadougou
Behavior	Post	INPUT20	alcoholic
Behavior	Post	INPUT21	frankenstein
Behavior	Post	INPUT22	eschewing
Behavior	Post	INPUT23	malaysia
Behavior	Post	INPUT24	Hermetic
Behavior	Post	INPUT25	spectacular
Behavior	Post	INPUT26	Gorgonzola
Behavior	Post	INPUT27	Hakusai
Behavior	Post	INPUT28	macroscopic
Behavior	Post	INPUT29	theologian
Behavior	Post	INPUT30	Incontrovertible
Behavior	Post	INPUT31	frabjuous
Behavior	Post	INPUT32	peculiar
Behavior	Post	INPUT33	Melanie
Behavior	Search	SEARCH4	adrenaline
Behavior	Search	SEARCH6	pampered
Behavior	Search	SEARCH9	condoms
Behavior	Search	SEARCH11	psychologist
Behavior	Search	SEARCH12	stroller
Behavior	Search	SEARCH13	sex toy
Behavior	Search	SEARCH14	agriculture
Behavior	Search	INPUT3	salamander
Behavior	Search	SEARCH23	pharmacy
Behavior	Search	LAT_MEX	-19\.4
Behavior	Search	LON_MEX	-99\.1
Behavior	Search	LAT	lat(?=[|i|=|:])
Behavior	Search	LON	lon(?=[|g|=|:])
Behavior	Search	ADDRESS3	Omaha
Behavior	Search	SEARCH28	Mediterranean
Behavior	Search	SEARCH29	Regency
Behavior	Search	SEARCH29	Marrakech
Behavior	Search	SEARCH30	Sepulveda
Behavior	Search	INPUT34	Boostrix
Behavior	Username	USERNAME1	baileylogan_202
Behavior	Username	USERNAME2	baileylogan202
Behavior	Username	USERNAME3	baileylogankik

Designation	Data Type	Canary	Term searched
Behavior	Username	USERNAME4	baileylogannike
Behavior	Username	USERNAME5	baileyloganpoints
Behavior	Username	USERNAME6	baileyloganskype
Behavior	Username	USERNAME7	baileylogansnap
PII	Zipcode	ZIPCODE	60604

Sensitive data that was searched for on iOS.

Designation	Data Type	Canary	Term searched
PII	Address	ZIPCODE	02459
PII	Address	ZIPCODE_WRONG	02549
PII	Address	ADDRESS	Newton
PII	Address	ZIPCODE	02459
PII	Birthday	BIRTHDATE1	1986
PII	Birthday	BIRTHDATE3	6/18/86
PII	Birthday	BIRTHDATE4	6-18-86
PII	Birthday	BIRTHDATE5	June 18
PII	Birthday	BIRTHDATE6	birth
PII	Phone Info	ICCID	[redacted]
PII	Phone Info	MAC	[redacted]
PII	Email	EMAIL	addisonhardwick823@gmail.com
Behavior	Employment	SEARCH32	Nurse
Behavior	Employment	SEARCH34	Car Mechanic
Behavior	Friend	FRIEND NAME1	Bailey
Behavior	Friend	FRIEND NAME2	Logan
Behavior	Friend	FRIEND	baileylogan202@gmail.com
Behavior	Friend	PHONE NUMBER4	312-498-2841
Behavior	Friend	PHONE NUMBER5	3124982841
Behavior	Friend	PHONE NUMBER6	\(312\)498-2841
PII	Gender	GENDER1	Female
PII	Gender	GENDER2	gender
Location	Location	SEARCH15/CITY1	Washington
Location	Location	LAT_DC	38\.9
Location	Location	LAT_DC2	38\.8
Location	Location	LON_DC	-77\.0
Location	Location	LON_DC2	-76\.9
Location	Location	LAT_MEX	-19\.4
Location	Location	LON_MEX	-99\.1
Behavior	Medical Info	SEARCH1	panini
Behavior	Medical Info	SEARCH2	watermelon
Behavior	Medical Info	SEARCH3	archery
Behavior	Medical Info	MEDICATION1	seasonique
Behavior	Medical Info	SEARCH4	linea nigra
Behavior	Medical Info	MEDINFO1	insomnia
Behavior	Medical Info	SEARCH6	miscarriage
Behavior	Medical Info	MEDINFO2	blood pressure

Designation	Data Type	Canary	Term searched
Behavior	Medical Info	MEDINFO3	pregnant
Behavior	Medical Info	MEDINFO4	pregnancy
Behavior	Medical Info	SEARCH16	fibromyalgia
Behavior	Medical Info	SEARCH17	urine-bloody
Behavior	Medical Info	MEDICATION2	Voltaren
Behavior	Medical Info	MEDICATION3	abilify
Behavior	Medical Info	MEDICATION4	diflucan
Behavior	Medical Info	CONDITION1	glomerulonephritis
Behavior	Medical Info	CONDITION2	vulvodynia
Behavior	Medical Info	CONDITION3	ovarian
Behavior	Medical Info	MEDINFO5	92\.23
Behavior	Medical Info	MEDINFO6	cervical
Behavior	Medical Info	MEDICATION5	amitryptiline
Behavior	Medical Info	SEARCH20	neurologist
Behavior	Medical Info	MEDICATION6	lipitor
Behavior	Medical Info	MEDICATION7	ibuprofen
Behavior	Medical Info	SEARCH31	Constipation
Behavior	Medical Info	CONDITION4	Fissures
Behavior	Medical Info	CONDITION5	Kidney
Behavior	Medical Info	CONDITION6	Intercourse
Behavior	Medical Info	CONDITION7	pain
Behavior	Medical Info	MEDINFO7	yogurt
Behavior	Medical Info	MEDINFO8	toast
Behavior	Medical Info	MEDINFO9	cardio
Behavior	Medical Info	CONDITION8	groin
Behavior	Medical Info	HEIGHT	height
Behavior	Medical Info	WEIGHT	weight
PII	Name	NAME1	Addison
PII	Name	NAME2	Hardwick
PII	Password	PASSCODE1	823823
PII	Password	PASSWORD	823P3nNs
PII	Phone Number	PHONE NUMBER1	617-678-9364
PII	Phone Number	PHONE NUMBER2	6176789364
PII	Phone Number	PHONE NUMBER3	\(617\)678-9364
PII	Phone Number	PHONE	^\D?(\d{3})\D?\D?(\d{3})\D?(\d{4})$
Behavior	Post	INPUT1	snickerdoodle
Behavior	Post	INPUT2	oysters
Behavior	Post	INPUT3	banality
Behavior	Post	INPUT4	Roquefort
Behavior	Post	INPUT5	Clarinet
Behavior	Post	INPUT6	Awwwww
Behavior	Post	INPUT7	ohmygodbeckylookatthat
Behavior	Post	INPUT8	bratwurst
Behavior	Post	INPUT9	flattery
Behavior	Post	INPUT10	obfuscation
Behavior	Post	INPUT11	penguin
Behavior	Post	INPUT12	transylvania

Designation	Data Type	Canary	Term searched
Behavior	Post	INPUT13	condom
Behavior	Post	INPUT14	jasmine
Behavior	Search	SEARCH9	Arlington
Behavior	Search	SEARCH10	croissant
Behavior	Search	SEARCH11	pharmacy
Behavior	Search	SEARCH12	Reagan
Behavior	Search	SEARCH13	nightlife
Behavior	Search	SEARCH14	Boston
Behavior	Search	SEARCH18	Sativa
Behavior	Search	SEARCH19	marijuana
Behavior	Search	SEARCH21	CollegeHumor
Behavior	Search	SEARCH22	Adidas
Behavior	Search	SEARCH23	briefcase
Behavior	Search	SEARCH24	barbie
Behavior	Search	SEARCH25	Nong Shim
Behavior	Search	SEARCH26	pineapple
Behavior	Search	SEARCH27	volkswagen
Behavior	Search	SEARCH28	stroller
Behavior	Search	SEARCH29	BabiesRUs
Behavior	Search	SEARCH30	onesie
Behavior	Search	SEARCH33	Chicago
Behavior	Search	LAT	lat(?[|i|:])
Behavior	Search	LON	lon(?[|g|:])
Behavior	Search	CITY2	Chicago
Behavior	Search	SEARCH36	Baltimore
Behavior	Username	USERNAME1	addie823
Behavior	Username	USERNAME2	addishardw

All Apps that were investigated on Android.

Date collected	App	Version	Reason for skipping
30-Jun	Job Search (Indeed.com)	2.3	
21-Jul	Facebook Pages Manager	5.0	[during the first go-round, we were unable to use social]
27-Jun	Job Search (Snagajob)	2.9.1	
27-Jun	Square Register		required having a Square/business and credit card information
27-Jun	QuickOffice		Was bought by Google and discontinued by the time it came to download
27-Jun	Don't Tap the White Tile	2.5.1	
27-Jun	Guess the Emoji: Emoji Pops	4.0	
	Angry Birds Epic		Wouldn't load
30-Jun	Candy Crush Saga	1.34.1	
30-Jun	Bubble Witch 2 Saga	1.4.2	

Date collected	App	Version	Reason for skipping
27-Jun	Monster Legends	1.7.3	
27-Jun	MyFitnessPal	3.3	
30-Jun	Fitbit	1.9.7	
27-Jun	iTriage Health	5.27	
27-Jun	WebMD	3.5	
27-Jun	Period Tracker	2.0.4.2	
27-Jun	Google Earth	7.1.3.1255	
21-Jul	Waze	3.8.1.0	Couldn't operate with full functionality inside of a building
27-Jun	Yelp	5.12.2	
27-Jun	Gasbuddy	4.2.2	
27-Jun	MapFactor: GPS Navigation	1.2.50	
18-Jul	Period Calendar	1.467	
18-Jul	RunKeeper	4.6.5	
18-Jul	Map My Walk	3.1.2	
18-Jul	Nike+ Running	1.4.1	
7/18 + 7/21	Lose It!	5.1.5	
21-Jul	MapQuest	2.6.0	
21-Jul	Scout	2.1.0.0313	
21-Jul	Priceline	3.4.34	
21-Jul	Expedia	3.6.1	
21-Jul	GPS Navigation BE-ON-ROAD	3.10.16	
21-Jul	American Well	7.3.0.005_01	
21-Jul	MyChart	3.3.1	Unable to use app because it requires a medical record at a participating hospital
21-Jul	Points2Shop	180.2.3-004	
21-Jul	GoodRX	2.2.0	
21-Jul	Epocrates	14.6	
18-Jul	Facebook	13.0.0.13.14	
18-Jul	Instagram	6.2.2	
18-Jul	Snapchat	5.0.27.3	
21-Jul	Twitter	5.18.0	Couldn't log in on the phone (no error message?)
18-Jul	Pinterest	3.1.2	
22-Jul	Tango	3.8.95706	
21-Jul	Vine	2.1.0	Didn't work because Twitter didn't work
22-Jul	ooVoo	2.2.1	"Error while connecting to service. Please try again later."
23-Jul	Emoji Smart Android Keyboard	1.02	
18-Jul	Tumblr	3.6.2.02	Couldn’t make an account
18-Jul	Facebook Messenger	8.0.0.20.14	
18-Jul	Skype	5.0.0.49715	
18-Jul	Kik	7.3.1.111	

Date collected	App	Version	Reason for skipping
21-Jul	WhatsApp	2.11.301	"Unable to connect"
23-Jul	Line	4.5.4	Said the current Wi-Fi network did not have connectivity
21-Jul	File Manager	1.17.0	
21-Jul	ADP Mobile		Unable to explore full functionality due to lack of employer
21-Jul	Box	3.1.2	
21-Jul	LinkedIN		Would have required impersonating a person to companies and other people
21-Jul	POF Free Online Dating		Would have required impersonating a person to people on a dating site
22-Jul	textPlus free	5.9.8	
22-Jul	Timehop	1.3.10	
22-Jul	Text Free	2.3.2	
23-Jul	Emoji Smart Keyboard	2.0	
23-Jul	Yahoo Mail		Mail apps could use alternative ports which were not being tracked
22-Jul	Viber	4.3.3.67	
23-Jul	Glide	Glide.v1.04.006	
5-Aug	KAYAK	6.1.2	
22-Jul	Drugs.com	1.23	
21-Jul	eBay	2.6.1.2	
21-Jul	Amazon	3.0.0	
21-Jul	Walgreens	4.4.1	
21-Jul	Groupon	4.3508	
21-Jul	Wish	3.7.0	

All Apps that were investigated on iOS.

Date collected	App	Version	Reason for skipping
27-Jun	Job Search (Indeed.com)	2.5	
27-Jun	LinkedIN Job Search	1.0.1	Would have required impersonating a person to companies and other people
3-Jul	Job Search (Snagajob)	2.5.1	
27-Jun	Adobe Reader	11.3.1	
27-Jun	OWA for iPhone		Wouldn't download for the version of iPhone we were using
3-Jul	ADP Mobile	1.9.3	
27-Jun	Fish Out of Water	1.2	
27-Jun	Guess The Emoji: Emoji Pops	4.0	
27-Jun	TwoDots	1.0.2	
3-Jul	Fruit Ninja	1.9.2	
27-Jun	Piano Tiles (Don't Tap the White Tile)	2.4	

Date collected	App	Version	Reason for skipping
27-Jun	MyFitnessPal	5.3.2	
27-Jun	Fitbit	2.3	
3-Jul	Map My Run	5.4.7	
27-Jun	WebMD	5	
3-Jul	RunKeeper	4.6	
3-Jul	Google Maps	3.1.2	
27-Jun	Waze	3.8.0	Couldn't operate with full functionality inside a building
3-Jul	MapQuest	4.3	
3-Jul	Scout	1.17.1	
3-Jul	Track Kit	1.1	
17-Jul	Nike+ Running	4.5.5	
17-Jul	Lose It!	5.2.2	
17-Jul	Period Tracker	9.2	
17-Jul	The Bump Pregnancy	2.1	
17-Jul	Pacer - Pedometer Plus	2.3.1	
17-Jul	Phone Tracker for iPhone	3.4.7	
17-Jul	Geocaching	2.4.1	
17-Jul	HopStop	2.6.1	SSL error (see screenshot)
17-Jul	Moovit	3.5.1	
17-Jul	Speedometer	6.2	Unable to use app w/o a car
17-Jul	GPS by Telenav	7.0.6	
22-Jul	INRIX	5.3.1	
18-Jul	Local Scope	4.2.1	
17-Jul	MyChart	3.2.1	Unable to use app because it requires a medical record at a participating hospital
18-Jul	Urgent Care	1.9	
18-Jul	GoodRx	3.6.1	
18-Jul	Ovia Fertility	3.3.3	
18-Jul	Follow My Health	1.2.7	Unable to use app because it requires a medical record at a participating hospital
18-Jul	CareZone Meds	3.6.4	URL error (see other screenshot)
18-Jul	Leafly Marijuana	2.1.2	
22-Jul	American Well	7.3.009	
18-Jul	Youtube	2.7.1	
18-Jul	Instagram	6.0.4	
18-Jul	Snapchat	7.0.4	
18-Jul	Flipagram	3.1	
18-Jul	InstaSize	2.9.3	
22-Jul	Facebook Messenger	2.0	
22-Jul	Facebook	12.1	
18-Jul	WhatsApp Messenger	2.11.8	Unable to fully test because the Android version was not functional
18-Jul	Twitter	6.9	URL error (see other screenshot)
18-Jul	Timehop	2.6.3	
18-Jul	Kik	7.2.1	

Date collected	App	Version	Reason for skipping
18-Jul	Emoji Keyboard	1.3	
18-Jul	Pinterest	3.6.5	
18-Jul	Skype	5.2.98	
18-Jul	Vine	2.1.1	URL error (see other screenshot)
22-Jul	Viber	4.2.1	
22-Jul	Tango	3.8.94526	
22-Jul	Tumblr	3.6.1	Refused to log in (no error message)
22-Jul	Hangouts	2.1.0	
18-Jul	Smart Scan Express	4.2	
18-Jul	Amazon App	4.0.1	
21-Jul	Groupon	3.4.1	
21-Jul	eBay	3.3.1	
21-Jul	Wish - Shopping Made Fun	3.6.0	
21-Jul	Walgreens	4.2.1	

Authors

Jinyan Zang is an experienced researcher on consumer protection, data security, and privacy issues as a Research Fellow at the Federal Trade Commission and a Research Analyst at Harvard University. He is currently working with Prof. Latanya Sweeney as the Managing Editor of Technology Science, the first open access, peer reviewed, online publication by Harvard University of research on the unforeseen consequences of technology on society. He graduated cum laude in 2013 from Harvard College with a BA in Economics.

Krysta Dummit is a first-year PhD candidate in Chemistry at the Massachusetts Institute of Technology. Her research interests include organometallic chemistry, synthesis, and catalyst design. While obtaining her BA at Princeton, she obtained a certificate in Computer Science and worked as a Research Fellow in Technology and Data Governance under Dr. Latanya Sweeney at the Federal Trade Commission.

Jim Graves is a PhD student in Engineering and Public Policy at Carnegie Mellon University, where his research focuses on the law and economics of data privacy. Before returning to school, he worked as a data security and networking professional for over 15 years. Jim earned his JD from William Mitchell College of Law, where he was Editor-in-Chief of the Law Review, and holds an M.S. in Information Networking and a B.S. in Mathematics and Computer Science, both from Carnegie Mellon University.

Paul Lisker '17 is a Computer Science student at Harvard College with a minor in Government. A former Technology and Data Governance fellow at the Federal Trade

Commission and software engineer at a data–use–protection start-up, he is passionate about the growing intersection of technology, privacy and government. He is a proud Mexican-American dual citizen also interested in international technology concerns.

Latanya Sweeney is Professor of Government and Technology in Residence at Harvard University, Director of the Data Privacy Lab at Harvard, Editor-in-Chief of Technology Science, and was formerly the Chief Technology Officer of the U.S. Federal Trade Commission. She earned her PhD in computer science from the Massachusetts Institute of Technology and her undergraduate degree from Harvard. More information about Dr. Sweeney is available at her website at latanyasweeney.org. As Editor-In-Chief of Technology Science, Professor Sweeeney was recused from the review of this paper.

This work was conducted at the Federal Trade Commission during the summer of 2014 as part of the Summer Research Fellows Program. All statements, analyses and conclusions are the authors' and do not necessarily reflect any position held by the Federal Trade Commission or any Commissioner.

Referring Editor: James Waldo

Citation

Zang J, Dummit K, Graves J, Lisker P, Sweeney L. Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps. Technology Science. 2015103001. October 30, 2015. http://techscience.org/a/2015103001

Data

The data is under classification review.