Institute for Software-Integrated Systems

<u>Technical Report</u>

TR#:	**ISIS-20-201**
Title:	**Google Data Collection**
Authors:	**<u>Douglas C. Schmidt</u>**

<u>**Google Data Collection**</u>

Professor Douglas C. Schmidt, Vanderbilt University

August 15, 2018

I. EXECUTIVE SUMMARY

1.   Google is the world's largest digital advertising company.[1] It also provides the #1 web browser,[2] the #1 mobile platform,[3] and the #1 search engine[4] worldwide. Google's video platform, email service, and map application have over 1 billion monthly active users each.[5] Google utilizes the tremendous reach of its products to collect detailed information about people's online and real-world behaviors, which it then uses to target them with paid advertising. Google's revenues increase significantly as the targeting technology and data are refined.

2.   Google collects user data in a variety of ways. The most obvious are "active," with the user directly and consciously communicating information to Google, as for example by signing in to any of its widely used applications such as YouTube, Gmail, Search etc. Less obvious ways for Google to collect data are "passive" means, whereby an application is instrumented to gather information while it's running, possibly without the user's knowledge. Google's passive data gathering methods arise from platforms (e.g. Android and Chrome), applications (e.g. Search, YouTube, Maps), publisher tools (e.g. Google Analytics, AdSense) and advertiser tools (e.g. AdMob, AdWords). The extent and magnitude of Google's passive data collection has largely been overlooked by past studies on this topic.[6]

3.   To understand what data Google collects, this study draws on four key sources:

   a. Google's My Activity[7] and Takeout[8] tools, which describe information collected during the use of Google's user-facing products;

   b. Data intercepted as it is sent to Google server domains while Google or 3rd-party products are used;

   c. Google's privacy policies (both general and product-specific); and

   d. Other 3rd-party research that has examined Google's data collection efforts.

4.   Through the combined use of above resources, this study provides a unique and comprehensive view of Google's data collection approaches and delves deeper into specific types of information it collects from users. This study highlights the following key findings:

---

[1] "Google and Facebook tighten grip on US digital ad market," *eMarketer*, Sept. 21, 2017, available at
https://www.emarketer.com/Article/Google-Facebook-Tighten-Grip-on-US-Digital-Ad-Market/1016494
[2] "Market share or leading internet browsers in the United States and worldwide as of February 2018," *Statista,* February 2018, available at https://www.statista.com/statistics/276738/worldwide-and-us-market-share-of-leading-internet-browsers/
[3] "Global OS market share in sales to end users from 1st quarter 2009 to 2nd quarter 2017," *Statista*, August 2017, available at https://www.statista.com/statistics/266136/global-market-share-held-by-smartphone-operating-systems/
[4] "Worldwide desktop market share of leading search engines from January 2010 to October 2017," *Statista,* Feb. 2018, available at https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/
[5] Google 10K filings with the SEC, 2017, available at https://abc.xyz/investor/pdf/20171231_alphabet_10K.pdf
[6] Please see Appendix section IX.F for a list of past studies/news reports on Google data collection
[7] "My Activity," *Google*, available at https://myactivity.google.com/myactivity
[8] "Download your data," *Google*, available at https://takeout.google.com/settings/takeout?pli=1

a. Google learns a great deal about a user's personal interests during even a single day of typical internet usage. In an example "day in the life" scenario, where a real user with a new Google account and an Android phone (with new SIM card) goes through her daily routine, Google collected data at numerous activity touchpoints, such as user location, routes taken, items purchased, and music listened to. Surprisingly, Google collected or inferred over two-thirds of the information through passive means. At the end of the day, Google identified user interests with remarkable accuracy.

b. Android is a key enabler of data collection for Google, with over 2 billion monthly active users worldwide.[9] While the Android OS is used by Original Equipment Manufacturers (OEMs) around the world, it is tightly connected with Google's ecosystem through Google Play Services. Android helps Google collect personal user information (e.g. name, mobile phone number, birthdate, zip code, and in many cases, credit card number), activity on the mobile phone (e.g. apps used, websites visited), and location coordinates. In the background, Android frequently sends Google user location and device-related information, such as apps usage, crash reports, device configuration, backups, and various device-related identifiers.

c. The Chrome browser helps Google collect user data from both mobile and desktop devices, with over 2 billion active installs worldwide.[10] The Chrome browser collects personal information (e.g. when a user completes online forms) and sends it to Google as part of the data synchronization process. It also tracks webpage visits and sends user location coordinates to Google.

d. Both Android and Chrome send data to Google even in the absence of *any* user interaction. Our experiments show that a dormant, stationary Android phone (with Chrome active in the background) communicated location information to Google 340 times during a 24-hour period, or at an average of 14 data communications per hour. In fact, location information constituted 35% of all the data samples sent to Google. In contrast, a similar experiment showed that on an iOS Apple device with Safari (where neither Android nor Chrome were used), Google could not collect any appreciable data (location or otherwise) in the absence of a user interaction with the device.

e. After a user starts interacting with an Android phone (e.g. moves around, visits webpages, uses apps), passive communications to Google server domains increase significantly, even in cases where the user did not use any prominent Google applications (i.e. no Google Search, no YouTube, no Gmail, and no Google Maps). This increase is driven largely by data activity from Google's publisher and advertiser products (e.g. Google Analytics, DoubleClick, AdWords)[11]. Such data constituted 46% of all requests

---

[9] Dave Burke, "Android: celebrating a big milestone together with you," *Google,* May 17, 2017, available at
https://www.blog.google/products/android/2bn-milestone/
[10] Frederic Lardinois, "Google says there are now 2 billion active Chrome installs," *TechCrunch,* Nov. 10, 2016, available at
https://techcrunch.com/2016/11/10/google-says-there-are-now-2-billion-active-chrome-installs/
[11] Google recently rebranded AdWords as "Google Ads" and DoubleClick as "Google Ad Manager"

to Google servers from the Android phone. Google collected location at a 1.4x higher rate compared to the stationary phone experiment with no user interaction. Magnitude wise, Google's servers communicated 11.6 MB of data per day (or 0.35 GB/month) with the Android device. This experiment suggests that even if a user does not interact with any key Google applications, Google is still able to collect considerable information through its advertiser and publisher products.

f. While using an iOS device, if a user decides to forgo the use of *any* Google product (i.e. no Android, no Chrome, no Google applications), and visits only non-Google webpages, the number of times data is communicated to Google servers still remains surprisingly high. This communication is driven purely by advertiser/publisher services. The number of times such Google services are called from an iOS device is similar to an Android device. In this experiment, the total magnitude of data communicated to Google servers from an iOS device is found to be approximately half of that from the Android device.

g. Advertising identifiers (which are purportedly "user anonymous" and collect activity data on apps and 3rd-party webpage visits) can get connected with a user's Google identity. This happens via passing of device-level identification information to Google servers by an Android device. Likewise, the DoubleClick cookie ID (which tracks a user's activity on the 3rd-party webpages) is another purportedly "user anonymous" identifier that Google can connect to a user's Google Account if a user accesses a Google application in the same browser in which a 3rd-party webpage was previously accessed. Overall, our findings indicate that Google has the ability to connect the anonymous data collected through passive means with the personal information of the user.

Contents

II.   A DAY IN THE LIFE OF A GOOGLE USER

5.   To illustrate the multitude of touchpoints between Google and an individual, as well as the extent of information collected during these interactions, an experiment was designed where a researcher carried an Android mobile phone device[12] during a day's activities. The mobile phone was wiped by conducting a factory data reset[13] and configured as a new device to avoid prior user information associated with the device.[14] A new Google account was created (username "Jane"), so that Google had no prior knowledge of the user and had no advertising interests associated with the account. Researcher then went about a normal day using the mobile phone associated with the new Google account.

6.   The data collected by Google was checked using two tools provided by Google: My Activity[15] and Takeout.[16] The My Activity tool shows data collected by Google from any Search-related activities, use of Google applications (e.g. YouTube video plays, Maps search, Google Assistant), visits to 3rd-party web pages (while logged in to Chrome), and clicks on advertisements. The Google Takeout tool provides a more comprehensive information about all historical user data collected via Google's applications (e.g. it contains all past email messages on Gmail, search queries, location collection, and YouTube videos watched). We synthesized the collected data and used it to depict key information collection events in the form of a "day in the life" of the user "Jane," as shown in Figure 1.

7.   In the activity shown in Figure 1, as well as throughout the rest of this document, the collected data is categorized in two broad subgroups: *active* and *passive*. Active data is defined as information directly exchanged between the user and a Google product, whereas passive data is defined as information exchanged in the background without any obvious notification to the user. An example of active data collection occurred when Jane submitted a keyword in the Search tool bar and that search query was collected by Google. An example of passive data collection occurred when Jane's location was sent to Google after she entered a search query.

---

[12] LG X Power device with Android 6.0 version installed

[13] The factory data reset deletes all login data for Google services and other accounts, system and app data and settings, all downloaded apps, digital rights management licenses, music, images, documents and backups, and other usage data from the internal storage of the device.

[14] Researchers used LG X Power device that was wiped clean to the default factory settings and given new SIM card in order to ensure that no data was stored on the phone and that phone numbers could not be linked with any past usage.

[15] "My Activity," Google, available at https://myactivity.google.com/myactivity

[16] "Download your data," Google, available at https://takeout.google.com/settings/takeout?pli=1

Figure 1: Day in the life of Jane, highlighting touchpoints where Google collects data

A DAY IN THE LIFE OF A TYPICAL
GOOGLE USER

MORNING ACTIVITY

1. Gets ready in the morning while listening to Google Play Music

2. Drops kids off at daycare before commuting to work

3. Checks the news while commuting to work on the subway

4. Searches for cold medicine while on the subway

AFTERNOON ACTIVITY

5. Walks from subway to work

6. Uses Maps to find a new lunch spot

7. Gets coffee from Starbucks using her Starbucks app

8. Schedules doctors appointment, Google creates a Calendar event from the confirmation email

9. Walks to Walgreens and purchases cold medicine using Google Pay

EVENING ACTIVITY

10. Takes an Uber home from work

11. Looks for hotels on Expedia for a weekend trip

12. Uses Google Home to play music for her kids

13. Watches late night TV videos on YouTube

cnn.com
Visited CNNPolitics - Political News, Analysis and Opinion

WSJ wsj.com
Visited California Takes Big Step to Require Solar on New Homes - WSJ

G Search
Searched for cold symptoms

G Search
Searched for cold vs allergies

G Search
Searched for cold medicine

DUANE READE #14480
$12.99
DATE
May 9
TIME
6:50 PM
CARD USED
Visa ••••1041

Web site tracking tools

`<script async id="tealium-tag-7110" src="//www.google-analytics.com/analytics.js"></script>`

`<iframe src="https://bid.g.doubleclick.net/xbbe/pixel?d=KAE" style="display: none;"></iframe>`

Assistant
Said play lullaby   ▶ PLAY
9:25 PM • Details

Assistant
Said lower the volume   ▶ PLAY
9:25 PM • Details

Item details
YouTube
Watched Jimmy Fallon Recaps the Met Gala
The Tonight Show Starring Jimmy Fallon
Details
Today at 10:19 PM

8.   Analysis of key touch points during a normal day in the life of Jane suggested that the number of "passive" data collection events outnumbered the "active" events by approximately two-to-one (a detailed breakdown of characterization of active vs passive data collected appears in Table 3 of the Appendix section IX. A).

9.   Google analyzes the collected data to assess user interests, which it then applies to target users with appropriate ads. For example, Google provides a list of interests that it has inferred from a user's activities, available via the "topics you like" section in the Google's Ad Personalization webpage.[17] Figure 2 shows such a list that Google associated with Jane's account after a day's worth of activity. In total, Google attributed 18 interests to Jane, eight of which (shown by colored borders) closely matched Jane's usage and activities.[18]

Figure 2: Google's assessment of Jane's interest at the end of the day

Interests inferred by Google

Action & Platform Games	Books & Literature	Business & Industrial
Computer & Video Games	Computers & Electronics	Education
Games	Home & Garden	Mobile & Wireless
Movies	Music & Audio	News
Online Communities	Parenting	Science
Sports Games	TV & Video	TV Comedies
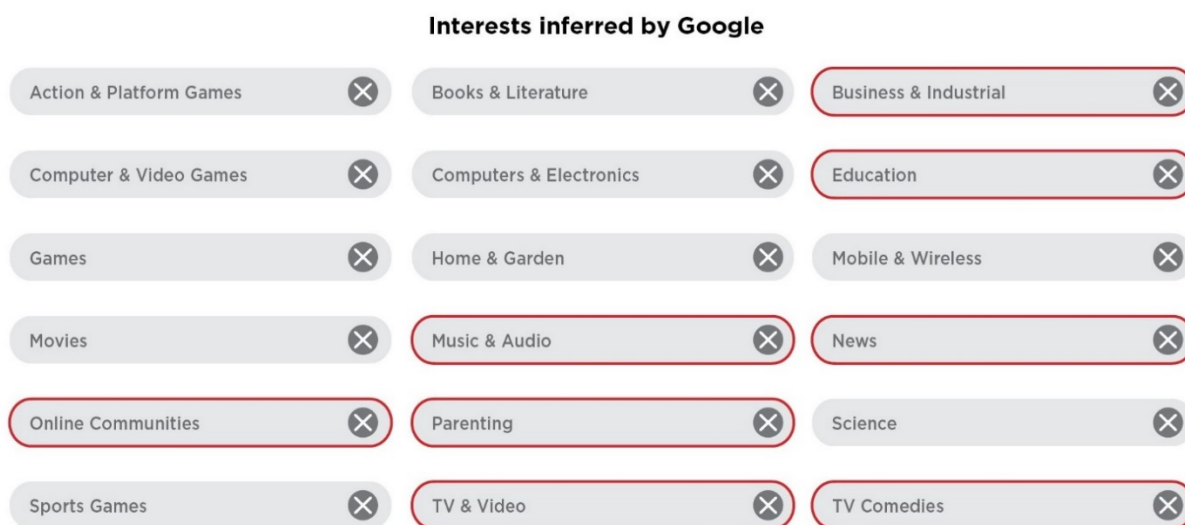
10.   Although My Activity and Takeout tools are helpful in assessing the amount of active data collected after a user interacts with Google's products, they do not paint a complete picture of the size and scale of Google's data collection. A comprehensive understanding of which requires a review of Google's product-specific privacy policies, as well as analyses of the actual data traffic passed to Google servers during the instances of a user's interaction with its services. Results derived from these resources are covered later in this report.

III.   DATA COLLECTION THROUGH ANDROID AND CHROME PLATFORMS

---

[17] "Ads personalization," *Google*, last accessed on August 15 2018, available at https://adssettings.google.com/authenticated
[18] It's unclear as to why other interests that have no connection with Jane's activities during the day show up in this list, though perhaps Google uses historical analysis of similar interests from other users to create associated recommendations.

11.   Android and Chrome are Google's key platforms that aid in significant user data collection due to their extensive reach and frequency of usage. By January 2018, Android captured 53% of the total US mobile OS market (Apple iOS held 45%)[19] and as of May 2017 there were more than 2 billion monthly active Android devices worldwide.[20]

12.   Google's Chrome browser held more than 60% share of all internet browser usage in the world with over 1 Billion monthly active users as reported in the 2017 Q4 10K filing.[21] Both platforms facilitate the use of Google and 3rd-party content (e.g. 3rd-party websites and 3rd-party apps) and hence provide Google access to a wide range of personal, web activity, and location information.

A.   Personal information and activity data collection

13.   To download and use apps from Google Play Store on an Android device, a user must have (or create) a Google Account, which becomes a key gateway through which Google collects personal information, including user name, email, and phone number. If a user registers for services such as Google Pay[22], Android also collects the user's credit card information, zip code, and birth date. All this information becomes part of a user's personal information associated with their Google Account.

14.   While Chrome does not mandate sharing additional personal information gathered from users, it does have the capability to capture such information. For example, Chrome collects a range of personal information via its form "autofill" feature, and such form fields typically include user name, address, phone number, login name, and passwords.[23] Chrome stores form fill information on a user's local drive, however, if the user logs in to Chrome using Google Account and enables its "Sync" feature, this information gets sent to and stored on Google servers. Chrome could also learn about the language(s) a person speaks during their interactions with its translate feature, which is enabled by default.[24]

15.   In addition to personal data, both Chrome and Android send Google information about a user's web browsing and mobile app activities, respectively. Any webpage visit is automatically tracked and collected under

---

[19] "Subscriber share held by smartphone operating systems in the United States from 2012 to 2018," *Statista,* May 2018, available at https://www.statista.com/statistics/266572/market-share-held-by-smartphone-platforms-in-the-united-states/
[20] Dave Burke, "Android: celebrating a big milestone together with you," *Google,* May 17, 2017, available at https://www.blog.google/products/android/2bn-milestone/
[21] Google 10K filings with the SEC
[22] "Google Chrome privacy whitepaper," *Google,* March 6, 2018, available at https://www.google.com/chrome/privacy/whitepaper.html#payments
[23] "Google Chrome privacy whitepaper," *Google,* March 6, 2018, available at https://www.google.com/chrome/privacy/whitepaper.html#autofill
[24] "Google Chrome privacy whitepaper," *Google,* March 6, 2018, available at https://www.google.com/chrome/privacy/whitepaper.html#translate
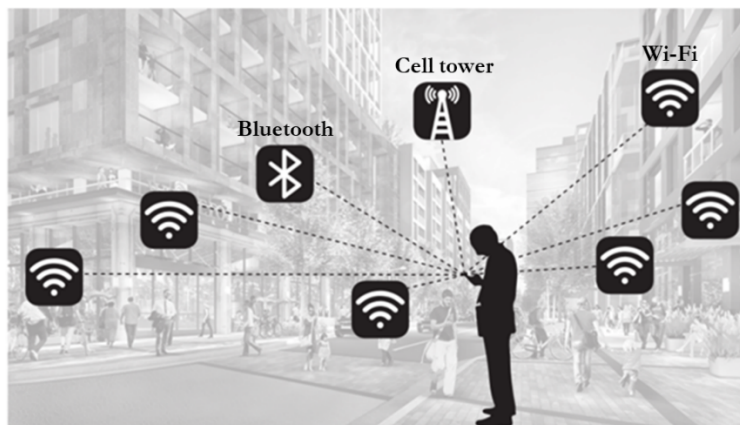
user credentials by Google if the user is signed in to Chrome. Chrome also collects information about a user's browsing history, passwords, website-specific permissions, cookies, download history, and add-on data.[25]

16.   Android sends periodic updates to Google servers, including device type, cell service carrier name, crash reports, and information about apps installed on the phone.[26] It also notifies Google whenever any app is accessed on the phone (e.g. Google knows when an Android user accesses their Uber app).

B.   User location data collection

17.   Android and Chrome platforms meticulously collect user location and movement information using a variety of sources, as depicted by Figure 3. For example, a "coarse location" assessment can be done by using GPS coordinates on an Android phone or through a network's IP address on a desktop/laptop device. The user location accuracy can be improved further ("fine location") through the use of nearby cell tower IDs or via scanning the device-specific BSSIDs or basic service set identifiers, assigned to the radio chipset used in nearby Wi-Fi access points.[27] Android phones can also use information from the Bluetooth beacons registered with Google's Proximity Beacon API.[28] These beacons not only provide user's geolocation coordinates, but could also pinpoint exact floor levels in buildings. [29]

Figure 3: Android and Chrome use multiple ways to locate a mobile user

---

[25] "Google Chrome Privacy Notice," *Google*, March 6, 2018, available at https://www.google.com/intl/en/chrome/browser/privacy
[26] https://policies.google.com/privacy?hl=en&gl=us#infocollect
[27] To understand how location data is sent to Google servers in more depth, our researchers analyzed the data traffic from a mobile phone from a user in motion, applying the method described in Appendix section VIII.C.
[28] "Google beacon platform, proximity beacon API," *Google,* last accessed on August 15 2018, available at https://developers.google.com/beacons/proximity/guides
[29] "Google beacon platform, proximity beacon API," *Google,* last accessed on August 15 2018, available at https://developers.google.com/beacons/proximity/guides

18.   It's hard for an Android mobile user to "opt out" of location tracking. For example, on an Android device, even if a user turns off the Wi-Fi, the device's location is still tracked via its Wi-Fi signal. To prevent such tracking, Wi-Fi scanning must be explicitly disabled in a separate user action, as shown in Figure 4.

Figure 4: Android collects data even if Wi-Fi is turned off by user
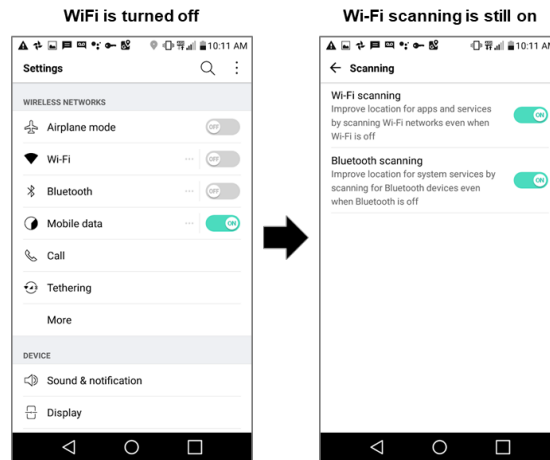
19.   The ubiquity of Wi-Fi hubs has made location tracking quite frequent. For example, during a short 15-minute walk around a residential neighborhood, an Android device sent nine location requests to Google. The request collectively contained ~100 unique BSSIDs of public and private Wi-Fi access points.

20.   Google can ascertain with a high degree of confidence whether a user is still, walking, running, bicycling, or riding on a train or a car. It achieves this by tracking an Android mobile user's location coordinates at frequent time intervals in combination with the data from onboard sensors (such as an accelerometer) on mobile phones. Figure 5 shows an example of such data communicated with the Google servers while the user was walking.

Figure 5: Snapshot from a Google user location upload

```
"activityReadings": [
    {
        "activities": [
            {
                "confidence": 99,
                "type": "onFoot"
            },
            {
                "confidence": 99,
                "type": "walking"
            },
            {
                "confidence": 1,
                "type": "unknown"
            }
        ],
        "timestampMs": 1527095517507
    },
```

### C. An assessment of passive data collection by Google through Android and Chrome

21.   Active data that Android or Chrome platforms collect and send to Google as a result of users' activities on these platforms can be assessed through the MyActivity and Takeout tools. Of potentially greater interest, however, is the passive data that these platforms collect, which goes beyond location data and which remains relatively unrecognized by the users. To assess the type and frequency of occurrence of such collection in greater detail an experiment was conducted that monitored traffic data sent to Google from mobile phones (both Android and iPhone) using the method discussed in Section IX.D in the Appendix. For comparison's sake, this experiment also included the analysis of data sent to Apple via an iPhone device.

22.   For simplicity, the phones were kept stationary, with no user interaction. On the Android phone a single Chrome browser session remained active in the background, whereas on the iPhone the Safari browser was used. This configuration provided an opportunity for systematic analysis of the background collection that Google performs purely through Android and Chrome, as well as collection that occurs in the absence of those (i.e. from iPhone device), without any additional collection requests generated by other products and applications (e.g. YouTube, Gmail, App usage).

23.   Figure 6 shows a summary of the results obtained from this experiment. The x-axis indicates the number of times the phones communicated with Google (or Apple) servers, whereas the y-axis indicates the phone type (Android or iPhone) and server domain type (Google or Apple) with which data packets were exchanged by the phones. The colored legend describes the broad categorization of the type of data requests identified by the domain address of the server. A complete list of domain addresses belonging within each category appears in Table 5 of Section IX.D of the Appendix.

24.   During a 24-hour time period the Android device communicated ~900 data samples to a variety of Google server endpoints. Of these, ~35% (or approximately 14/hour) were location-related. Google ad domains received only ~3% of the traffic, which is mainly due to the fact that the mobile browser was not actively used during the collection period. The remaining ~62% of communications with the Google server domains were roughly divided between requests to Google's Play App store, Android's uploads of device-related data (such as crash reports and device authorization), and other data which were predominantly in the category of Google services background calls and refreshes.

Figure 6: Traffic data sent from idle Android and iPhone mobiles
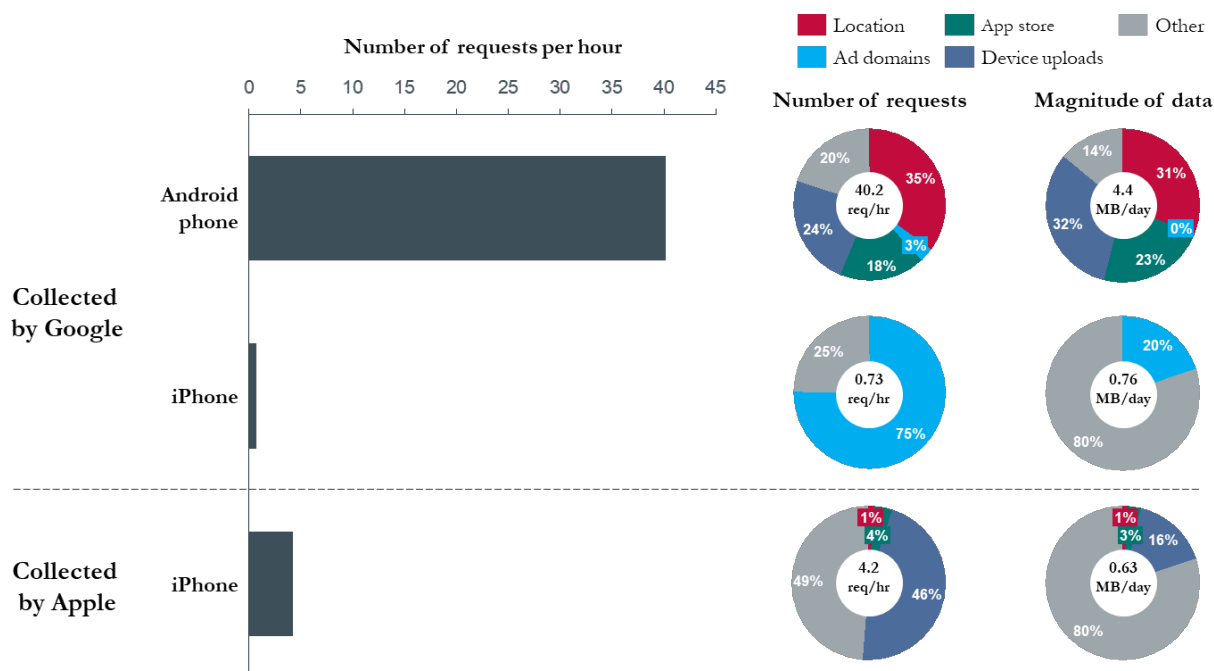
25.   Figure 6 shows that the iPhone device communicated with Google domains at more than an order of magnitude (~50x) lower frequency than the Android device, and that Google did not collect any user location during the 24-hour experiment timeframe via iPhone. This result highlights the fact that the Android and Chrome platforms play an important role in Google's data collection.

26.   Additionally, the iPhone device's communication with Apple's servers were 10x less frequent than the Android device's communications with Google. Location data made up a very small fraction (~1%) of the net data sent to Apple servers from the iPhone, with Apple receiving location-related communications once every day on an average.

27.   Magnitude wise, Android phones communicated 4.4 MB of data per day (~130MB per month) with Google servers, which is 6x more than what Google servers communicated through the iPhone device.

28.   As a reminder, this experiment was conducted using a stationary phone with no user interaction. As a user becomes mobile and starts interacting with their phone, the frequency of communications with Google's servers increases considerably. Section V of this report summarizes results from such an experiment.

IV.   DATA COLLECTION THROUGH PUBLISHER AND ADVERTISER TECHNOLOGIES

29.   A major source for Google's user activity data collection stems from its publisher- and advertiser-focused tools, such as Google Analytics, DoubleClick, AdSense, AdWords, and AdMob. These tools have tremendous reach, e.g. over 1 million mobile apps use AdMob,[30] over 1 million advertisers use AdWords,[31] over 15 million websites use AdSense,[32] and over 30 million websites use Google Analytics.[33]

30.   During the writing of this report Google rebranded AdWords as "Google Ads" and DoubleClick as "Google Ad Manager", however there were no changes instituted in the core product functionalities including information collection by these products.[34] Therefore, for the purpose of this report the names are kept unchanged to avoid confusion that may occur with related domain names (such as doubleclick.net).

31.   There are two main groups of users of Google's publisher- and advertiser-focused tools:

o   *Website and app publishers*, which are organizations that own websites and create mobile apps. These entities use Google's tools to (1) make money by allowing the display of ads to visitors on their websites or apps, and (2) better track and understand who is visiting their websites and using their apps. Google's tools place cookies and run scripts in the browsers of website visitors that help determine a user's identity, track their interest in content, and follow their online behavior. Google's mobile app libraries track use of apps on mobile phones.

o   *Advertisers*, which are organizations that pay to have banner, video, or other ads delivered to users as they browse the Internet or use apps. These entities apply Google's tools to target specific profiles of people for advertisements to increase the return on their marketing investments (better targeted ads generally yield higher click-through rates and conversions). Such tools also enable advertisers to analyze their audiences and measure the efficacy of their digital advertising by tracking which ads were clicked with what frequency and by providing insight into the profiles of people who clicked on ads.

---

[30] "AdMob by Google," *Google*, last accessed on August 15 2018, available at https://www.google.com/admob/
[31] "Hear form our happy customers," *Google,* last accessed on August 15 2018, available at
https://adwords.google.com/home/resources/success-stories/
[32] "Websites using Google Adsense," *BuiltWith,* last accessed on August 15 2018, available at
https://trends.builtwith.com/websitelist/Google-Adsense
[33] Google Analytics usage statistics," *BuiltWith,* April 2018, available at https://trends.builtwith.com/analytics/Google-Analytics
[34] Garett Sloane, "Google to retire Doubleclick and AdWords names in a rebrand of its as business," *Ad Age,* available at
http://adage.com/article/digital/google-waves-goodbye-doubleclick-ad-business-evolves/314046/

32. Together, these tools collect information about user activities on websites and in apps, such as content visited, and ads clicked. They work in the background—largely unnoticeable by users. Figure 7 shows some of these key tools, with arrows indicating data collected from users and ads served to users.

Figure 7: Google products aimed at publishers and advertisers[35]
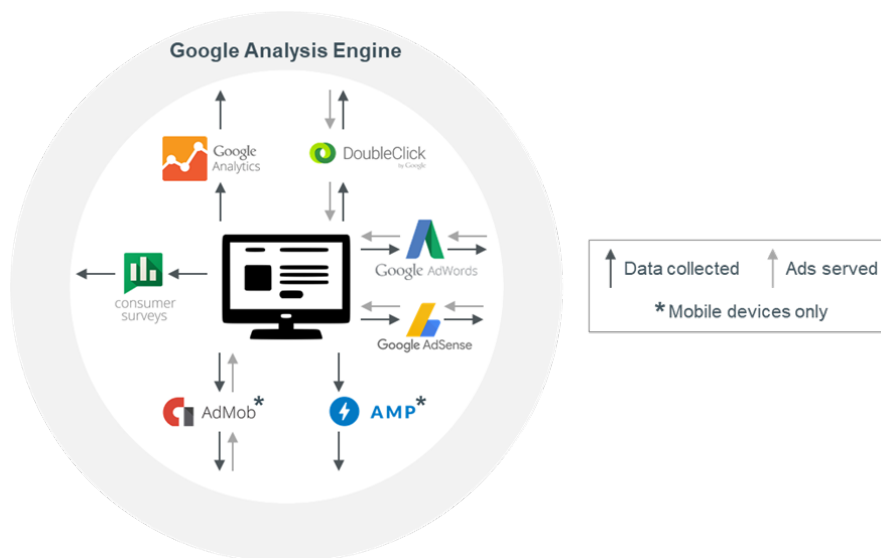
33. The information collected by such tools includes a non-personal identifier that Google can use to send targeted advertisements without identifying the unique individual's personal information. These identifiers can be device- or session-specific, as well as permanent or semi-permanent. Table 1 lists a set of such identifiers. To provide users greater anonymity during information collection for ad targeting, Google has recently shifted towards using semi-permanent device unique identifiers (e.g. GAIDs).[36] Further sections go in detail about how these tools collect user data and the use of such identifiers during the data collection process.

Table 1: Identifiers passed to Google

Identifier	Type	Description
GAID/IDFA	Semi-Permanent	Alphanumeric string for Android / iOS devices to allow targeted mobile ads. Resettable by users.
Client ID	Semi-Permanent	ID created the first time a cookie is stored on the browser. Used to link browsing sessions together. Resets when browser cookies are cleared.
IP address	Semi-Permanent	A unique string of number that identifies the network through which a device is accessing the Internet.

Android device ID	Semi-Permanent	Randomly generated number when a device is first booted up. Used to identify the device. It is in the process of being phased out of advertising. Resets with a factory reset of a device.
Google Services Framework (GSF)	Semi-Permanent	Randomly assigned number when a user first logs into Google services on a device. Used to identify a unique device. Resets with a factory reset of a device.
IEMI / MEID	Permanent	Identifier used in mobile communication standards. Unique for each mobile phone.
MAC address	Permanent	Unique 12-character identifier for a piece of hardware (e.g. router).
Serial number	Permanent	Alphanumeric string used to identify a device.

A. Google Analytics and DoubleClick

34. DoubleClick and Google Analytics (GA) are Google's leading products in user behavior tracking and webpage traffic analyses on desktop and mobile devices. GA is used by ~75% of the top 100,000 most visited websites.[37] DoubleClick cookies are associated with more than 1.6 million websites.[38]

35. GA uses short pieces of tracking code (called "page tags") embedded in a website's HTML code.[39] After a webpage loads per a user's request, the GA code calls an "analytics.js" file residing on Google's servers. This program transfers a "default" snapshot of user data at that moment, which includes visited webpage address, page title, browser information, current location (derived from IP address), and user language settings. GA scripts use cookies to track user behavior.

36. GA script, the first time when it's run, generates and stores a browser-specific cookie on the user's computer. This cookie has a unique client identifier or Client ID (see Table 1 for details).[40] Google uses the unique identifier to link previously stored cookies that capture a user's activity on a particular domain as long as the cookie does not expire, or the user does not clear the cookies cached on their browser.[41]

37. While a GA cookie is specific to the particular domain of the website that user visits (called a "1st-party cookie"), a DoubleClick cookie is typically associated with a common 3rd-party domain (such as

---

[37] Google Analytics usage statistics," *BuiltWith,* April 2018, available at https://trends.builtwith.com/analytics/Google-Analytics
[38] "DoubleClick market share," *Datanyze,* last accessed on August 15 2018, available at https://www.datanyze.com/market-share/ad-exchanges/doubleclick-market-share
[39] GA or other tags can also be implemented through Google Tag Manage (GTM) without changing the functionality of the page tag
[40] "Cookies and user identification," *Google*, last accessed on August 15 2018, available at https://developers.google.com/analytics/devguides/collection/analyticsjs/cookies-user-id
[41] "Cookies and user identification," *Google*, last accessed on August 15 2018, available at https://developers.google.com/analytics/devguides/collection/analyticsjs/cookies-user-id

doubleclick.net). Google uses such cookies to track user interaction across multiple 3rd-party websites.[42] When a user interacts with an advertisement on a website, DoubleClick's conversion tracking tools (e.g. Floodlight) places cookies on a user's computer and generates a unique client ID.[43] Thereafter, if the user visits the advertised website, the stored cookie information gets accessed by the DoubleClick server, thereby recording the visit as a valid conversion.

B. AdSense, AdWords and AdMob

38.   AdSense and AdWords are Google tools that serve ads on websites and in Google Search results, respectively. More than 15 million websites have AdSense installed to display sponsored ads.[44] Likewise, more than 2 million websites and apps that make up the Google Display Network (GDN) and reach over 90% of Internet users[45] display AdWords ads.

39.   AdSense collects information about whether an ad was displayed on the publisher's webpage. It also collects how the user interacted with the ad, such as clicking an ad or tracking the curser movement over an ad.[46] AdWords enables advertisers to serve search ads on Google Search, display ads on publisher pages, and overlay ads on YouTube videos. To track user click-through and conversion rates, AdWords ads place a cookie on users' browsers to identify the same user if they later visit the advertiser's website or complete a purchase.[47]

40.   While AdSense and AdWords collect data on mobile devices as well, their ability to get user information on mobile devices is limited since mobile apps do not share cookie data between them, an isolation technique known as 'sandboxing,'[48] which makes it hard for advertisers to track user behavior across mobile apps.

41.   To address this issue, Google and other companies use mobile "ad libraries" (such as AdMob) that are integrated into the apps by their developers for serving ads in mobile apps. These libraries compile and run with the apps and send to Google data that is specific to the app to which they belong, including GPS locations, device make, and device model when apps have the appropriate permissions. As observed through the data

---

[42] "DoubleClick search help," *Google,* last accessed on August 15 2018, available at
https://support.google.com/ds/answer/7298761?hl=en
[43] "DoubleClick search help," *Google,* last accessed on August 15 2018, available at
https://support.google.com/ds/answer/2903014?hl=en&ref_topic=6054260
[44] "Websites using Google Adsense," *BuiltWith,* last accessed on August 15 2018, available at
https://trends.builtwith.com/websitelist/Google-Adsense
[45] "Google Ads Help," *Google,* last accessed on August 15 2018, available at https://support.google.com/google-ads/answer/2404191?hl=en
[46] "AdSense help, privacy and security," *Google,* last accessed on August 15 2018, available at
https://support.google.com/adsense/answer/9897?hl=en
[47] "Evaluating ad performance on the Search Network," *Google,* last accessed on August 15 2018, available at
https://support.google.com/adwords/answer/2404037?hl=en; "About conversion tracking," *Google,* last accessed on August 15 2018, available at https://support.google.com/adwords/answer/1722022?hl=en
[48] This approach is similar to desktops, where cookies are not shared between browsers.

traffic analyses (Figure 8), and confirmed through Google's own developer webpages,[49] such libraries can also send user-personal data, such as age and gender, to Google whenever app developers explicitly pass these values to the library.

Figure 8: Snapshot of information sent back to Google when an application is launched
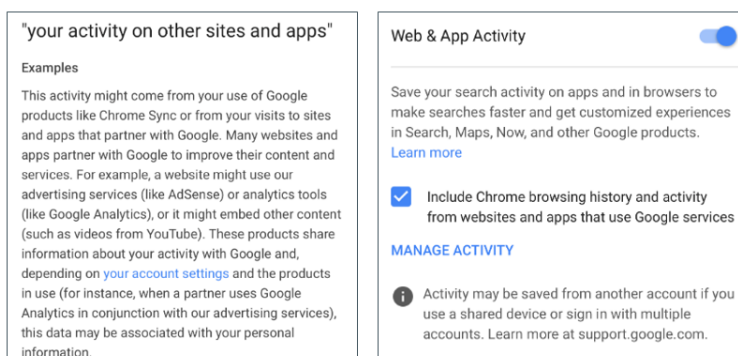

C. Association of passively collected data with personal information

42. As discussed above, Google collects data through publisher and advertiser products and associate such data with a variety of semi-permanent, anonymous identifiers. Google however, has the ability to associate these IDs with a user's personal information. This is insinuated by the statements made in Google's privacy policy, excerpts of which are shown in Figure 9. The left text box clearly states that Google may associate data from advertising services and analytics tools with a user's personal information, depending upon the user's account settings. This arrangement is enabled by default, as shown in the right text box.

Figure 9: Google's privacy page for 3rd-party websites collection and association with personal information[5051]

---

[49] "Google APIs for Android," *Google,* last accessed on August 15 2018, available at
https://developers.google.com/android/reference/com/google/android/gms/ads/doubleclick/PublisherAdRequest.Builder
[50] "Google privacy and terms," *Google,* last accessed on August 15 2018, available at
https://policies.google.com/privacy/example/your-activity-on-other-sites-and-apps
[51] "Google https://myaccount.google.com/activitycontrols

43.   Furthermore, an analysis of data traffic exchanged with Google servers (summarized below) identified two key examples (one on Android and the other in Chrome) that point to Google's ability to correlate anonymously collected data with users' personal information.

1)   Mobile advertising identifier may get de-anonymized through data sent to Google by Android

44.   Analyses of data traffic communicated between an Android phone and Google server domains suggest a possible way through which anonymous identifiers (GAID in this case) can get associated with a user's Google Account. Figure 10 describes this process through a series of three key steps.

45.   In step 1, a "checkin" data is sent to the URL *android.clients.google.com/checkin*. This particular communication provides an Android data sync to Google servers and contains Android log information (e.g. recovery log), kernel messages, crash dumps, and other device-related identifiers. A snapshot of a partially decoded checkin request sent to Google's server from Android is shown in Figure 10.

Figure 10: Device identifiers are sent together with account information in Android check in requests
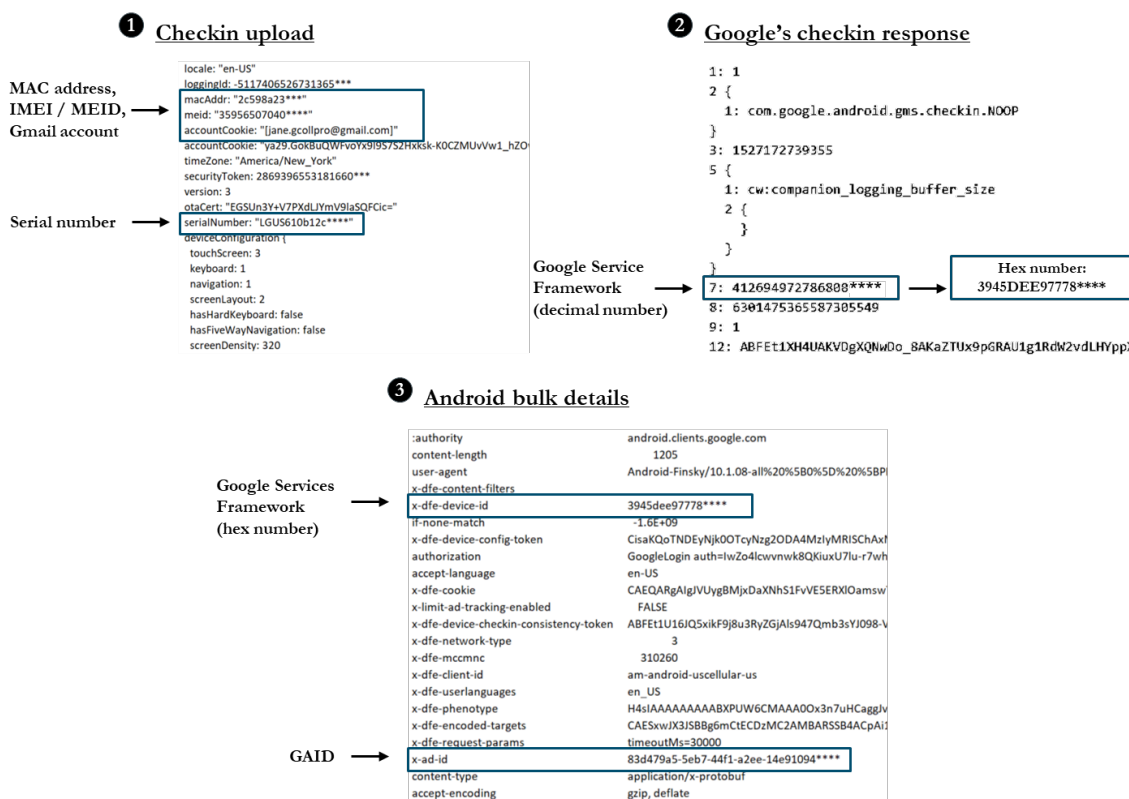
46.   As pointed out by the labeled boxes, through the checkin process, Android sends to Google a variety of important device-related permanent identifiers, including device MAC address, IMEI /MEID, and device

serial number. Moreover, these requests also contain the Android user's Gmail ID. The data present in checkin uploads enable Google to connect a user's personal information with Android device permanent identifiers.

47. In step 2, the reply to the checkin request comes from the Google server. This message contains a Google services framework identifier (GSF ID)[52] that is similar to the actual "Android ID"[53] (see Table 1 for descriptions).

48. Step 3 entails another instance of communication where the same GSF ID (from step 2) is sent to Google together with the GAID. Figure 10 shows one such data transmit to *android.clients.google.com/fdfe/bulkDetails ?au=1*.

49. Through the above three data exchanges, Google receives the information needed to connect a GAID with permanent device identifiers as well as users' Google Account IDs.

50. These intercepted data exchanges with Google servers from an Android phone show how Google can connect anonymized information collected on an Android mobile device via DoubleClick, Analytics or AdMob tools with the user's personal identity. During the 24-hour data collection from a stationary and dormant Android phone two instances of checkin communications with Google servers were observed. Additional analysis is needed, however, to determine if such information exchange occurs with a certain periodicity or if it is triggered by specific activities on the phones.

2) DoubleClick cookie ID gets connected with user's personal information on Google Account

51. The previous section explained how Google can de-anonymize user identity via the passive, anonymized data it collects from an Android mobile device. This section shows how such de-anonymization can also occur on a desktop/laptop device.

52. Anonymized data on desktops/laptops is collected via cookie-based identifiers (e.g. Cookie ID), which are typically generated by Google's ad and publisher products (e.g. DoubleClick) and stored on a user's local mass storage. The experiment presented below assessed whether Google can connect such identifiers (and hence information associated with them) with a user's personal information. This experiment involved the following ordered steps:

    1. Opened a new (no saved cookies, e.g. Private or Incognito) browser session (Chrome or other),

---

[52] "Difference between Android ID and device ID," *Stack Exchange,* Dec. 2016, available at
https://android.stackexchange.com/questions/162448/difference-between-android-id-and-device-id
[53] Patrick Ahlbrecht, "What's the difference between the GSF ID and the Android ID," *Onyxbits,* March 2016, available at
https://blog.onyxbits.de/whats-the-difference-between-the-gsf-id-and-the-android-id-208/

2. Visited a 3rd-party website that used Google's DoubleClick ad network,

3. Visited the website of a widely used Google service (Gmail in this case),

4. Signed in to Gmail.

53.   After completion of step 1 and 2, as part of the page load process, the DoubleClick server received a request when the user first visited the 3rd-party website. This request was part of a series of requests comprising the DoubleClick initialization process started by the publisher website, which resulted in the Chrome browser setting a cookie for the DoubleClick domain. This cookie stayed on user's computer until it expired or until the user manually cleared coolies via the browser settings.

54.   Thereafter, in step 3, when the user visited Gmail, they are prompted to log in with their Google credentials. Google manages identity using a "single sign on (SSO)" architecture, whereby credentials are supplied to an account service (signified by *accounts.google.com*) in exchange for an "authentication token," which can then be presented to other Google services to identify the users. In step 4, when a user accesses their Gmail account, they are effectively signing into their Google Account, which then provides Gmail with an authorization token to verify the user's identity. [54] This process is outlined by Figure 24 in Section IX.E in the Appendix.

55.   In the last step of this sign-on process, a request is sent to the DoubleClick domain. This request contains both the authentication token provided by Google and the tracking cookie set when the user visited the 3rd-party website in step 2 (this communication is shown in Figure 11). This allows Google to connect the user's Google credentials with a DoubleClick cookie ID. Therefore, if the users do not clear browser cookies regularly, their browsing information on 3rd-party webpages that use DoubleClick services could get associated with their personal information on Google Account.

Figure 11: Request to DoubleClick.net includes Google's authentication token and past cookies

---

[54] The advantage of the extra authentication step is that the user's browser can later use the same authentication token to confirm user identity on other Google services (due to this process a sign-on in any particular Google application enables an automatic sign-on all others in the same browser session).

56.   It has thus far been established that Google collects a wide variety of user data through its publisher and advertiser tools, without a direct knowledge of the user. While such data is collected with user-anonymous identifiers, Google has the ability to connect this collected information with a user's personal credentials stored in their Google Account.

57.   It's worth pointing out that Google's passive user data collection from 3rd-party webpages cannot be prevented using popular ad blocking tools,[55] as such tools are designed primarily to prevent the occurrence of advertisements while users browse through 3rd-party webpages.[56] The next section takes a closer look at the magnitude of such data collection.

V.   AMOUNT OF DATA COLLECTED DURING A MINIMAL USE OF GOOGLE PRODUCTS

58.   This section examines the details surrounding Google's data collection through its publisher and advertiser services. To understand such data collection, an experiment is designed which entailed a user going through her daily life using a mobile phone (akin to "day in the life" described before), while deliberately *avoiding* the use of any direct Google products (i.e. avoiding Search, Gmail, YouTube, Maps, etc.), except for the Chrome browser.

59.   To keep the experiment as realistic as possible, various consumer usage studies[57,58] were used to form a daily usage profile of a typical mobile phone user, thereafter, any direct interactions with Google's products were omitted from the profile. Section IX.F in the Appendix describes the websites and apps used in this experiment.

60.   The experiment was replicated on both Android and iOS devices and the HTTPS data sent to Google and Apple servers were monitored and analyzed using a similar method explained in previous sections. The results are summarized in Figure 12. During the 24-hour time period (which includes the night time stationary/dormant timeframe), the majority of calls from the Android phone were made to Google's location and publisher/advertisement service domains (e.g. DoubleClick, Analytics). Google collected user location in ~450 instances, which is ~1.4x times the experiment presented in Section III.C, which involved a stationary phone.

---

[55] "How many users block Google Analytics, measured in Google Analytics," *Quantable,* Dec. 2017, available at https://www.quantable.com/analytics/how-many-users-block-google-analytics/
[56] "Ad blocking: who blocks ads, why and how to win them back," *iab.,* 2016, available at https://www.iab.com/wp-content/uploads/2016/07/IAB-Ad-Blocking-2016-Who-Blocks-Ads-Why-and-How-to-Win-Them-Back.pdf
[57] The average person visited 88 webpages per day in 2010. "Nielsen provides topline U.S. web data for March 2010," *Nielsen,* April 2010, available at http://www.nielsen.com/us/en/insights/news/2010/nielsen-provides-topline-u-s-web-data-for-march-2010.html
[58] 55% of web traffic comes from mobile devices. Eric Enge, "Mobile vs desktop usage: mobile grows but desktop still a big player in 2017," *Stone Temple,* April 2017, available at https://www.stonetemple.com/mobile-vs-desktop-usage-mobile-grows-but-desktop-still-a-big-player-in-2017/

Figure 12: Information requests from mobile devices during a day of typical use
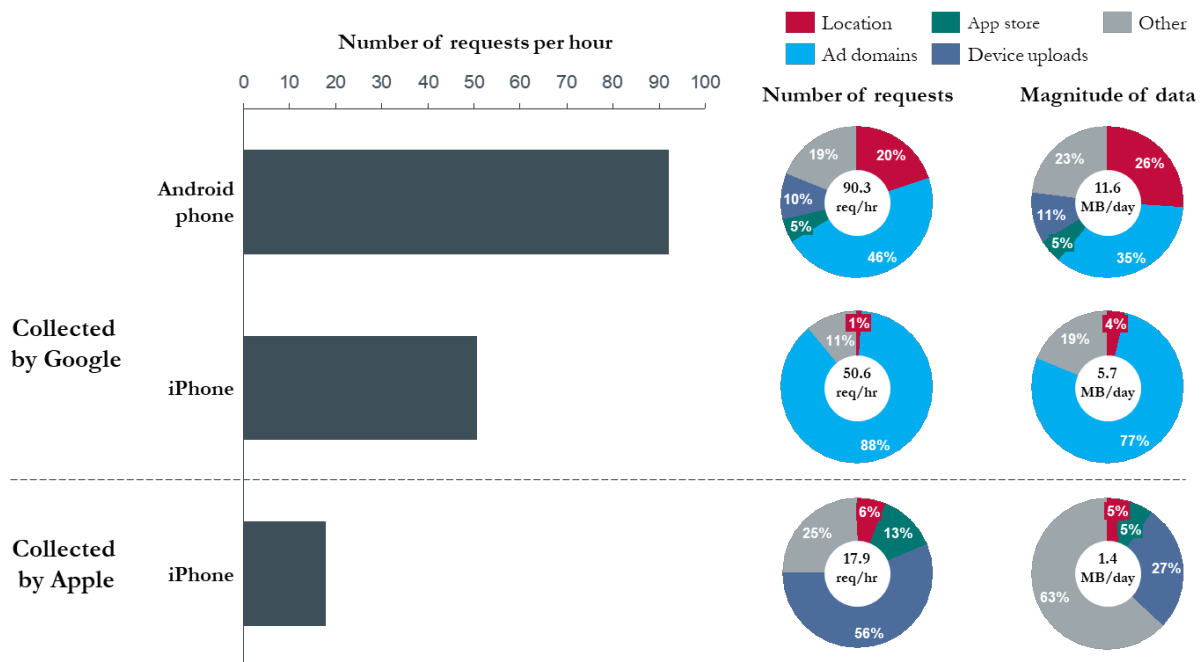
61. Google servers communicated significantly lower number of times with an iPhone device compared to Android (45% less). However, the number of calls to Google's advertising domains were similar from both devices - an expected outcome since the usage of 3rd-party webpages and apps was similar on both devices. One notable difference was that the location data sent to Google from an iOS device is practically non-existent. In the absence of Android and Chrome platforms—or the use of any other Google product—Google becomes significantly limited in its ability to track the user location.

62. The total number of calls to Apple servers from an iOS device was much lower, just 19% the number of calls to Google servers from an Android device. Moreover, there are no ad-related calls to Apple servers, which may stem from the fact that Apple's business model is not as dependent on advertising as Google's. Although Apple does obtain some user location data from iOS devices, the volume of data collected is much (16x) lower than what Google collects from Android.

63. Magnitude wise, Android phones communicated 11.6 MB of data per day (~350 MB per month) with Google servers. On the other hand, the iPhone device communicated just half that amount. The amount of data particularly associated with Google's ad domains remained very similar across both the devices.

64. The iPhone device communicated an order of magnitude less data to Apple servers than what the Android device exchanged with Google servers.

65. Overall, even in the absence of user interaction with Google's most popular applications, a user of an Android phone and the Chrome browser still sends a significant amount of data to Google, the majority of which is associated with location and calls to ad server domains. Although an iPhone user is insulated from Google's location collection in this narrow experiment, Google still captures a similar amount of ad-related data.

66. The next section describes the data collected by Google's popular applications, such as Gmail, YouTube, Maps, and Search.

VI.   DATA COLLECTED FROM GOOGLE'S KEY POPULAR APPLICATIONS AIMED AT INDIVIDUALS

67. Google has dozens of constantly evolving products and services (a list is available in Table 4 in Section IX.B of the Appendix). These products are often accessed through—or associated with—a Google Account, which enables Google to directly link user activity details from its application-oriented products and services to a user profile. In addition to product usage data, Google also collects device-related identifiers and location data when Google's products and services are accessed.[59]

68. Some of Google's applications (e.g. YouTube, Search, Gmail, and Maps) are central to the basic tasks that many users conduct daily through their desktop or mobile devices. Table 2 describes the reach of these key products. This section explains how each of these prominent applications collect user information.

Table 2: Worldwide reach of Google's top application products

Product	Active users
Search	Greater than 1B monthly active users, 90.6% search engine market share[60]
YouTube	Greater than 1.8 billion logged-in monthly active users[61]
Maps	Greater than 1 billion monthly active users[62]
Gmail	1.2 billion registered users[63]

[59] "Google privacy and terms," *Google,* last accessed on August 15 2018, available at https://policies.google.com/privacy

[60] "Search engine market share worldwide," *StatCounter Global Stats,* April 2018, available at http://gs.statcounter.com/search-engine-market-share#monthly-201704-201804

[61] Devindra Hardawar, "YouTube gets 1.8 billion logged-in viewers monthly," *Engadget,* May 3, 2018, available at https://www.engadget.com/2018/05/03/youtube-1-8-billion-viewers/

[62] Google 10K filings with the SEC, 2017, available at https://abc.xyz/investor/pdf/20171231_alphabet_10K.pdf

[63] Motek Moyen, "Gmail is very popular but Google still won't fix a security vulnerability," *Seeking Alpha,* July 17, 2017, available at https://seekingalpha.com/article/4088241-gmail-popular-google-still-fix-security-vulnerability

A.   Search

69.   Google Search is the most popular web search engine in the world,[64] with over 11 billion search queries per month in the United States alone.[65] In addition to serving ranked webpage results in response to users' general queries, Google operates other search-based tools, such as Google Finance, Flights, News, Scholar, Patents, Books, Images, Videos, and Hotels. Google uses its search products to collect data related to search queries, browsing history, and ad-click/purchase activity. For example, Google Finance collects information on the type of stocks users may be tracking, whereas Google Flights tracks users' travel bookings and search requests.

70.   Whenever Search is used, Google collects location data via various means of assessing locations on mobile or desktop devices, as discussed in previous sections. Google records all search activity a user conducts and links it back to their Google Account if the user is logged in. Figure 13 shows an example of information collected by Google about a user's keyword search and page visit.

Figure 13: An example search data collection taken from user's My Activity page
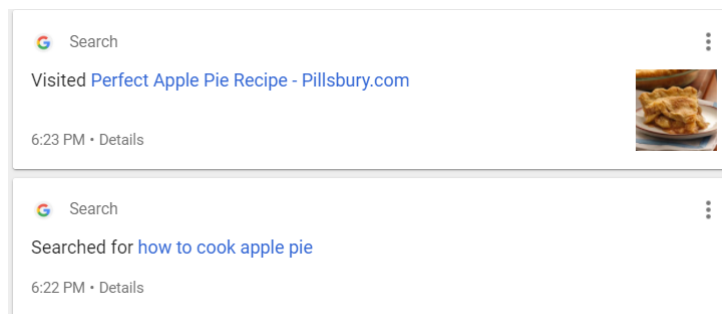
71.   In addition to being the default search engine on Chrome and Google devices, Google Search is also the default option on other 3rd-party browsers and applications through various distribution agreements. For example, Google recently became the default search engine on Mozilla's Firefox browser[66] in key geographic locations (including US and Canada), a position owned by Yahoo previously. Similarly, Apple switched from

---

[64] "Search engine market share worldwide," *StatCounter Global Stats,* April 2018, available at http://gs.statcounter.com/search-engine-market-share#monthly-201704-201804
[65] "Number of explicit core search queries powered by search engines in the United States as of January 2018 (in billions)," *Statista,* Feb. 2018, available at https://www.statista.com/statistics/265796/us-search-engines-ranked-by-number-of-core-searches/
[66] Denelle Dixon, "Firefox features Google as default search provider in the U.S., Canada, Hong Kong and Taiwan," *The Mozilla Blog,* Nov. 14, 2017, available at https://blog.mozilla.org/blog/2017/11/14/firefox-features-google-as-default-search-provider-in-the-u-s-canada-hong-kong-and-taiwan/

Microsoft's Bing to Google for Siri web search results on iOS and Mac devices.[67] Google has similar agreements in place with OEMs,[68] which helps reach mobile customers.

B.   YouTube

72.   YouTube provides users a platform for uploading and viewing video content. It has more than 180 million users in the USA alone and has the distinction of being the second-most visited website in the US,[69] ranked only behind Google Search. Among online streaming media companies, YouTube has almost 80% market share in terms of monthly user visits (as shown in Figure 14). The amount of content uploaded and viewed on YouTube is substantial; ~400 hours of video are uploaded every minute[70] and ~1 billion hours of video are watched daily on the YouTube platform.[71]

Figure 14: Comparison of leading multimedia websites monthly visits in the United States[72]
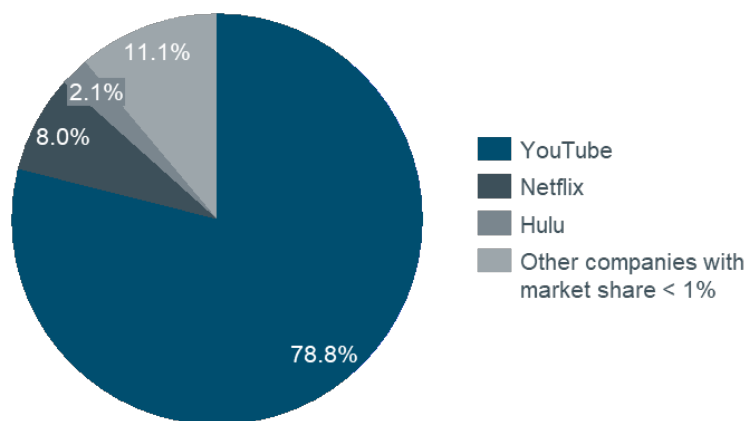
73.   YouTube can be accessed by users via desktops (web browser), mobile devices (app and/or web browser), and Google Home (through a paid subscription service called YouTube Red). Google collects and stores search history, watch history, playlists, subscriptions, and comments on videos. All this information is marked with a date and time stamp of when the activity took place.

---

[67] Matthew Panzarino, "Apple switches from Bing to Google for Siri web search results on iOS and Spotlight on Mac," Sept. 25, 2017, available at https://techcrunch.com/2017/09/25/apple-switches-from-bing-to-google-for-siri-web-search-results-on-ios-and-spotlight-on-mac/
[68] "Google's Android mobile application distribution agreement with OEMs leaked, reveals lots of strict conditions," *Microsoft and Technology News,* Feb. 13, 2014, available at https://mspoweruser.com/mobile-application-distribution-agreement/
[69] "Top Sites in United States," *Alexa,* available at https://www.alexa.com/topsites/countries/US
[70] "Hours of video uploaded to YouTube every minute as of July 2015," *Statista,* July 2015, available at https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/
[71] Darrell Etherington, "People now watch 1 billion hours of YouTube per day," *TeachCrunch*, Feb. 28, 2017, available at https://techcrunch.com/2017/02/28/people-now-watch-1-billion-hours-of-youtube-per-day/
[72] "Leading multimedia websites in the United States in November 2016, based on market share of visits," *Statista,* Dec. 2016, available at https://www.statista.com/statistics/266201/us-market-share-of-leading-internet-video-portals/

74.   If a user is signs into their Google Account on any Google application inside a browser (e.g. Chrome, Firefox, Safari), Google recognizes the user's identity, even if the video is accessed through a non-Google website (e.g. YouTube videos played through CNN.com). This feature allows Google to track a user's YouTube usage across multiple 3rd-party platforms. Figure 15 shows an example of YouTube data collected.

Figure 15: An example of YouTube data collection from My Activity
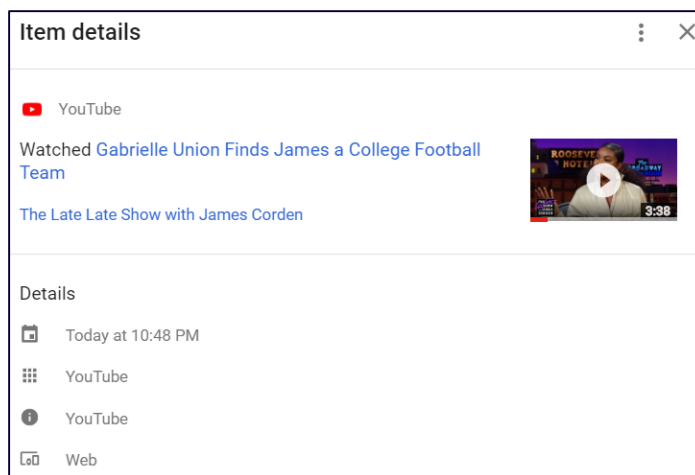
75.   Google also offers a separate YouTube product for children, known as YouTube Kids, which is intended as a "family friendly" version of YouTube with parental control features and video filters. Google collects information from YouTube Kids, including device type, operating system, unique device identifiers, log information, and details of how the service was used. Google then uses this information to deliver limited advertisements that are non-clickable, and which have restrictions on format, time length, and site-served.[73]

C.   Maps

76.   Maps is Google's flagship navigation app. Google Maps can ascertain user's travel routes, speed, and places that a user visits frequently (e.g. home, work, restaurants, and businesses). This information provides Google with a window into a user's interests (e.g. food and shopping preferences), movement, and behavior.

77.   Maps uses IP address, GPS, cell signal, and Wi-Fi access point data to calculate a device's location. The latter two are collected from the device through which Maps is used and sent to Google for location assessment through its Location API. This API provides rich details about a user, including geographic coordinates, whether the user is stationary or moving, speed, and probabilistic determination of user's mode of transport (e.g. bike, car, train, etc.).

---

[73] "Advertising on YouTube Kids," *Google,* last accessed on August 15 2018, available at https://support.google.com/youtube/answer/6168681?hl=en

78.   Maps stores a historical timeline of places visited by a user signed in to Maps using their Google account. Figure 16 shows an example of such a user's timeline.[74] The red dots indicate location coordinates captured by Maps while the user is on the move; the blue connecting lines are Maps projection of the actual route the user took.

Figure 16: Example Google Maps "Timeline" from an actual user
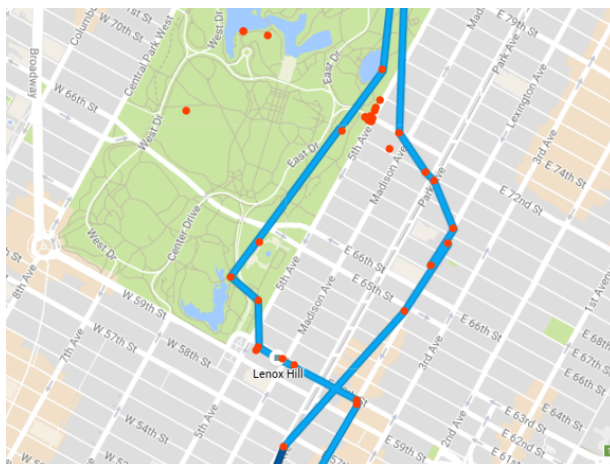
79.   The accuracy of location information captured by navigation applications enables Google to not only target ad audiences, but also helps deliver ads to users as they approach stores.[75] In addition, Google Maps uses this information to generate real-time traffic updates.[76]

D.   Gmail

80.   Gmail stores all messages (sent/received), sender name, email address, and date/time of messages sent or received. Since Gmail acts as a central mail repository for many people, it can ascertain their interests by scanning email content, identifying merchant addresses through their promotional emails or sales receipts sent to emails, and learn about a user's plans (e.g. dinner reservations, doctor's appointments,). Since users may use their Gmail ID for other 3rd-party platforms (e.g. Facebook, LinkedIn), Gmail can scan any content that comes from them in the form of an email (e.g. notifications, messages).

81.   From its inception in 2004 until at least late 2017, Google may have scanned the contents of Gmail emails to improve ad targeting and search results, as well as filter spam. In the summer of 2016, Google went a step further and changed its privacy policy to enable it to combine formerly anonymous web-browsing data

---

[74] "My Activity," *Google*, available at https://myactivity.google.com/myactivity
[75] "The Home Depot earns 8X in-store ROI with mobile display ads," *Google,* Sept. 2016, available at
https://www.thinkwithgoogle.com/intl/en-aunz/advertising-channels/mobile/home-depot-roi-mobile-display-ads/
[76] "Google Map's real-time traffic layer…," *Spatial Unlimited,* March 2011, available at
https://shreerangpatwardhan.blogspot.com/2011/03/google-maps-real-time-traffic-layer.html

of its subsidiary DoubleClick (which serves customized ads across the Internet) with the personally-identifying data Google has through its other products, including Gmail.[77] The result was that "…the DoubleClick ads that follow people around on the web may now be customized to them based on keywords they used in their Gmail. It also meant that Google could now build a complete portrait of a user by name, based on everything they write in email, every website they visit, and the searches they conduct."[78]

82.   Toward the end of 2017, Google announced it would discontinue the practice of Gmail message-based personalization of ads.[79] Recently, however, Google clarified that it is still scanning Gmail messages for some purposes.[80]

VII.   PRODUCTS WITH HIGH FUTURE POTENTIAL FOR DATA AGGREGATION

83.   Google has additional products that show future potential for market adoption and data collection, including AMP, Photos, Chromebook, Assistant, and Pay. Additionally, Google is able to use third party data vendors to collect user information. The following sections describe these in greater detail.

84.   There are other Google applications that may not be widely used, however for completeness, data collection through them is presented in Section VIII.B of the Appendix.

A.   Accelerated Mobile Pages (AMP)

85.   Accelerated Mobile Pages (AMP) is an open-source initiative spearheaded by Google to enable quicker load times for websites and ads. AMP converts conventional HTML and JavaScript code into a more simplified version developed by Google[81] and caches the AMP-validated webpages in Google's network of servers for faster access.[82] AMP delivers page links through Google search results, as well as 3rd-party platforms, such as LinkedIn and Twitter. As the AMP page reports: "AMP's ecosystem includes 25 million domains, 100+ technology providers and leading platforms, that span the areas of publishing, advertising, e-commerce, local and small businesses, and more!"[83]

---

[77] Julia Angwin, "Google has quietly dropped ban on personally identifiable web tracking," *ProPublica,* Oct. 21, 2016, available at
https://www.propublica.org/article/google-has-quietly-dropped-ban-on-personally-identifiable-web-tracking
[78] Suzanne Monyak, "Google changed a major privacy policy four months ago, and no one really noticed," *Slate,* Oct. 21, 2016,
available at
http://www.slate.com/blogs/future_tense/2016/10/21/google_changed_a_major_privacy_policy_and_no_one_really_noticed.html
[79] Mark Bergen, "Google will stop reading your emails for Gmail ads," *Bloomberg,* June 23, 2017, available at
https://www.bloomberg.com/news/articles/2017-06-23/google-will-stop-reading-your-emails-for-gmail-ads
[80] Ben Popken, "Google sells the future, powered by your personal data," *NBC News,* May 10, 2018, available at
https://www.nbcnews.com/tech/tech-news/google-sells-future-powered-your-personal-data-n870501
[81] "AMP HTML specification," *AMP,* last accessed on August 15 2018, available at
https://www.ampproject.org/docs/fundamentals/spec
[82] "Load AMP pages quickly with Google AMP Cache," *Google,* last accessed on August 15 2018, available at
https://developers.google.com/amp/cache
[83] "An open source effort to improve the content ecosystem for everyone," *AMP*, last accessed on August 15 2018, available at
https://www.ampproject.org/learn/overview

86.   Figure 17a describes the steps leading to the delivery of an AMP page accessed via Google Search. Please note that the provider of content through AMP does not need to provide their own a cache server, as this is something that Google provides for securing optimal delivery speeds to users. Since the AMP cache is hosted on Google servers, when an AMP link is clicked through Google Search, the domain address shows up from a Google.com domain rather than from a publisher's own domain. This is shown through snapshots taken from an example keyword search in Figure 17b.

Figure 17: Regular web page vs AMP page
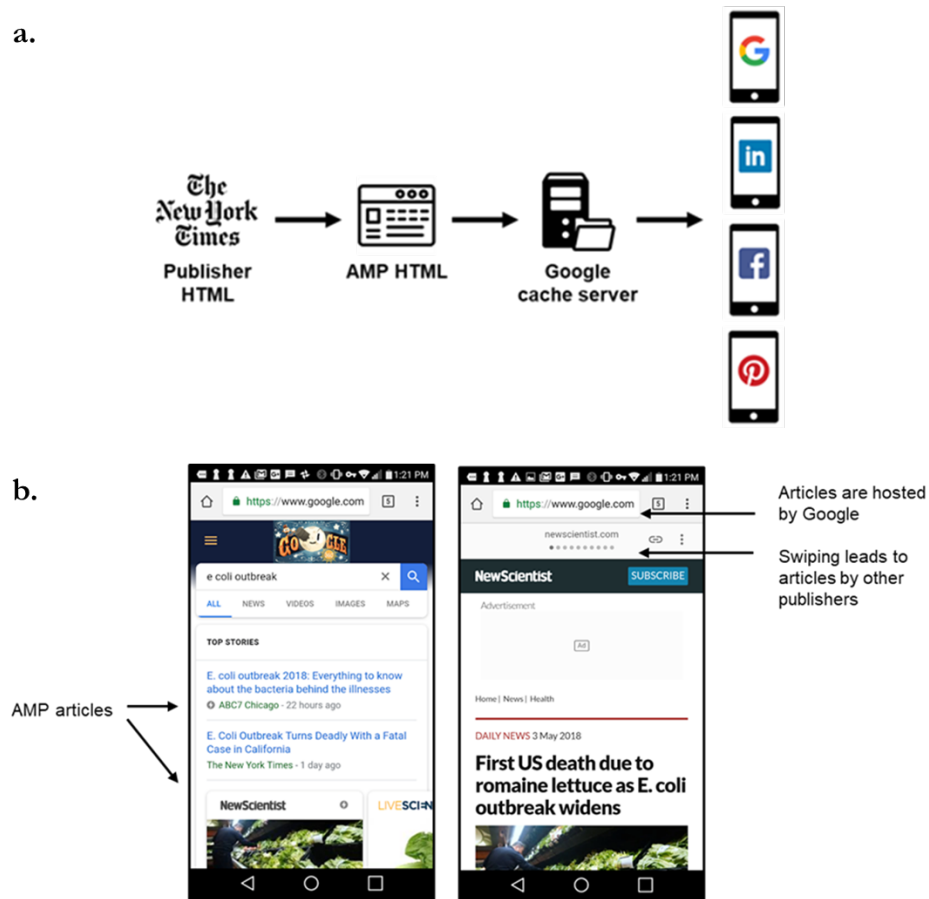
87.   Users can access content from multiple publishers whose articles appear in search results while navigating the AMP carousel, all while staying within the Google domain. In effect, the AMP cache operates as a content delivery network (CDN) owned and operated by Google.

88.   By creating an open-source tool, complete with a CDN, Google has attracted a large user base for serving mobile websites and advertisements that constitute a significant amount of information (e.g. the content itself, page views, ads served, and information on whom that content is being delivered). All of this information

31

is available to Google by virtue of it residing on Google's CDN servers, thereby providing Google far more data than it otherwise could access.

89.   AMP is highly user-centric, i.e. it delivers a much faster and improved browsing experience to users without the clutter of pop-ups and sidebars. Although AMP is a major shift in the way content is cached and delivered to users, Google's privacy policy associated with AMP is quite general.[84] In particular, Google is able to collect webpage usage information (e.g. server logs and IP address) from requests sent to AMP cache servers. Moreover, regular pages are converted into AMP via the use of AMP APIs.[85] Google can therefore access applications or websites ("API clients") and use any submitted information through the API in accordance with its general policies.[86]

90.   Like regular webpages, AMP webpages track usage data via Google Analytics and DoubleClick. In particular, they collect information on page data (e.g. domain, path, and page title), user data (e.g. client ID, time zone), browsing data (e.g. unique page view ID and referrer), browser info, and interaction and events data.[87] Although Google's modes of data collection have not changed with AMP, the *amount* of data collected has increased since visitors are spending 35% more time on web content that loads with Google AMP versus standard mobile pages.[88]

B.   Google Assistant

91.   Google Assistant is a virtual personal assistant accessed through mobile phones and smart devices. It is a popular virtual assistant, alongside Apple's Siri, Amazon's Alexa, and Microsoft's Cortana.[89] Google Assistant is accessed through the home button of mobile devices with Android 6.0 or higher. It can also be accessed through a dedicated app on iOS devices[90], as well as smart speakers, such as Google Home. Google Assistant performs numerous functions, such as sending texts, looking up emails, controlling music, searching photos, getting answers to questions about the weather or traffic, and controlling smart home devices.[91]

92.   Google collects all Google Assistant queries, whether audio or typed. It also collects the location where the query occurred. Figure 18 shows the content of a query stored by Google. In addition to its use on Google's

---

[84] "AMP on Google privacy and terms," *Google,* last accessed on August 15 2018, available at https://developers.google.com/amp/cache/policies
[85] "Link to AMP content," *Google,* last accessed on August 15 2018, available at https://developers.google.com/amp/cache/use-amp-url
[86] "Google APIs terms of service," *Google*, last accessed on August 15 2018, available at https://developers.google.com/terms
[87] "Tracking accelerated mobile pages (AMP)," *Google,* last accessed on August 15 2018, available at https://support.google.com/analytics/answer/6343176?hl=en&ref_topic=7378717
[88] John Saroff, "The new speed of mobile engagement," *Chartbeat,* June 5, 2017, available at http://blog.chartbeat.com/2017/06/05/the-new-speed-of-mobile-engagement
[89] Tripp Mickle, "'I'm not sure I understand' – how Apple's Siri lost her mojo," *The Wall Street Journal,* June 7, 2017, available at https://www.wsj.com/articles/apples-siri-once-an-original-now-struggles-to-be-heard-above-the-crowd-1496849095
[90] "Google Assistant," *Google,* last accessed on August 15 2018, available at https://assistant.google.com/platforms/phones/
[91] "Google Assistant," *Google,* last accessed on August 15 2018, available at https://assistant.google.com/platforms/phones/

Home speakers, Google Assistant is enabled on various other speakers produced by 3rd-parties (e.g. Bose wireless headphones). Overall, Google Assistant is available on more than 400 million devices.[92] Google can collect data via all these devices since Assistant queries go through Google's servers.

Figure 18: Example of detail collected from Google Assistant query
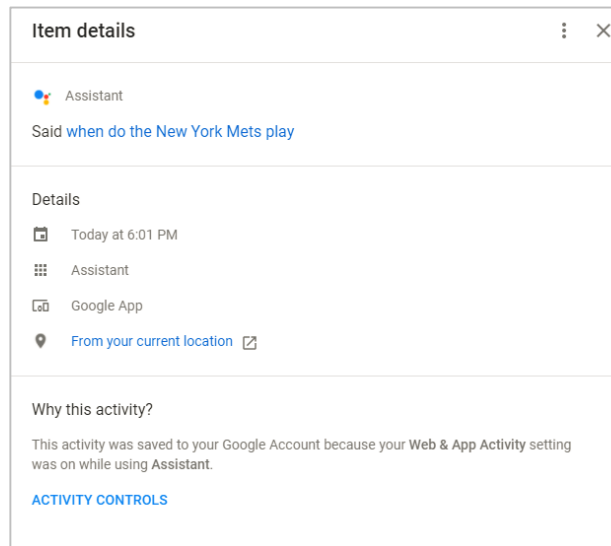
C. **Photos**

93.   Google Photos is used by more than 500 million people globally and stores more than 1.2 billion photos and videos every day.[93] Google records the time and GPS coordinates for every photo taken. Google uploads images to the Google cloud and conducts image analysis to identify a broad set of objects, such as modes of transportation, animals, logos, landmarks, text, and faces.[94] Google's face detection capabilities even enable the detection of emotional states associated with faces in photos uploaded and stored in their cloud.[95]

94.   Google Photos conducts this image analysis by default when the product is used, but will not distinguish between individual people unless the user gives the app permission.[96] If a user provides permission for Google to group similar faces together, Google identifies different people using facial recognition

---

[92] Scott Huffman, "New devices and more: what's in store for the Google Assistant this year." *Google,* Jan. 9, 2018, available at https://www.blog.google/products/assistant/new-devices-more-google-assistant-ces-2018
[93] Anil Sabharwal, "500 million people using Google Photos, and three new ways to share," *Google,* May 17, 2017, available at https://blog.google/products/photos/google-photos-500-million-new-sharing/
[94] "Cloud vision API," *Google,* last accessed on August 15 2018, available at https://cloud.google.com/vision/
[95] "Cloud vision API," *Google,* last accessed on August 15 2018, available at https://cloud.google.com/vision/
[96] "Find people, things, and places in your photos," *Google,* last accessed on August 15 2018, available at https://support.google.com/photos/answer/6128838?co=GENIE.Platform%3DAndroid&hl=en

technology and enables users to share photos based on its "face grouping" technology.[97,98] Examples of Google's image classification capabilities with and without face grouping permission from the user are shown in Figure 19. Google uses Photos to assemble a vast trove of identifying facial information, which has become the subject of recent lawsuits[99] by certain states.

Figure 19: Example image recognition in Google Photos
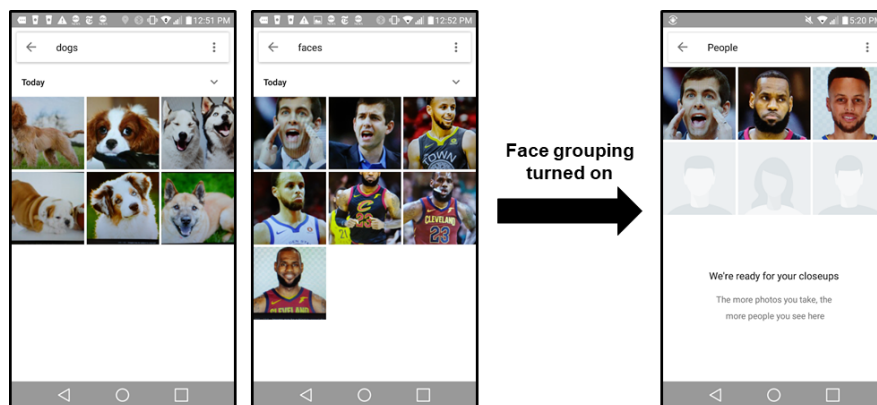
D.   Chromebook

95.   Chromebook is Google's tablet computer running on the Chrome operating system (Chrome OS) which allows users to access applications on the cloud. While Chromebook holds a very small fraction of the PC market, it's growing rapidly, especially in computing devices for the K-12 category, where it held 59.8% of the market in Q2 2017.[100] Since Chromebook is accessed through a Google Account, a user is always signed on to all of Google's applications while accessing them on a Chromebook device. Chromebook's data collection is similar to the Google Chrome browser, which is covered in section II.A. Chromebooks also allow cookies from Google and 3rd-party domains to track user activity, similar to any other notebook or PC device.

96.   Many K-12 schools use Chromebooks to access Google's products via its GSuite for Education service. Google states that data collected from such use is not used for targeted advertising.[101] Students are shown ads, however, if they use additional services (such as YouTube or Blogger) on Chromebooks provided through their educational institutions.

[97] Anil Sabharwal, "500 million people using Google Photos, and three new ways to share," *Google,* May 17, 2017, available at https://blog.google/products/photos/google-photos-500-million-new-sharing/
[98] "Share your Google Photos library with a partner," *Google*, last accessed on August 15 2018, available at https://support.google.com/photos/answer/7378858#filterbyface
[99] Amy Korte, "Federal court in Illinois rules biometric privacy lawsuit against Google can proceed," *Illinois Privacy,* March 8, 2017, available at https://www.illinoispolicy.org/federal-court-in-illinois-rules-biometric-privacy-lawsuit-against-google-can-proceed/
[100] "Mobile PC sales in to US K-12 education starting to slow as the market looking toward replacement cycles," *Future Source Consulting,* Dec. 6, 2017, available at https://www.futuresource-consulting.com/Press-Q3-2017-Mobile-PC-Sales-in-Education-1217.html
[101] "Chromebooks privacy and security," *Google,* last accessed on August 15 2018, available at https://drive.google.com/file/d/0B__OTXR_u3RbcFB3Y01xUVhaalU/view

E.   Google Pay

97.   Google Pay is a digital payments service that allows users to store credit card, bank account, and PayPal information to make payments in stores, on websites, or within apps using Google Chrome or a connected Android device.[102] Pay is the means by which Google collects verified user address and phone numbers, as these are associated with charge accounts. In addition to personal information, Pay also collects transaction information, such as date and amount of transaction, merchant location and description, type of payment used, descriptions of the items purchased, any photo that a user choose to associate with the transaction, names and email addresses of the seller and buyer, the user's description of the reason for transaction, and any offers associated with the transaction.[103] Google treats this information as personal information under its general privacy policy. It therefore can use this information across its products and services for enriched advertising.[104] Google's privacy settings allow for such a use of collected data by default.[105]

F.   User data collected from 3rd-party data vendors

98.   Google collects 3rd-party data in addition to information they collect directly from their services and applications. For example, in 2014 Google announced that it would begin tracking sales in brick-and-mortar stores by buying credit and debit card transaction data. Such data covered 70% of all credit and debit transactions in the US.[106] It contained the name of the individual, as well as the time, location, and amount of their purchase.[107]

99.   3rd-party data is also used to support Google Pay, including verification services, information arising from Google Pay transactions at merchant locations, payment methods, identity of card issuers, information regarding access to balances in the Google payment account, carrier and operator billing information, and consumer reports.[108] For sellers, Google may obtain information from credit bureaus or business information services.

100.   Although the 3rd-party user information that Google currently receives is limited in scope, it has already attracted the attention of governmental authorities. For example, the FTC announced an injunction against

---

[102] "Google Pay," *Google,* last accessed on August 15 2018, available at https://pay.google.com/about/
[103] "Google Payments privacy notice," *Google*, Dec. 14, 2017, available at https://payments.google.com/payments/apis-secure/u/0/get_legal_document?ldo=0&ldt=privacynotice&ldl=en
[104] "Google Payments privacy notice," *Google*, Dec. 14, 2017, available at https://payments.google.com/payments/apis-secure/u/0/get_legal_document?ldo=0&ldt=privacynotice&ldl=en
[105] "Google payments center," *Google,* available at
https://payments.google.com/payments/home?page=privacySettings#privacySettings:
[106] "Google plans to track credit card spending," *BBC,* May 26, 2017, available at http://www.bbc.com/news/technology-40027706
[107] Elizabeth Dwoskin and Craig Timberg, "Google now knows when its users go to the store and buy stuff," *The Washington Post,* May 23, 2017, available at https://www.washingtonpost.com/news/the-switch/wp/2017/05/23/google-now-knows-when-you-are-at-a-cash-register-and-how-much-you-are-spending/?utm_term=.281715f8e215
[108] "Google Payments privacy notice," *Google*, Dec. 14, 2017, available at https://payments.google.com/payments/apis-secure/u/0/get_legal_document?ldo=0&ldt=privacynotice&ldl=en

Google in July 2017 with respect to how Google's collection of consumer purchasing data infringes upon electronic privacy.[109] The injunction challenges Google's claim that they can protect consumer privacy throughout the process using their algorithm. Although further action has not yet occurred, the FTC injunction is an example of public concern with the amount of consumer data that Google collects.

VIII. CONCLUSION

101.   Google counts a large percentage of the world's population as its direct customers, with multiple products leading their markets globally and many surpassing 1 billion monthly active users. These products are able to collect user data through a variety of techniques that may not be easily graspable by a general user. A major part of Google's data collection occurs while a user is not directly engaged with any of its products. The magnitude of such collection is significant, especially on Android mobile devices. And while such information is typically collected without identifying a unique user, Google distinctively possesses the ability to utilize data collected from other sources to de-anonymize such a collection.

---

[109] FTC Complaint, request for investigation, injunction, and other relief submitted by The Electronic Privacy Information Center, available at https://epic.org/privacy/ftc/google/EPIC-FTC-Google-Purchase-Tracking-Complaint.pdf

IX.   APPENDIX

A.   Characterization of active vs passive data collection from "day in the life" of a user

Table 3: Active and passive Google data collection

Number	Description	Active collection	Passive collection
1	Gets ready in the morning while listening to Google Play Music	• Music interests	• Morning location
2	Drops kids off at daycare before commuting to work		• Walked to a daycare location
3	Checks the news while commuting to work on the subway		• Traveling on the subway • News pages visited
4	Searches for cold medicine while on the subway	• Records search queries	• Traveling on the subway
5	Walks from subway to work		• Commute path to work address
6	Uses Maps to find a new lunch spot	• Destination entered into Maps	• Dining interests
7	Gets coffee from Starbucks using her Starbucks app		• Walks to a Starbucks • Opens Starbucks app
8	Schedules doctor's appointment, Google creates a Calendar event from the confirmation email		• Event details of the doctor appointment
9	Walks to Walgreens and purchases cold medicine using Google Pay	• Purchase details	• Walks to a Walgreens
10	Takes an Uber home from work		• Commute path to home address via car • Use of Uber app
11	Looks for hotels on Expedia for a weekend trip		• Webpage interaction via DoubleClick cookies & Google Analytics
12	Uses Google Home to play music for her kids	• Google Home search query	• Location of Google Home

13	Watches videos on YouTube	• YouTube activity

B. List of Google products

Table 4: List of Google Products

Category	Products
1. Applications	**Watch, Listen and Play** YouTube, Google Play Music, Chromecast, Google Play Movies and TV
	Browser Chrome
	Search Search, Finance, Flights, News, Scholar, Patents, Books, Images, Videos, Hotels
	Navigation Maps, Waze
	Productivity tools Drive, Docs, Sheets, Slides, Forms
	Social & communications Gmail, Allo, Hangouts, Duo, Google+, Translate
	Storage and organization Photos, Contacts, Calendar, Keep
	Personal Assistant Google voice assistant
2. Operating systems	**Android** Phones, Wear, Auto
	Chrome Chromebook
3. Services	Fiber, DNS, Project Fi, Google pay
4. Devices	Home, Wi-Fi router, Chromecast, Nest, Daydream View

C. Data collection from other prominent Google products

   a)   Google Play Music and Play Movies and TV

102.   Google Play Music, Play Movies and TV are on-demand services that offer streaming of music, podcasts, TV shows, and movies. These services can be thought of as the Google equivalent of Apple iTunes. Like YouTube, these services collect information about user search, bought/rent/played content, information about users' geography (through IP address), and device information.

b)   <u>Waze</u>

103.   Waze got acquired by Google in 2013 and operates as a Google subsidiary. In contrast to Maps, Waze is a crowd-sourced app where user-supplied data (such as GPS location coordinates, travel times, traffic info, accidents, police monitoring, blocked roads, and construction) is analyzed to provide routing and real-time travel condition updates. In addition to location information collected through the mobile device, Waze collects information about use of its services from the device Waze is installed on, including mobile device name, operating system, web page visits, information viewed on app, app content use/created, and ads viewed and clicked.

104.   Waze works like a social network and provides users the ability to befriend other users and create an online community of local drivers.[110] Users can link their phone's contact list, Facebook, or Twitter accounts, which Waze then uses to match with friends on platforms who are also using Waze's service. Overall, Waze gives Google the ability to access more real-time user data, as well as information on acquaintances that may or may not be using a Google Account.

c)   <u>Google Docs and Drive</u>

105.   Google's productivity tools (Docs, Sheets, Slides, Forms, and Drive – which are part of the broader "G Suite" of products) are cloud-based applications used by both individuals and enterprises. Google's G Suite privacy policies[111] prohibit Google from scanning stored data for advertising purposes on enterprise versions. In contrast, the free versions of these tools are governed by Google's general privacy policy, so Google may access the information for ad targeting.

106.   After an individual registers for a Google Account, Google provides free storage space (currently 15GB) on Google Drive to share across products including Gmail, Photos, and Docs. Google's Terms of Service indicate that Google retains the license to use the stored data in variety of ways, including reproducing,

---

[110] J.D. Biersdorfer, "Getting social with Waze," *The New York Times,* Sept. 20, 2017, available at
https://www.nytimes.com/2017/09/20/technology/personaltech/getting-social-with-waze.html
[111] "Google Cloud help, privacy," *Google,* last accessed on August 15 2018, available at
https://support.google.com/googlecloud/answer/6056650?hl=en

modifying, communicating, and publishing.[112] Although all user data stored in Google Drive is encrypted, it's not a "zero-knowledge encryption"[113] since Google manages the data encryption key.

<p style="text-align:center">d)   <u>Video chat and social messaging apps</u></p>

107.   Google Hangouts is a communication platform, akin to Skype, and a part of Google's G Suite cloud-connected apps. Users can start and join video conferences or group conversations from the Hangout app available on Android and iOS, through a web browser, from the Hangouts desktop app or Chrome extension, and from other Google products (e.g. Gmail and Calendar).[114] Google stores details of these exchanges (including conference call and conversation time stamps), participant information, and message content. These details are available to users and can be downloaded via the Google Takeout tool.[115] Figure 20 shows some data recorded from a Google Hangouts conversation, including participant names, participant IDs, and the contents of the conversation.

Figure 20: Google Takeout recordings of a Google Hangouts Conversation

108.   In addition to the enterprise-focused Hangouts, Google also offers an instant messaging app called Allo, which is available for Android and iOS or through web browsers.[116] Google records and stores all messages communicated through Allo by default (unless the user invokes incognito mode).[117]

---

[112] "Google Drive terms of service," *Google,* last accessed on August 15 2018, available at https://www.google.com/drive/terms-of-service

[113] Fergus O'Sullivan, "What exactly is zero-knowledge in the cloud and how does it work?," *Cloudwards,* June 16, 2017, available at https://www.cloudwards.net/what-exactly-is-zero-knowledge-in-the-cloud-and-how-does-it-work

[114] "GSuite learning center," *Google,* last accessed on August 15 2018, available at https://gsuite.google.com/learning-center/products/hangouts/get-started/#!/

[115] "Download your data," *Google*, available at https://takeout.google.com/settings/takeout?pli=1

[116] Sean Keach, "Google Allo just got a major upgrade – but do we really need it?," *Trusted Reviews,* Aug. 16, 2017, available at http://www.trustedreviews.com/news/google-allo-3261336

[117] Russell Brandom, "Google backs off on previously announced Allo privacy feature," *The Verge,* Sept. 21, 2016, available at https://www.theverge.com/2016/9/21/12994362/allo-privacy-message-logs-google

109.   Google offers an app called Google Duo that is a dedicated mobile video chat app. It is available both on Android and iOS, but not on desktop/laptop computers. Users can call or video chat each other and even connect with users on the Android operating system who have not downloaded the app.[118] Once installed, Google Duo has access to a user's profile data, contacts, camera, microphone, Wi-Fi connection information, and device ID and call information.[119] Google also stores the time stamps of when the apps are used and provides this info to the users via Google Takeout.

e)   Google+

110.   Google+ is a social media network launched in 2011 as a competitor to Facebook. Although Google does not release statistics on Google+'s active users, it is now primarily a place for niche communities.[120] A user can choose to follow other users, organize the users they follow into groups (e.g. best friends vs. work colleagues), start "Communities," or join existing ones (e.g. "Photography enthusiasts"). Posts from followed users and communities then appear in the user's home feed.

111.   Google Takeout shows several types of Google+ data that Google stores. Google compiles vCards from the profiles of all the people a user follows (in Takeout, this information is contained within the "Google+ Circles" folder). Google also collects in HTML format all content posted by a user, which is contained in the "Google+ Stream" folder of Takeout, as shown in Figure 21. Posted photos also are saved within Google Photos.

Figure 21: Example of Google+ posted content saved by Google

John Maple - Apr 26, 2018 - Google+ - Limited - Permalink
What are the best DSLR cameras?
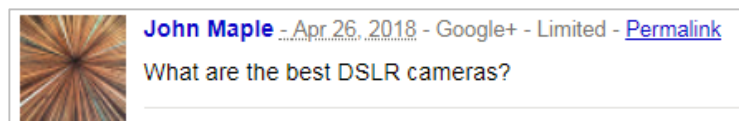
f)   Translate

112.   Google Translate is a free machine translation service supporting over 100 languages that is available on the web and through apps for Android and iOS. It is also integrated into Google Assistant and Google Chrome, as well as being available to 3rd-party developers through a paid API.[121] Altogether, it serves more than 500 million monthly users.[122] If users have the Google Translate app on their phone, they can use the app to

---

[118] Swapna Krishna, "Google Duo allows you to call people who don't have the app," *Engadget,* Jan. 12, 2018, available at https://www.engadget.com/2018/01/12/google-duo-call-android-users-without-app
[119] Google informs the user of this access when the user downloads the app.
[120] Karissa Bell, "Google+ isn't dead and these are the people astill using it the most," *Mashable,* Jan. 18, 2017, available at https://mashable.com/2017/01/18/who-is-using-google-plus-anyway/#_lP0PXr.eiqZ
[121] "Google Cloud, pricing," *Google,* last accessed on August 15 2018, available at https://cloud.google.com/translate/pricing
[122] Gideon Lewis-Kraus, "The great A.I. awakening," *The New York Times,* Dec. 14, 2016, available at https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

translate languages within other apps, such as WhatsApp.[123] Although Google states that it does not track Google Translate web queries[124]—and these queries do not appear in a user's Search history or elsewhere in Google Takeout—Google's privacy policy does allow it to store them for short period of time and occasionally for longer period for debugging and other testing.[125]

g)   Calendar

113.   Google Calendar is a scheduling tool that enables users to keep track of their daily and weekly activities. It is widely used on both desktop and mobile devices, with more than 500 million downloads from the Google Play Store.[126] Personal information, such as an individual's name and contact information, is often associated with a Calendar user.

114.   By using calendars, users provide Google with details (such as the time, location, and contact information) for all participants of an event. Google states that the app has access to "read calendar events plus confidential information."[127] As part of this information collection, Google reads and stores the email addresses of all members included in a calendar event, regardless of their Google affiliation.[128] As long as one person on a calendar invitation is using Google Calendars, therefore, Google records the contact information for every other person on the invitation.

h)   Keep

115.   Google Keep is a note-taking tool that allows the user to take notes that sync between devices associated with their Google account. Keep has been downloaded more than 100 million times from the Google Play Store.[129] Google collects and stores all contents of the notes, as well as the time they were created.[130] Figure 22 shows Google Keep notes recorded by Google. Google scans and classifies the notes that are created based on their contents. Examples of classification categories are food, places, and travel.

[123] "Google Translate," *Google,* available at https://translate.google.com/intl/en/about
[124] "Google Translate help," *Google,* last accessed on August 15 2018, available at https://support.google.com/translate/answer/6142479?co=GENIE.Platform%3DDesktop&hl=en&oco=0
[125] "Access to the Google Cloud API," *Google,* last accessed on August 15 2018, available at https://cloud.google.com/translate/faq
[126] "Google Calendar," *Google Play Store,* last accessed on August 15 2018, available at https://play.google.com/store/apps/details?id=com.google.android.calendar
[127] "Google Calendar," *Google Play Store,* last accessed on August 15 2018, available at https://play.google.com/store/apps/details?id=com.google.android.calendar
[128] "Download your data," *Google,* available at https://takeout.google.com/settings/takeout?pli=1
[129] "Google Keep," *Google Play Store,* last accessed on August 15 2018, available at https://play.google.com/store/apps/details?id=com.google.android.keep&hl
[130] "Download your data," *Google,* available at https://takeout.google.com/settings/takeout?pli=1

Figure 22: Example of Google Keep notes, accessed from Google take out

```
        </style></head>
<body><div class="note"><div class="heading"><div class="meta-icons">

</div>
Apr 26, 2018, 11:06:04 AM</div>

<div class="content">Grocery list:<br>- pasta<br>- tomatoes<br>- lettuce <br>- salad
dressing <br>- salmon <br>- lemon<br>- capers <br>- Bread<br>- butter<br>- milk<br>-
eggs<br>- bacon<br>- OJ<br>- </div>

</div></body></html>
```

i) Chromecast

116.   Like Apple TV, Google Chromecast is a device that acts as an interface to watch videos from a variety of applications (e.g. Netflix, YouTube, Hulu, Play Store), as well as to project video from smartphones and computers onto larger televisions and monitors. Every Cast device has a unique identifier that is associated with a user's Google Account during registration.

117.   Google uses Chromecast to collect system activity, crash reports, and usage data, such as details about the use of casting functionality of devices, including the apps and domains that are casted.[131] Chromecast uses Google Cast, which is a software platform that enables seamless streaming of audio/video between devices on the same network.[132] In addition to a multitude of other Google products (e.g. Google Home) that use Cast functionality, the Cast platform is also used by 3rd-party devices using "Chromecast built-in" (e.g. TVs and video gaming consoles[133]), that offer similar functionality. Google also collects usage data and crash reports from 3rd-party Cast devices.[134]

j) Google DNS

118.   Google launched a free Domain Name System (DNS)[135] service, called Google Public DNS, in December of 2009. Google DNS is aimed at improving web browsing experience by enhancing speed, security, and accuracy.[136] In December 2014, Google Public DNS was reported to be serving 400 billion responses.[137]

119.   To detect abuse (such as DDoS attacks) and to fix problems, Google Public DNS keeps a temporary log of full IP addresses that it deletes within 24-48 hours. For longer-term efforts to debug and prevent abuse,

---

[131] "Chromecast help," *Google,* last accessed on August 15 2018, available at
https://support.google.com/chromecast/answer/6076570?hl=en
[132] "Google Cast," *Google,* last accessed on August 15 2018, available at https://developers.google.com/cast
[133] "What is Google Cast and Chromecast," *Shield,* last accessed on August 15 2018, available at https://shield.nvidia.com/blog/what-is-googlecast-chromecast
[134] "Chromecast help," *Google,* last accessed on August 15 2018, available at
https://support.google.com/chromecast/answer/6076570?hl=en
[135] DNS services translate domain names into IP addresses and thus are necessary to navigate the web.
[136] "Google Public DNS," *Google,* last accessed on August 15 2018, available at https://developers.google.com/speed/public-dns/faq
[137] "Google Public DNS and location-sensitive DNS responses," *Google Webmaster Central Blog,* Dec. 15, 2014, available at https://webmasters.googleblog.com/2014/12/google-public-dns-and-location.html

Google keeps non-personally-identifiable city/metro-level location information for two weeks, and randomly samples a small subset for permanent storage.[138]

k) Google Wi-Fi router

120. Google began rolling out a mesh Wi-Fi router, Google Wi-Fi, in December 2016. Mesh routers allow a user to extend access to a Wi-Fi network across a larger area with additional connected routers. By December of 2017, Google Wi-Fi became the best mesh Wi-Fi system in the USA according to data from the NPD Group.[139] The information it collects[140] falls into the following three categories:

- Cloud services, which include broadcast information from connected devices (such as a device name like "Jane's iPhone"), infer information from connected devices (such as manufacture's name like Samsung), connection status, data transfer speed and historical consumption, network settings and information about wireless environment (e.g. other routers in the area). It does collect connected device information and data usage.
- Wi-Fi stats, which include anonymous data usage, crash reports and device performance information.
- Wi-Fi app stats, which includes usage and crash reports.

121. According to Google, the Wi-Fi router does not track the websites visited or collect the content of network traffic.[141]

l) Nest

122. In January 2014, Google acquired Nest, which is a home automation company.[142] Nest's product line includes many smart home devices, including thermostats, cameras, doorbells, alarm systems, locks and smoke/CO detectors.[143] In addition, Nest lists more than 100 3rd-party products that can integrate with Nest, ranging from refrigerators to beds.[144]

123. Nest devices collect a variety of information, including not only users' direct adjustments to the devices, but also data on the environment within the home. For example, the Nest Learning Thermostat collects data (such as temperature, humidity, ambient light, and movement) and thus knows when people are at home and even in what rooms.[145] When a user connect 3rd-party products that integrate with Nest, Nest shares

---

[138] "Google Public DNS," *Google,* available at https://developers.google.com/speed/public-dns/privacy
[139] Jillian D'Onfro, "The surprising use case that has made Google Wifi one of the company's sleeper hits," *CNBC,* Dec. 18, 2017, available at https://www.cnbc.com/2017/12/18/google-wifi-mesh-technology-sales-stats.html
[140] "Google Wifi help," *Google,* last accessed on August 15 2018, available at https://support.google.com/wifi/answer/6246642?hl=en
[141] "Google Wifi help," *Google,* last accessed on August 15 2018, available at https://support.google.com/wifi/answer/6246642?hl=en
[142] "Google to acquire Nest," *Alphabet,* Jan. 13, 2014, available at https://abc.xyz/investor/news/releases/2014/0113.html
[143] Nest homepage, available at https://nest.com/
[144] "Works with Nest," Nest, last accessed on August 15 2018, available at https://nest.com/works-with-nest/
[145] "Privacy statement for Nest products and services," *Nest,* last accessed on August 15 2018, available at https://nest.com/legal/privacy-statement-for-nest-products-and-services/

information with those parties but informs the user about what information is being shared.[146] Nest does not share data with other 3rd-parties, such as partner energy or insurance companies, without first gaining a user's consent.[147]

124.   On its website, Nest states that Nest accounts and Google accounts are not linked (unless a user chooses to integrate with Google products and services) and that Google does not sell Nest data. Recently, however, data sharing concerns have arisen from an announcement[148] by Google concerning the merger of Nest and Google hardware teams.[149] In addition, concerns have been raised in the media about the future links between Google, Nest, and 3rd-party electrical and insurance companies.[150]

<div align="center">

m)   <u>Google Fiber</u>

</div>

125.   Google Fiber is a broadband, Internet protocol (IP) TV, and Voice-Over-Internet-Protocol (VOIP) phone service connecting users via ultra-high-speed fiber-optic networks that extend all the way to their residences,[151] known as Fiber Internet, Fiber TV, and Fiber Phone, respectively. Google's efforts to deploy extensive networks of fiber optic cables were hampered by physical costs and negotiations with local municipalities. Google Fiber is therefore now expanding to new cities via Webpass (an Internet service provider acquired by Google), which delivers similarly high speeds through existing wireless and Ethernet technologies.[152]

126.   Fiber Internet collects technical information connected to a user's Google Account, but does not share account details with other Google properties without the user's additional consent.[153] Fiber Internet requires user consent before associating a user's Google Account with other information, such as sites visited or content of communications.[154]

---

[146] "Privacy statement for Nest products and services," *Nest,* last accessed on August 15 2018, available at https://nest.com/legal/privacy-statement-for-nest-products-and-services/

[147] "Privacy statement for Nest products and services," *Nest,* last accessed on August 15 2018, available at https://nest.com/legal/privacy-statement-for-nest-products-and-services/

[148] Rick Osterloh, "Nest to join forces with Google's hardware team," *Nest,* Feb. 7, 2018, available at https://blog.google/topics/hardware/nest-join-forces-googles-hardware-team/

[149] Leo Kelion, "Google-Nest merger raises privacy issues," *BBC,* Feb. 8, 2018, available at http://www.bbc.com/news/technology-42989073

[150] Casey Johnston, "What Google can really do with Nest, or really, Nest's data," *Ars Technica,* Jan. 15, 2014, available at https://arstechnica.com/information-technology/2014/01/what-google-can-really-do-with-nest-or-really-nests-data/

[151] Ryan Waniata, "Comcast killer: Google Fiber touches down in Austin with its new TV and internet devices," *Digital Trends,* Dec. 3, 2014, available at https://www.digitaltrends.com/home-theater/google-fiber-tv-hands-on

[152] Nick Statt, "Google Fiber-owned Webpass is bringing its wireless gigabit internet to Denver," *The Verge,* Feb. 22, 2017, available at https://www.theverge.com/2017/2/22/14703142/google-fiber-webpass-denver-expansion-gigabit-Internet and https://gizmodo.com/what-happened-to-google-fiber-1792440779

[153] "Google Fiber privacy notice," *Google,* last accessed on August 15 2018, available at https://fiber.google.com/legal/privacy

[154] "Google Fiber privacy notice," *Google,* last accessed on August 15 2018, available at https://fiber.google.com/legal/privacy

127.   Fiber TV collects information on programs and applications used and associates it with the user's Google Account.[155] Likewise, Fiber Phone collects usage information (e.g. logs of call history, voicemails, SMS messages, and recorded conversations) and associates it with a user's Google Account.[156] Although this information is not shared with 3rd-parties unless a user provides consent, information may be shared with 3rd-parties for external processing and for legal reasons. While Google does not explicitly mention the use of Fiber TV/phone for ad targeting, its general privacy policy does allow the use of collected personal information for targeting purposes. Moreover, Google states it will share Google Fiber user's non-personal identifiable information publicly with content providers, publishers, advertisers, and/or connected sites.[157]

D.   Method for location traffic monitoring

128.   To capture the requests being sent to Google server domains from a mobile phone, our study employs a "man-in-the-middle" (MITM) technique using the "MITM Proxy" tool. While previous studies that analyze similar data employed a Wi-Fi hotspot to act as an intermediary between mobile phone and Google servers, the present study uses a virtual private network (VPN) on a mobile phone to analyze data sent through Wi-Fi networks as well as the cellular network.

Figure 23: Example VPN setup used to analyze data shared with Google
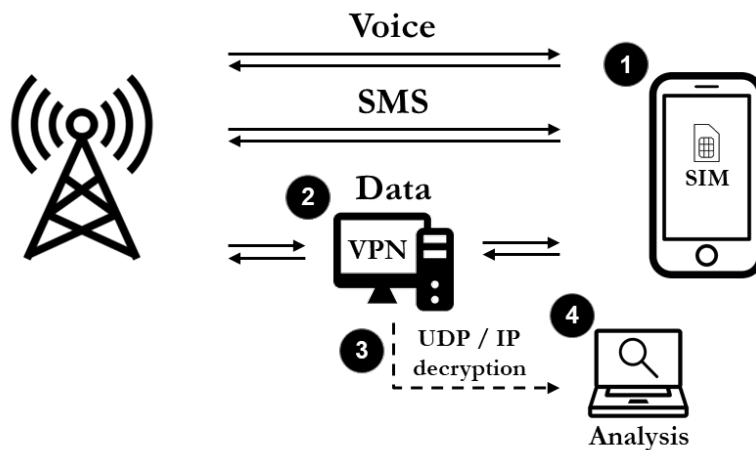
The specific steps performed to configure and conduct the experiments are described below:

1.   The mobile devices used in the data collection experiments[158] were factory reset to ensure that no previously installed applications or adjusted settings affected traffic to and from the phone. Upon

---

[155] "Google Fiber privacy notice," *Google,* last accessed on August 15 2018, available at https://fiber.google.com/legal/privacy
[156] "Google Fiber privacy notice," *Google,* last accessed on August 15 2018, available at https://fiber.google.com/legal/privacy
[157] "Google Fiber privacy notice," *Google,* last accessed on August 15 2018, available at https://fiber.google.com/legal/privacy
[158] Devices include an LG X Power with Android 6.0 version installed and an iPhone 5 with iOS 10.3.3 installed

reactivation, the devices were configured with the suggested default settings. The devices were then equipped with new SIM cards to obtain new cell phone numbers.

2. A VPN connection was setup with a remote proxy computer using VPN settings/functionality provided by the phone operating system. An IPsec/LT2P with a PSK authentication setup was chosen. The VPN configuration enabled the remote proxy to intercept and record the data transmitted from the mobile phones. Due to the nature of the type of signals emitted from a phone, the VPN set up is unable to intercept voice and SMS data sent from the mobile devices. However, it captured all TCP traffic to and from the device, including HTTP and HTTPS traffic.

3. HTTPS software certificates were installed on the mobile devices to enable decryption of the data traffic captured. VPN configuration allowed the routing of all HTTP and HTTPS traffic through the mitmproxy program[159] using iptables.[160] This program then performed SSL decryption using its own certificate to decrypt the traffic and dump it to HAR (HTTP Archive) files for analysis.

4. Analysis of decrypted HTTP and HTTPS traffic data mainly involved categorization of server requests into key segments using the request header info. Tables 5 and 6 detail the traffic headers that were identified as transmitting data to/from Google and Apple.

129.   In specific cases, requests to Google were further decoded to analyze the information that was passed at a more granular level. One specific request to Google that was further decoded was the "Google location API," designated by the /loc/m/api endpoint. The location specifications were reverse engineered by removing the message header and decoding the compressed protobuf message.[161] The decoded location API contained Wi-Fi and network scans that were used to determine the location of the device.

Table 5: Notable Google server domains communicating with mobile phone devices

Segment	Pathway header	Description
Ad domains	doubleclick.net	Sends data to and from DoubleClick
	google-analytics.com	Sends data to and from GA
	googletagmanager.com	Implements webpage tags
	googletagservices.com	Implements webpage tags
	googlesyndication.com	Retrieves and displays ads
	adservice.google.com	Calls the AdWords network

---

[159] https://mitmproxy.org/
[160] "iptables(8) – Linux man page," *Die.net,* available at https://linux.die.net/man/8/iptables
[161] Additional information on the decoding method can be found here: "Reverse engineering: Google Location protobuf specifications," *Esther Codes,* accessed March 2018, available at
https://web.archive.org/web/20180213201547/https://esther.codes/reverse-engineering-google-location-gms-specification/

|		
	google.com/ads/	Ad measurement and user lists
	gstatic.com	Loads ads on the page
	google.com/adsense	Calls AdSense
	google.com/pagead	Serves page ads
Location	google.com/loc/m/api	Sends back nearby network and Wi-Fi information
	googleapis.com/userlocation/v1/reports	Sends back user movement information (i.e. walking, running, biking, driving, etc.)
	googleapis.com/placesandroid	Matches determined location with businesses, etc. (Google Places)
	maps.googleapis.com/maps	Retrieves detailed information based on a place ID or a place search
	clients4.google.com/glm/mmap	Sends user's location information to retrieve map data
Google Play API	play.googleapis.com/log/batch	Device activity logging information
	play.googleapis.com/play/log/timestamp	Updates a cookie and reports the time
	play.googleapis.com/play/log?format	Play Store/Services log upload
Device auth. and upload	googleapis.com/batch	Device information and updates
	clients4.google.com/chrome-sync	Chrome browser synchronization
	googleapis.com/experimentsandconfigs	testing/experiment config download
	android.clients.google.com/backup/backup	Device backup
	android.clients.google.com/auth	Device authorization
	android.clients.google.com/checkin	Device identifying information and activity
	android.clients.google.com/	Authentication with Google services
Other	cdn.ampproject.org	Retrieving data from the AMP CDN
	google.com/xjs	Communication with Google Search
	google.com/gen	Communications with Google Search that transmit cookie data
	clients4.google.com/ukm	Chrome speed information
	inbox.google.com/sync	Mail synchronization
	mail.google.com/mail/ads	Mail ad refresh
	google.com/complete/search	Sends character level information to enable the search autocomplete function
	Googleapi.com/	Other Google APIs include Google services (i.e. YouTube, Calendar, etc.)

	Google.com/	Other miscellaneous calls to Google's domain

Table 6: Notable Apple server domains communicating with mobile phone devices

Segment	Pathway header	Description
Location	geosrc=wifi,73.xxx&kb_ime=en_US&key=beagle1626&latlng=40.xxx,-74.xxx&locale=en_US	Passes location coordinates back to Apple when using the browser
	cl2.apple.com	Calls to the core location server
	gs-loc.apple.com	Apple location services
App store	us-east-1.blobstore.apple.com/apple	Communicates with the Apple store; includes Apple ID
Device auth. and upload	mesu.apple.com/assets/com_apple_MobileAsset	Details mobile configurations and settings
	bookmarks.icloud.com	Syncs mobile behavior with the cloud
	ckdatabase.icloud.com/api	Communicates device authorization tokens and syncs with iCloud
	keyvalueservice.icloud.com/sync	Syncs mobile device behavior
Other	api-glb-nyc.smoot.apple.com/	Miscellaneous Apple APIs
	gsp64-ssl.ls.apple.com	Provides device information when the phone accesses websites via Safari
	gspe35-ssl.ls.apple.com/geo_manifest/dynamic/config?application=geod	Loads map tiles, but does not pass location information
	configuration.apple.com/configurations/pep/config/geo/networkDefaults	Communicates the settings of the location collection tools

E. **Google sign in authentication sequence**

Figure 24: Authentication sequence
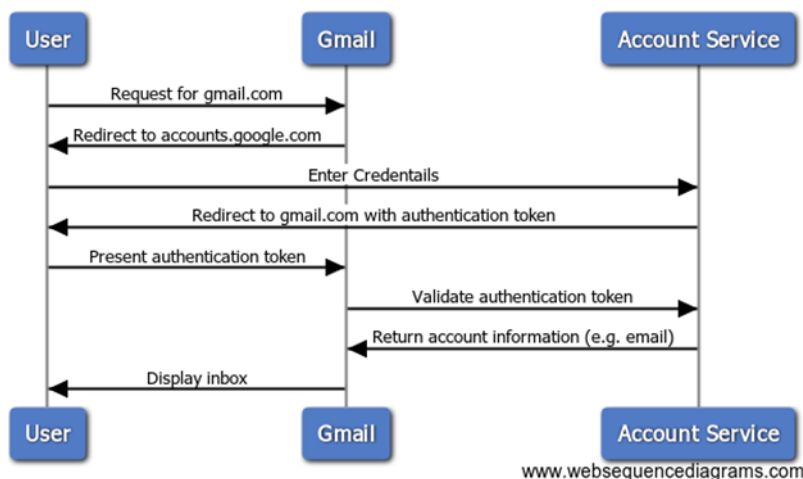
F. Usage profile for mobile data collection experiments

130. A usage profile was designed to simulate a typical user's interaction with their mobile phone throughout the course of a day. A variety of statistics that describe peoples' online behavior and mobile phone usage were integrated to create the profile, as described below.

131. The designed profile visited 45 webpages during the course of the day based on website visit statistics from a 2010 Nielsen study, which indicates that the average person visits 88 webpages per day[162] and a 2017 Stone Temple study, which states that roughly 50-55% of webpage visits come from mobile devices.[163] The 45 webpage visits were evenly split between 5 top non-Google news and sports domains.[164] When the phone was not being used to visit webpages the browser was left running in the background of the phone. The resulting usage profile represents a conservative user as the number of webpage visits per day is likely to have increased since the 2010 Nielsen study.

132. The usage profile also included a variety of non-Google mobile applications. Top non-Google applications were selected from the social media, shopping, travel, and health categories. These applications included Facebook, Instagram, Snapchat, Pinterest, Amazon Shopping, Walmart, Starbucks, Yelp, and Six Pack

---

[162] "Nielsen provides topline U.S. web data for March 2010," Nielsen, April 2010, available at
http://www.nielsen.com/us/en/insights/news/2010/nielsen-provides-topline-u-s-web-data-for-march-2010.html
[163] Eric Enge, "Mobile vs desktop usage: mobile grows but desktop still a big player in 2017," *Stone Temple,* April 2017, available at
https://www.stonetemple.com/mobile-vs-desktop-usage-mobile-grows-but-desktop-still-a-big-player-in-2017/
[164] The domains selected were New York Time, CNN, The Guardian, ESPN, and Crickbuzz. The websites were identified by using Alexa's lists, available at https://www.alexa.com/topsites/category

in 30 Days. These apps were opened periodically throughout the day to simulate a typical user who spends approximately 2.5 hours in mobile app per day, as reported by eMarketer.[165]

G. Past articles that relate to Google's data collection practices

Table 7: Summary of other Google data collection studies

Title	Relevant findings	Author, Date
AP Exclusive: Google tracks your movements, like it or not[166]	Google is tracking users' location even when location services are disabled	Ryan Nakashima August 2018
Australian regulator investigates Google data harvesting from Android phones[167]	Google "harvest" about 1GB of data from Android devices per month	Oracle May 2018
How to Keep Google From Owning Your Online Life[168]	It is very difficult for the average consumer to avoid Google products	WSJ, May 2018
Google tracking phones even when they are disconnected?[169]	Google tracks phones even when phones are "disconnected" (no SIM cards, airplane mode, Wi-Fi off)	Fox News, Feb. 2018
Google collects Android users' locations even when location services are disabled[170]	Google collects Android location when location services are turned off	Quartz Nov. 2017

---

[165] "eMarketer reveals new estimates for mobile app usage," *eMarketer,* April 2017, available at https://www.emarketer.com/Article/eMarketer-Unveils-New-Estimates-Mobile-App-Usage/1015611

[166] Ryan Nakashima, "AP Exclusive: Google tracks your movements, like it or not," *AP*, August 13, 2018, available at https://apnews.com/828aefab64d4411bac257a07c1af0ecb

[167] Anne Davis, "Australian regulator investigates Google data harvesting from Android phones," *The Guardian,* May 13, 2018, available at https://www.theguardian.com/technology/2018/may/14/australian-regulator-investigates-google-data-harvesting-from-android-phones

[168] David Pierce, "How to Keep Google From Owning Your Online Life," *The Wall Street Journal,* May 8, 2018, available at https://www.wsj.com/articles/how-to-keep-google-from-owning-your-online-life-1525795372

[169] Brett Larson, "Google tracking phones even when they are disconnected?," *Fox News,* Feb 11, 2018, available at http://video.foxnews.com/v/5731183327001/?#sp=show-clips

[170] Keith Collins, "Google collects Android users' locations even when location services are disabled," *Quartz,* November 17, 2017, available at https://qz.com/1131515/google-collects-android-users-locations-even-when-location-services-are-disabled/

Google is permanently nerfing all Home Minis because mine spied on everything I said 24/7[171]	The Google Home mini was saving recording when the device was not activated with "OK Google" *Google claims to have resolved the issue	Artem Russakovskii Oct. 2017
Online Tracking: A 1-million-site Measurement and Analysis[172]	Google can track users ~80% of websites using its cookies.	Princeton University 2016
Why Do Android Smartphones Guzzle the Most Data?[173]	Android devices consume more data (2.2GB/month) than other smartphones	Ericsson Dec. 2013
Data leakage from Android smartphones[174]	Android passes anonymous IDs along with devices IDs such as Mac address and IMIE	Lasse Øverlier June 2012

H.   Clarifications

133.   Our understanding of the data being sent to Google through its Android platform is limited to Android 6.0 version only. This study does not capture any updates/patches that may have been implemented on later versions that may affect Android's communications with the Google servers. While new versions of Android are currently present in the market, Android 6.0 is still the most widely used version.[175] Additionally, while we took utmost precaution for classifying the pathway headers by their purpose (e.g., location, ad, device upload, app store), it is possible that some headers may serve multiple purposes (e.g., ad as well as location). These aspects are not captured in our study. Consequently, the description presented for these headers may not be exhaustive with respect to the purpose they serve.

I.   About the author

134.   Professor Douglas Schmidt is a software system expert with over 30+ years conducting, supervising, and researching the development of software for distributed middleware systems and their applications in

---

[171] Artem Russakovskii, "Google is permanently nerfing all Home Minis because mine spied on everything I said 24/7," *Android Police,* October 10, 2017, available at https://www.androidpolice.com/2017/10/10/google-nerfing-home-minis-mine-spied-everything-said-247/

[172] Englehardt, Steven, and Arvind Narayana, "Online Tracking: A 1-million-site Measurement and Analysis," *ACM CCS,* 2016, available at http://randomwalker.info/publications/OpenWPM_1_million_site_tracking_measurement.pdf

[173] Brian Chen, "Why Do Android Smartphones Guzzle the Most Data?," *The New York Times,* December 31, 2013, available at https://bits.blogs.nytimes.com/2013/12/31/why-do-android-smartphones-guzzle-the-most-data/

[174] Lasse Øverlier , "Data leakage from Android smartphones," *Norwegian Defense Research Establishment,* June 6, 2012, available at https://www.ffi.no/no/Rapporter/12-00275.pdf

[175] "Mobile & Tablet Android Version Market Share Worldwide," *statcounte*r, available at http://gs.statcounter.com/android-version-market-share/mobile-tablet/worldwide

networking and security, machine learning and smart-grid, design patterns, and more. He has authored 10+ books and 600+ papers that have been collectively cited over 38,000 times. Professor Schmidt has over 30 years of experience teaching these concepts both in-classroom and online to over 200,000 students in total.

135.   Professor Schmidt has participated in 20+ prior expert engagements spanning expert report production and oral testimony both through deposition and at trial. His consulting work has ranged from patent and copyright litigation matters to advising both private and public entities on various issues relating to software infrastructure and design. He earned his PhD and MS in Computer Science from the University of California, Irvine in 1994 and 1990, respectively.

136.   Professor Schmidt is currently the Cornelius Vanderbilt Professor of Computer Science at Vanderbilt University. Prior to Vanderbilt, he held several directorial and C-level positions in both academic and industry settings including the Software Engineering Institute at Carnegie Mellon University, the Information Technology Office at the Defense Advanced Research Project Agency (DARPA), the Federal Government, as well as several technology start-ups.