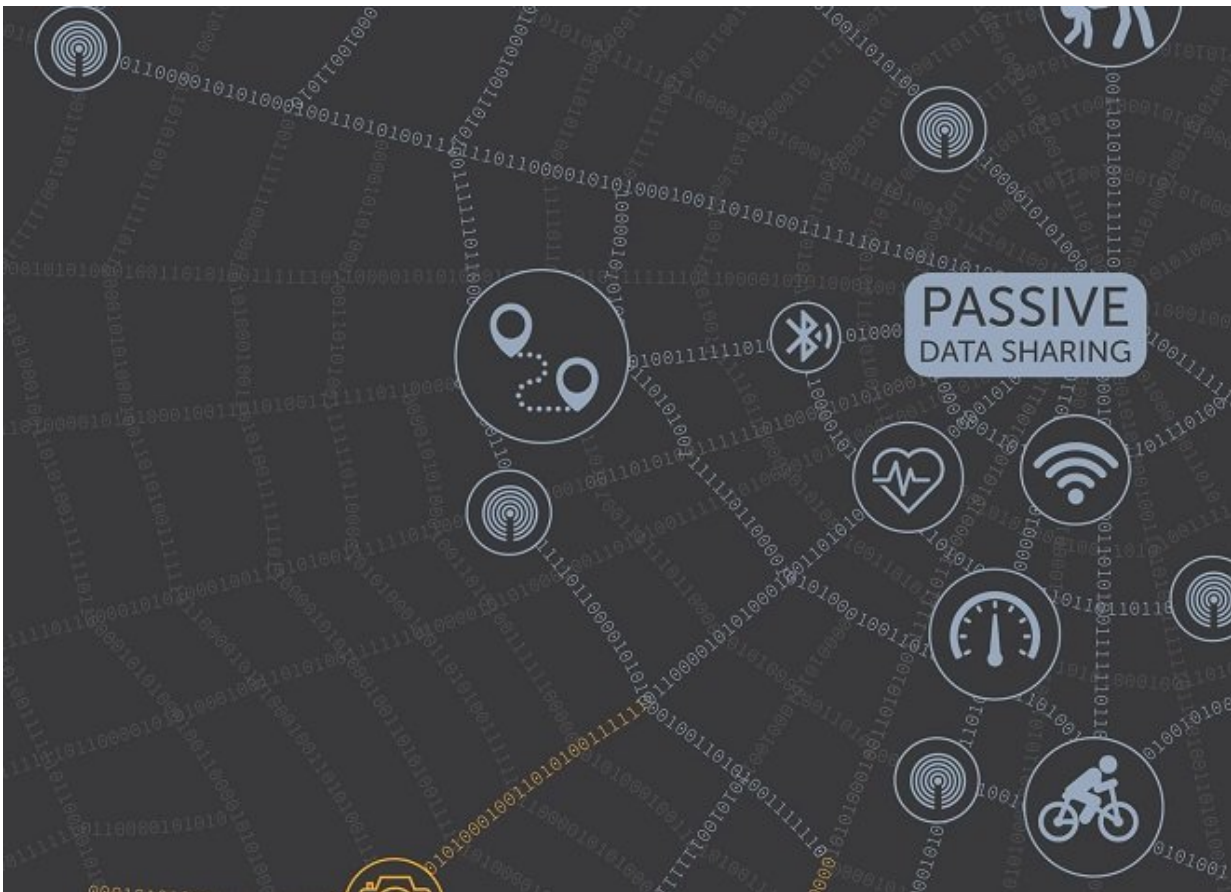


# Study of Google data collection comes amid increased scrutiny over digital privacy

November 2 2018, by Liz Entman

---



The study for Digital Content Next found much of the data Google collects from Android-enabled devices happens when the user is not interacting with Google products. Credit: Pamela Saxon

Google may not know whether you've been bad or good but it knows when you're sleeping and when you're awake. If you use an Android device with the Chrome browser running, the tech giant knows whether you are traveling by foot or car, where you shop, how often you use your Starbucks app and when you've made a doctor's appointment.

Cornelius Vanderbilt Professor of Engineering Douglas C. Schmidt studied Google's [data collection](#) practices under a "day in the life" scenario of an Android phone user. The 55-page study, commissioned by Digital Content Next, a trade group representing digital publishers, also detailed data mining over a 24-hour period from an idle Android phone with Chrome running in the background.

The stationary smartphone running Google's Android operating system and Chrome sent data to the company's servers an average of 14 times an hour, 24 hours a day.

"These products are able to collect user data through a variety of techniques that may not be easily graspable by a general user," Schmidt concluded in the paper, released in August 2018. "A major part of Google's data collection occurs while a user is not directly engaged with any of its products."

## **Mounting privacy concerns**

The study comes amid growing scrutiny of how Google collects data, including lawsuits by consumers who claim the company misled them over its practices when they used their devices in "incognito" mode and attempted to turn off their location history settings.

Also escalating is a larger debate about digital privacy and consideration in Washington D.C. of stricter privacy regulation, a step the European Union took in May 2018. Facebook, too, is under pressure for a range of

practices, including how it gathers data even when people aren't using the social media network – through third-party websites that have Facebook "like" and "share" buttons.

"The national conversation about personal data collection by various companies is intensifying, with Americans beginning to understand who's invested in knowing their online behaviors," Schmidt said. "As more information becomes available about which companies are monitoring our online behavior and for what purpose, laws and regulations will need to keep up."

After the study's release, Google questioned its credibility.

"This report is commissioned by a professional lobbyist group, and written by a witness for Oracle in their ongoing copyright litigation with Google. So, it's no surprise that it contains wildly misleading information," the company said in a statement.

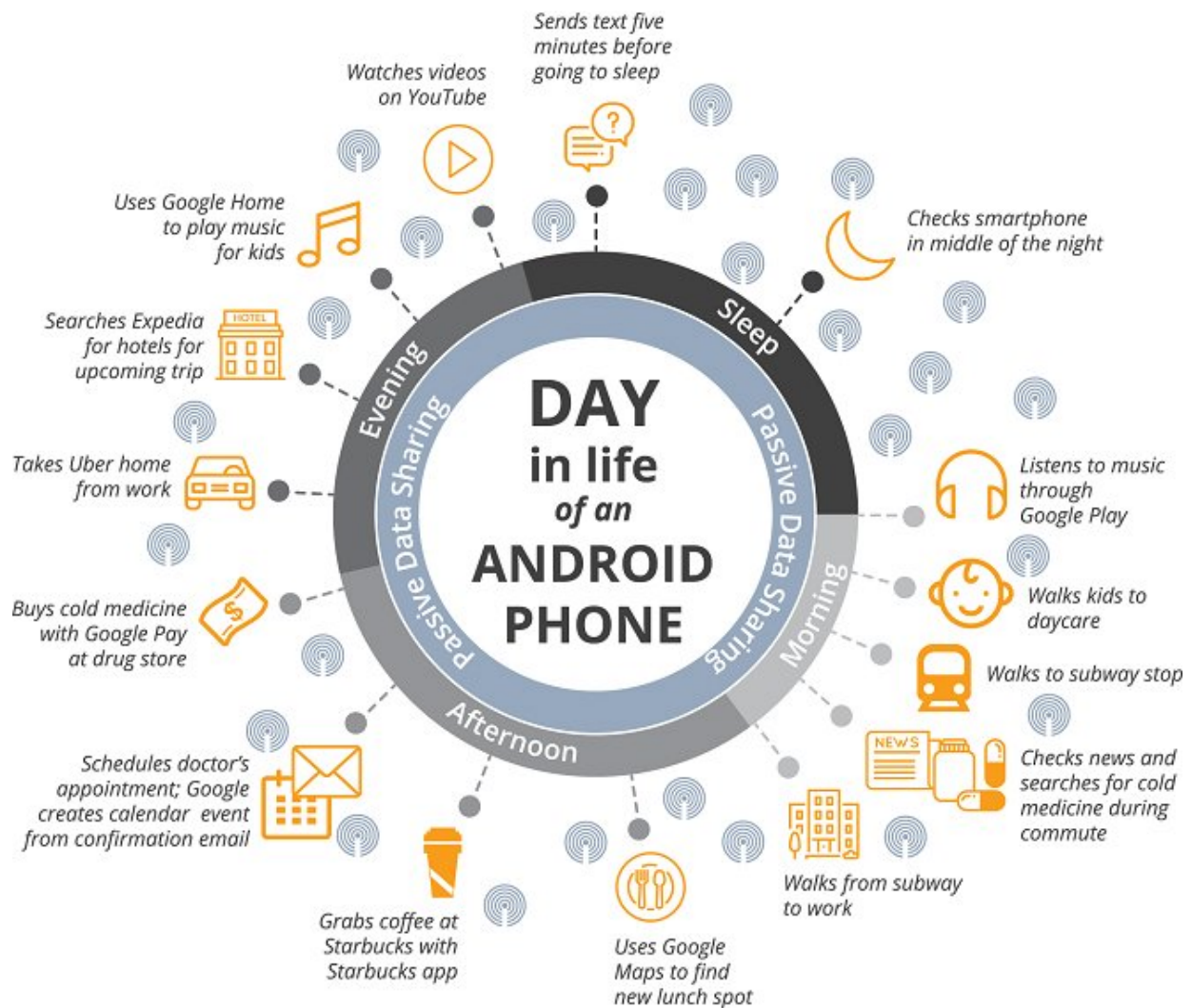
"In May of 2016 I was a witness for the Oracle vs. Google 'Fair Use Copyright,' trial (which had nothing to do with Google's data collection practices), but have not been involved with this case since then," Schmidt replied. "Moreover, Google has not been able to identify any specific aspects of my report's methods or conclusions as erroneous."

## **Phoning home – often**

Schmidt studied data gathering from all Google platforms and products, such as Android mobile devices, the Chrome browser, YouTube and Google Photos, plus the company's publishing and advertising services, such as DoubleClick and AdWords.

In the study's scenario, a researcher created a new Google account as "Jane" and carried a factory-reset Android mobile phone with a new SIM

card throughout a normal day. While riding the subway to work, she searched for cold medicine and later scheduled a doctor's appointment. From the appointment confirmation email, Google created a calendar event.



The gray “pings” represent passive data collection during a typical day of an Android phone user. Credit: Pamela Saxon

She searched for a new lunch spot, took Uber home from work, used Google Play and Google Home for music and watched videos on YouTube.

In all those instances Jane was actively engaged with Google products. The study distinguishes active data collection and "passive data collection," which occurs when the user is not using Google products directly.

Surprisingly, Schmidt wrote, "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."

What qualifies as passive data? With Chrome running and location enabled, an Android phone is "pinged" throughout the day by other wireless networks, hot spots, cell towers and Bluetooth beacons. During a short 15-minute walk around a residential neighborhood, for example, Jane's phone sent nine location requests to Google. The requests collected 100 unique identifiers from public and private Wi-Fi access points.

"Android phones can also use information from the Bluetooth beacons registered with Google's Proximity Beacon feature," Schmidt said. "These beacons not only provide user's geolocation coordinates, but could also pinpoint exact floor levels in buildings."

Even when a consumer does not use Google Maps, Google Search, Gmail or YouTube, the company's publisher and ad products collect data as she visits web pages, uses apps and clicks ads. The number of passive data collection events was twice that of active ones.

## Comparing iPhone data

The study also compared data collection from an idle Android phone running Chrome with an idle iPhone running Apple's operating system and the Safari browser. Google did not collect user location information during the 24-hour time frame. The Android phone communicated with Google twice as often as the iPhone did.

"I found that an idle Android phone running the Chrome browser sends back to Google nearly 50 times as many data requests per hour as an idle iOS phone running Safari," Schmidt said. "I also found that idle Android devices communicate with Google nearly 10 times more frequently as Apple devices communicate with Apple servers. These results highlight the fact that Android and Chrome platforms are critical vehicles for Google's passive data collection."

Schmidt found Google has the ability to identify specific users by combining "user-anonymous" advertiser data with its own collected data. The study could not determine whether the company takes such steps to link de-anonymized data when a user logs into Gmail or other Google services. In its statement, Google said it does not connect the data sources or identify users.

Not using Google's devices or services does limit data collection, but the company's dominant advertising network and tight integration of the Android platform, Chrome browser and other products makes it nearly impossible to block Google from collecting some data, the study said.

"Overall, I found that a major part of Google's data collection occurs while a user is not directly engaged with any of its products," said Schmidt. "The magnitude of Google's data collection is significant, especially on Android mobile devices, arguably the most popular personal accessory now carried 24/7 by more than 2 billion people."

The study, "Google's Data Collection," was made available to the public at Schmidt's request. Visit [Digital Content Next](#) to download the report.

## Sources used in the study

- Google's My Activity and Takeout tools, which describe information collected during use of Google's user-facing products
- Data intercepted as it is sent to Google server domains while Google or third-party products are used
- Google's privacy policies, both general and product-specific
- Other third-party research that has examined Google's data collection efforts

Provided by Vanderbilt University

Citation: Study of Google data collection comes amid increased scrutiny over digital privacy (2018, November 2) retrieved 11 July 2024 from <https://phys.org/news/2018-11-google-scrutiny-digital-privacy.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.