API Privacy: A Look at G Suite Marketplace Permissions and Policies

Irwin Reyes
Two Six Labs, LLC
Arlington, VA, USA
irwin.reyes@twosixlabs.com

Michael Lack
Two Six Labs, LLC
Arlington, VA, USA
michael.lack@twosixlabs.com

***Abstract*—Software developers routinely use application programming interfaces (APIs) to leverage existing data and functionality offered by external services. Online services such as Facebook and Google offer their own APIs for that purpose, and allow developers to access private user information like messages, files, and calendars if given proper user authorization. This has, however, produced serious privacy breaches, most notably Cambridge Analytica's unexpectedly broad collection of user data through the Facebook API. In light of this, we examined a corpus of 987 Google API apps on the G Suite Marketplace. We found that nearly half of those apps are able to communicate with outside services, whose identities aren't reliably disclosed to users. Additionally, our data suggests that app auditing measures meant to protect users from potential API misuse may fall short: a new user limit placed on potentially risky unverified apps is not rigidly enforced, and thousands of users will nonetheless authorize risky apps if allowed. We offer potential directions for improvement of this ecosystem and hope to spur further investigations of online APIs as a whole.**

***Index Terms*—privacy, APIs, disclosure, measurement**

Numerous online services expose application programming interfaces (APIs) that allow third-party software developers to take advantage of those services' existing capabilities. Such APIs may also allow external software to obtain, modify, or otherwise interact with sensitive data that users have provided to the service. For example, a business chat application may request access to a code repository's API in order to send automated alerts to team members.

Software developers routinely employ the notion of "notice and consent" when requesting access to sensitive user data. Prior investigations into this practice, however, has shown that users often consent to the requested terms merely out of habit instead of actual comprehension. Indeed, this has resulted in grave breaches of consumer privacy. This happened most notably in 2016 when the political consulting firm Cambridge Analytica collected and exploited data from 87 million Facebook users without their full understanding. Cambridge Analytica obtained this data through the Facebook web app "thisisyourdigitallife," initially developed as a personality quiz for academic purposes. Users authorized this app to access various parts of their Facebook data, such as their public profile, city, and page likes, as well as that of their friends. This sparked widespread backlash against Facebook.

In response to the subsequent outcry from lawmakers and the public at large, Facebook implemented stricter limits on the personal data third-party apps may access through the Facebook API, as well as limiting how long apps have access to consumer data without user interaction. These measures, however, came about following sustained negative press coverage and intense scrutiny from regulators, and only after initial resistance from Facebook in light of the Cambridge Analytica revelations. Although Facebook is a massive online service with a user base measuring in the billions, it is not unique in its scale, scope of data collection, or offering third-party apps access to consumer data via API.

This work is a preliminary investigation into analogous risks to consumer data posed by Google's API and the various disclosure mechanisms surrounding it. Like Facebook, Google has an active user base in the billions, competes with Facebook in the social media and advertising space, and allows third-party apps to integrate with Google functionality and user data via an API. Unlike Facebook, Google has not been subjected to the same level of criticism prompted by the Cambridge Analytica scandal. This work intends to motivate further examinations into how online services as a whole give third-party apps programmatic access to user data, as well as how consumers are informed of those privileges.

As an initial investigation,[1] we examined the third-party uses of the Google API to identify potential risks to consumer data, as well as how developers and Google themselves communicate those risks. We analyzed a corpus of 987 web apps listed on the G Suite Marketplace, and found that half are able to communicate with undisclosed external services, with a portion of those apps also holding permission to access users' Google Drive files, emails, or contacts. Additionally, while Google requires developers to submit apps for review if they use "sensitive" API functions, those products may still be listed on the Marketplace as "unverified." We found that the restriction on unverified apps gaining new users is not rigidly enforced. Unverified apps will continue to draw many new users—on the order of thousands in our 16-day observation period—despite warnings to do otherwise. We believe that even after a major scandal stemming from the abuse of an API provided by a competitor, our results show that there is still substantial risk in these systems, and recognize broad opportunities for improvement in how online services such as Google expose user data for programmatic use by third-parties.

[1]This work was funded by Two Six Labs, LLC