I never signed up for this! Privacy implications of email tracking

Steven Englehardt and Arvind Narayanan
FTC PrivacyCon 2018 Talk Proposal

Authors

Steven Englehardt

Arvind Narayanan

Publication

I never signed up for this! Privacy implications of email tracking
Steven Englehardt, Jeffrey Han, and Arvind Narayanan
Appearing at Privacy Enhancing Technologies 2018 (PETS 2018)

Paper available at: https://senglehardt.com/papers/pets18_email_tracking.pdf
Additional analysis of the companies involved is provided in a supplemental blog post: https://freedom-to-tinker.com/2017/09/28/i-never-signed-up-for-this-privacy-implications-of-email-tracking/

Talk Abstract

The proposed talk will be based on the paper *I never signed up for this! Privacy implications of email tracking*. The abstract of the paper is below. In the talk, our discussion will focus on the use of hashed email addresses as tracking identifiers. In particular, we will discuss how email address hashes are collected when a user opens or interacts with an email, and will explore the privacy risks associated with tracking identifiers based on email address.

We found a number of trackers collecting and sharing the recipient's email address when an email is viewed, both in plaintext and hashed. In some cases, hashed email addresses were shared directly with trackers that have a large web presence. In others, the addresses were collected by trackers that offer personalization and identity services. We will explore these cases to show that the privacy impact of email tracking extends far beyond a sender learning that an email has been read.

Email-address-based tracking identifiers are persistent, cross-device, and unique. Users will very rarely change their email address, and will share it with companies when they sign up for online accounts or make in-person purchases at a store. Trackers that collect this information can use it to connect a user's activities across devices, to reconnect tracking profiles after a user has taken steps to clear their on-device identifiers, and to "onboard" data from offline sources. By analyzing the privacy policies and marketing materials of the email trackers, we show evidence that email-address-based are used in this way.

Lastly, we will show that hashing does little to protect the privacy of a user's email address. When user records in a database are keyed by hashed email address, looking up the record for a given email address is trivial: simply hash it first and look it up (indeed, this is the whole point of storing hashed email addresses at all). Data associated with a hash of an unknown email address is also likely to be recoverable -- during the talk we will provide estimates of the cost and effort necessary to break common email addresses.

Paper Abstract

We show that the simple act of viewing emails contains privacy pitfalls for the unwary. We assembled a corpus of commercial mailing-list emails, and find a network of hundreds of third parties that track email recipients via methods such as embedded pixels. About 30% of emails leak the recipient's email address to one or more of these third parties when they are viewed. In the majority of cases, these leaks are intentional on the part of email senders, and further leaks occur if the recipient clicks links in emails. Mail servers and clients may employ a variety of defenses, but we analyze 16 servers and clients and find that they are far from comprehensive. We propose, prototype, and evaluate a new defense, namely stripping tracking tags from emails based on enhanced versions of existing web tracking protection lists.