Transparency and Choice: Protecting Consumer Privacy in an Online World

Alma Whitten[a] Sean Harvey[b] Ian Fette[c] Betsy Masiello[c] Jochen Eisinger[d] Jane Horvath[e]
{alma,sharvey,ifette,betsym,eisinger,janehorvath}@google.com

[a] Google Inc., Belgrave House, 76 Buckingham Palace Road, London SW1W 9TQ, UK
[b] Google Inc., 76 Ninth Avenue, New York, NY 10011, USA
[c] Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
[d] Google Germany GmbH, Dienerstr. 12, D-80331 München, Germany
[e] Google Inc., 1101 New York Avenue, N.W., Second Floor, Washington, DC 20005, USA

Abstract

There have been concerns raised recently about online tracking. There are a variety of mechanisms by which data is collected online, and for which it is used, and it is unclear which of these are intended to be addressed by "Do Not Track" mechanisms. Tracking is often data collection that helps ensure the security and integrity of data, determines relevancy of served content and also helps create innovation opportunities. This value ought to be central in any "Do Not Track" discussions.

1. Introduction

The idea of a "Do Not Track" mechanism has inspired debate among those concerned about online tracking. Several mechanisms and solutions exist or have been proposed to provide users with choice and control over the profiling they experience online. Each of these has limitations and consequences—none appears to be a panacea for concerns about tracking. Perhaps most significantly, there is a wide range of definitions of "tracking" and thus no uniform problem to solve for. If implemented carelessly, solutions to prevent tracking could have the unintended consequence of diminishing the online experience for users and stifling the growth of online publishing without meaningfully improving user privacy.

2. Tracking

There are many types of data collection that occur when a user browses the Web, and these occur for many different reasons. In discussions about "Do Not Track" it is important to be concrete about what is meant by "track."

Mechanisms of data collection that create streams of information about a particular user or browser include HTML cookies, javascript, authentication, or advanced types of "fingerprinting." Data is

collected by first-party publishers to serve personalized content to users. It can also be collected by third-party content providers for the same reason. Data may also be used to monetize content, either by serving contextually targeted advertisements to a user, or by inferring interests that a particular browser or user is likely to have and serving ads targeted to those interests. Importantly, the data collected may be used not only to personalize content and advertisements, but also to protect and secure services from fraud and abuse.

3. Existing approaches

There are a variety of existing approaches to preventing tracking. The longest-standing approach relies on cookie settings in the browser: users of most major browsers can choose to block all cookies from being set, block third party cookies from being set, or in some cases block cookies from specific domains from being set. More recently a number of approaches have sought to build on the cookie infrastructure but enable users a more global option for preventing ad targeting.

A second variety of approach is browser extensions and features that block network requests altogether or block the display of network content. This network-level blocking is effectively based on a list of domains to which network requests cannot be sent if an extension or feature is turned on, or from which content cannot be viewed.

A third and newly proposed approach is an HTTP header. The idea is for users to signal their preference to not be tracked universally to all websites they visit.

3. Analysis

What is sometimes referred to as tracking is often data collection that helps ensure the security and integrity of data, determines relevancy of served content and also helps create innovation opportunities. It is important not to let a single negatively-loaded term obscure the fact that data collection is the source for the creation of value as well as the legitimate concerns of different parties.

A common assertion made in discussions about tracking is that average users do not want to be tracked and do not understand the tracking ecosystem. As we set out goals for the workshop, one question to ask is whether one goal should be helping users understand the data collection that occurs online as well as its risks and benefits.

One observation of the range of solutions described above is that most, though not all, focus on providing the user a simple decision interface for an inherently complex ecosystem: turn on a header or don't, block third party cookies or don't, install a content blocking feature or don't. However, some solutions have features that focus on enhancing transparency to the end user. We should analyze the impact of improved transparency on the user's online experience, as well as their understanding of the ecosystem, and whether improved transparency influences the decisions they make about various

forms of tracking.

In addition to focusing on the user experience, there needs to be a focus on user value. An important observation is that the tracking used to monetize online services may be invisible to the user, and yet provide the user immense value. Advertising may be less annoying or intrusive if it is useful and relevant to the user. As a simple example, data collection enables advertisers to do frequency capping, which ensures that the same ad is not shown repeatedly to a given browser or user. Tracking allows advertising companies to monitor for fraudulent services or deceptive ads, further ensuring that irrelevant messages or offers are not intruding a user's experience. To ask a user to make a decision about tracking without incorporating both sides of the equation—the value they get from advertising-supported content as well as their concerns about tracking—would put at risk aspects of the online experience that millions of users have come to expect and value.

4. Conclusions

To address the concerns of users and give them effective tools to improve their online privacy, it is important to (a) be transparent about what data is collected and how it is used, and (b) offer users meaningful choices that are understandable to both users and sites being notified of these choices, without mysterious or unintended side effects. A commonly agreed-upon definition of tracking would be an important step forward. Tracking ought to consider the connection of information about online behavior to the offline world, as well as the rise of cloud-based computing and the ever growing mobile market. Most importantly, solutions should put the user first: what should the user understand about tracking if they are to make an informed decision, what expectations do they have about their online experience that they may unknowingly compromise without that understanding, and what is the value that users derive from tracking?