The current state of tracking on the web

Web tracking is used to collect information about, and even identify, users. It is now ubiquitous across the web; according to a study by WhoTracks.me in April 2018, 71% of the 1330 most visited websites contained tracking, with an average of 8 trackers per site, and an average number of 17 tracking requests per page load (WhoTracks.Me: Shedding light on the opaque world of online tracking: https://arxiv.org/abs/1804.08959).

There has been increasing awareness amongst web users of privacy concerns. 62% of Americans say it's not possible to go through daily life without companies collecting data about them, 81% said they have little to no control over the data collected by companies, and 79% were very or somewhat concerned about how their data is collected (June 2019, PEW Research Centre https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/). The same study showed a trend over time, with 70% of respondents feeling that their personal information was less secure than five years ago.

The basics of trackers

Web tracking can be divided into two categories. First party tracking is when the site a user is visiting tracks them directly, often to provide a better user experience, such as by storing a user's selected language in a cookie, so it can be remembered if the user revisits the site.

Third party tracking is when someone other than the site a user is currently visiting tracks them. It can allow companies to build up detailed profiles of users, containing large amounts of personal information, and can sometimes even be linked to their real identity, all typically without the users consent or knowledge. This data can then be used for a variety of purposes, such as price gouging, it may be compounded and sold, and is often used for targeted advertising. The nonconsensual and privacy violating nature of this data collection and use makes it particularly worth fighting against, both as an individual concerned about their privacy, and as a company.

There are several ways users are tracked across the web:

Cookies

Cookies are small pieces of data that are stored by browsers whilst using a website. First party cookies typically serve as local storage for a website to provide a better user experience (e.g. to keep you logged in).

Third party cookies are mostly used to track users and can be used to show ads to people after they've left the site, such as showing users ads for a product they've already viewed (called retargeting).

Tracking domains

Usually implemented as pixels, an invisible single pixel image embedded in a webpage, that once the page is loaded, transmits data about the user's browser to a tracking domain.

Fingerprinting

By using a combination of attributes of your device and browser (e.g. which browser and what version you use, what device you are on, the window size, your location, language settings, etc.), it's possible to combine this information to uniquely identify users (i.e like a "fingerprint" that is unique to each person), and hence track you across different websites.

Tools such as Cover Your Tracks (formally Panopticlick) can tell you how unique your browser's fingerprint is.

The basics of tracker blocking

It's possible to block most third-party tracking. Some browsers stop tracking through cookies by blocking some third-party cookies, either by using a block list of known tracking domains, or through an analysis of their behavior to identify domains that use cookies for tracking (such as Safari's ITP). Some browsers have even started blocking all third-party cookies.

Blocking connections to known tracking domains can be done either on the client (e.g. in the browser), or at any other layer of the network, such as through DNS filtering, or through a VPN. Identification of which domains to block can be done either with some automatic classifier (e.g. a machine learning algorithm), or by using a large list of known trackers, such as notrack, or the DuckDuckGo tracker radar.

Fingerprinting can be blocked by using a fingerprint resistant browser. These work by either making your fingerprint less unique, or less consistent. For example, browsers can block JavaScript on some sites (and therefore stop sites from reading a lot of data that is used for fingerprinting), block sites from accessing certain APIs (such as the battery API or blocking access to canvas data without permission), or can periodically change the user agent.

The privacy landscape

There are a number of companies with products targeting users concerned about tracking. Several organizations have their own tracker blocking browsers: DuckDuckGo currently offers a mobile browser, Brave also have their own tracker blocking browser for both desktop and mobile, and Mozilla's Firefox browser also includes tracker blocking.

There are also many companies that have browser extensions to block trackers, such as Ghostery, Privacy Badger, and DuckDuckGo. DuckDuckGo also maintain the DuckDuckGo Tracker Radar, an automatically generated and updated data set of trackers on the web.

Some non-privacy focused browsers also have some level of tracker blocking built in now, Safari now blocks all third party cookies, with Google planning to do likewise in the future.