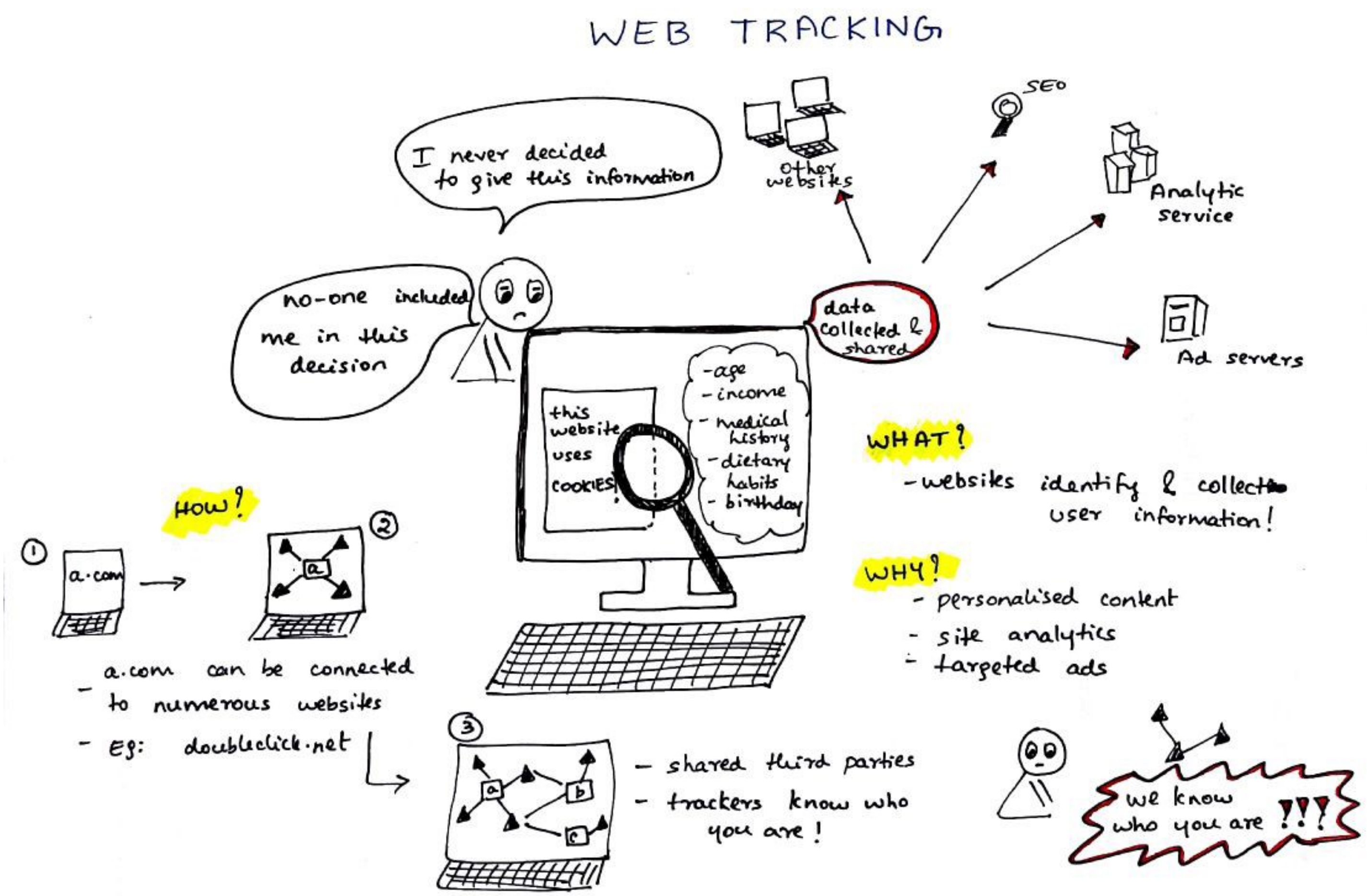
Web Tracking: What You Should Know About Your Privacy Online

Princiya Follow
Apr 23, 2018 · 6 min read



An overview of web tracking

The wake of the Facebook scandal has all of us concerned and debating over the rise of online privacy. Here is my attempt to explain how web tracking works in a nutshell.

What you will find in this article:

What is web tracking?

Web tracking is the practice by which websites identify and collect information about users. This is generally in the form of some subset of web browsing history.

How does it work?

Whenever you use the Internet, you leave a record of the websites you visit, along with each and every thing you click. To track this information, many websites save a small piece of data, embed invisible objects, or use your user accounts and hardware configuration.

More on this is described below in the "tracking mechanisms" section.

Why is it done?

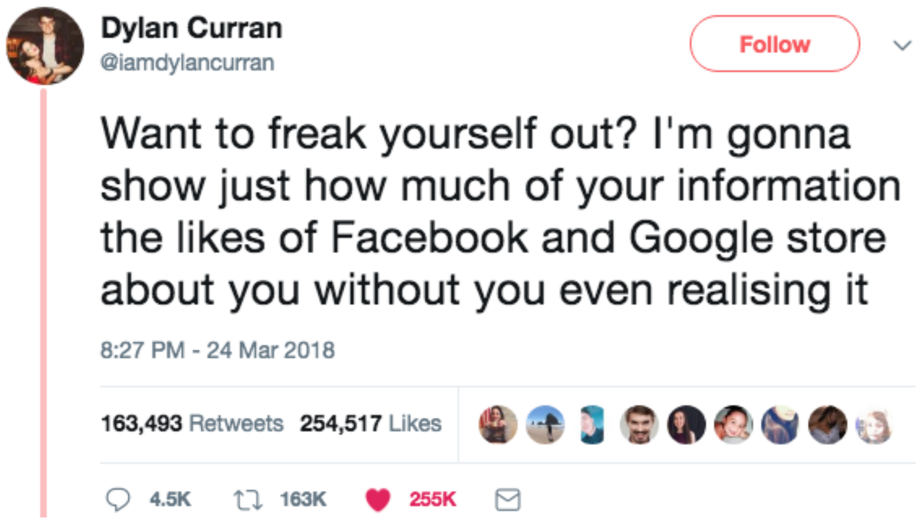
From the perspective of website owners and of trackers, it provides desirable functionality, including personalization, site analytics, and targeted advertising.

Without trackers, an e-commerce website will have to treat every user as a stranger and would be unable to present personalized content.

Is web tracking evil?

Web tracking isn't 100% evil, but its workings remain poorly understood. After you switch websites, advertisements for products you've just looked at, or products you looked at a few weeks ago, reappear! The greatest concern involves trackers from third-party websites.

This Twitter thread describes how much of our information is being collected by Google and Facebook.



Dylan Curran
@iamdylancurran

Want to freak yourself out? I'm gonna show just how much of your information the likes of Facebook and Google store about you without you even realising it

8:27 PM - 24 Mar 2018

163,493 Retweets 254,517 Likes

4.5K 163K 255K

First-party vs third-party web tracking

Say for example, you go to nytimes.com. The New York Times knows you've visited and knows which article you're reading. In this case, the New York Times is a "first-party."

Because you choose to visit a first-party, we are not particularly concerned about what the first-party knows from your visit. A third-party tracker like doubleclick.net—embedded by nytimes.com to provide, for example, targeted advertising—can log the user's visit to nytimes.com.

The number of trackers that exist in any website depends on what the website owner has decided.

What is third-party tracking?

Third-party web tracking refers to the practice by which an entity (the tracker), other than the website directly visited by the user, tracks or assists in tracking the user's visit to the site.

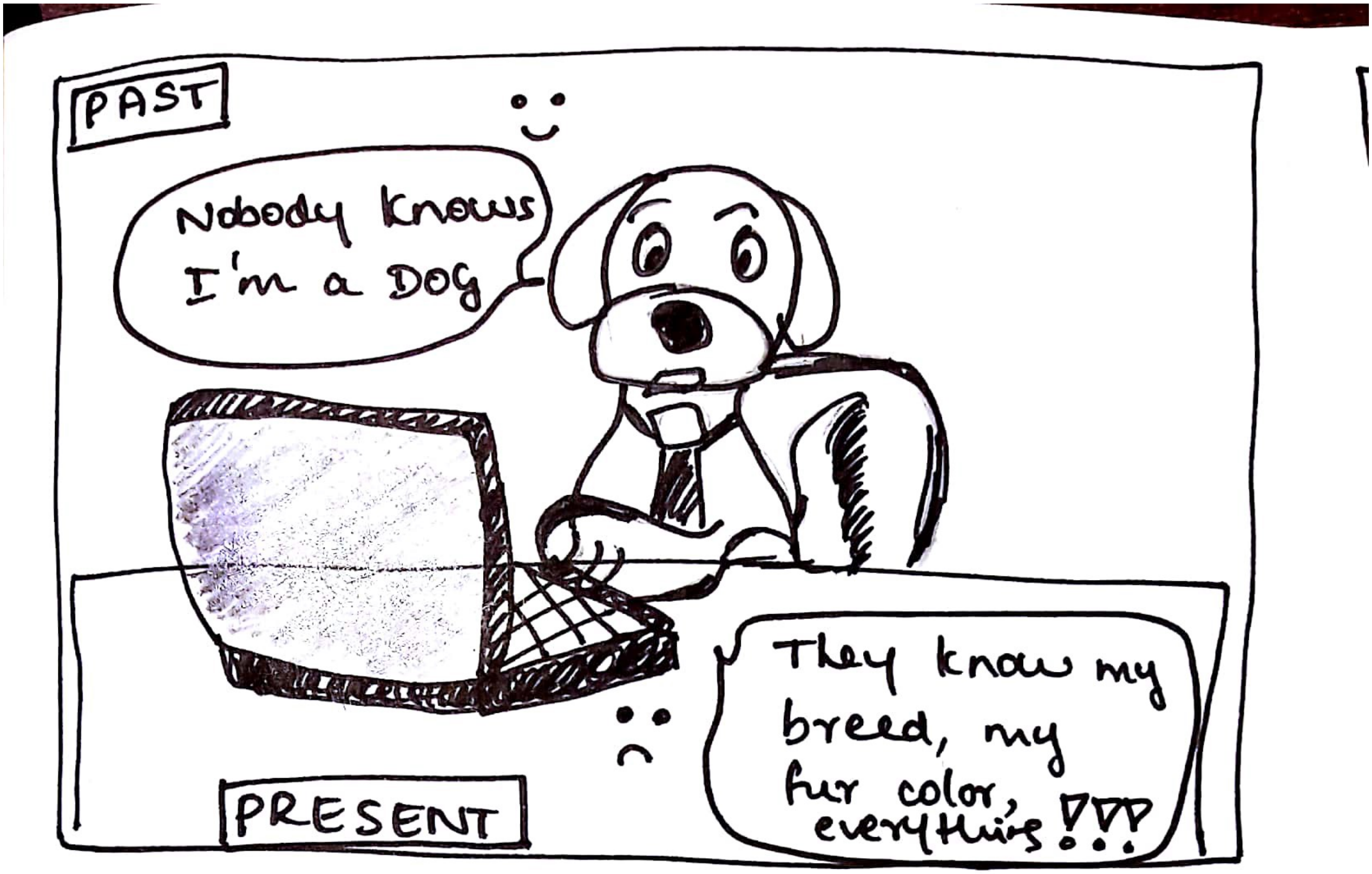
Third-party trackers are creepy.

Once there is one third-party on a page, that third-party has the ability to turn around and invite any number of other third-parties to the first-party webpage.

Your personal information is valuable, and **it's your right to know what data is being collected about you**—your age, income, family's ages and income, medical history, dietary habits, favorite web sites, your birthday…the list goes on.

The trick is in taking this data and shacking up with third-parties to help them come up with new ways to convince you to spend money, sign up for services, and give up more information. It would be fine if you decided to give up this information for a tangible benefit, but you may never see a benefit aside from an ad, and no one's including you in the decision.

Tracking is not anonymous



Cartoon depicting tracking isn't anonymous

You might think that this tracking is anonymous, since your real name is not attached to it. But many third-parties do know your real identity.

For example, when Facebook acts as a third-party tracker, they can know your identity as long as you've created a Facebook account and are logged in—and perhaps even if you aren't logged in.

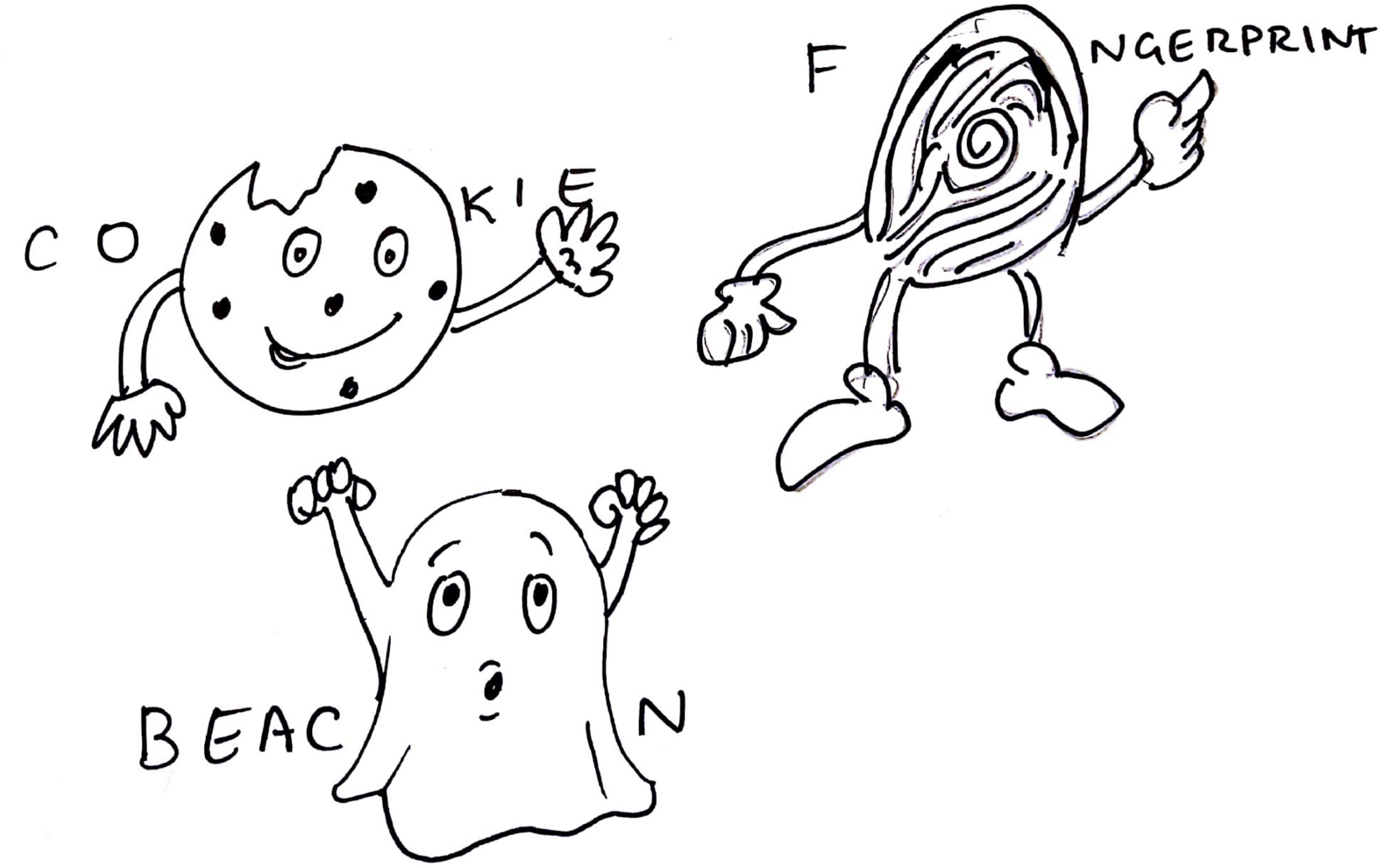
It is also possible for a tracker to de-anonymize a user by algorithmically exploiting the statistical similarity between their browsing history and their social media profile.

Visible third-party trackers

While most third parties are invisible, visible page elements such as Facebook Like buttons, embedded Twitter feeds, and a variety of other commercial widgets are all modes of third-party tracking.

Tracking mechanisms

Below are the most common tracking mechanisms:



Cookies, fingerprinting and beacons — tracking mechanisms

Cookies are the most widely known method to identify a user. They use small pieces of data (each limited to 4 KB) placed in a browser storage by the web server. When a user visits a website for the first time, a cookie file with a unique user identifier (could be randomly generated) is stored on the user's computer.

Subsequent visits to the Facebook page do not require you to login, because your details will be remembered by the browser through a cookie stored during your first login.

Browser fingerprinting is a highly accurate way to identify and track users whenever they go online. The information collected is quite comprehensive, and often includes the browser type and version, operating system and version, screen resolution, supported fonts, plugins, time zone, language and font preferences, and even hardware configurations.

These identifiers may seem generic and not at all personally identifying. But, typically only one in several million people have exactly the same specifications as you.

Web beacons are very small, usually invisible objects embedded into a web page or email. Web beacons are also referred to as "web bugs," which also go by the names "tags," "page tags," "tracking bugs," "pixel trackers," or "pixel gifs."

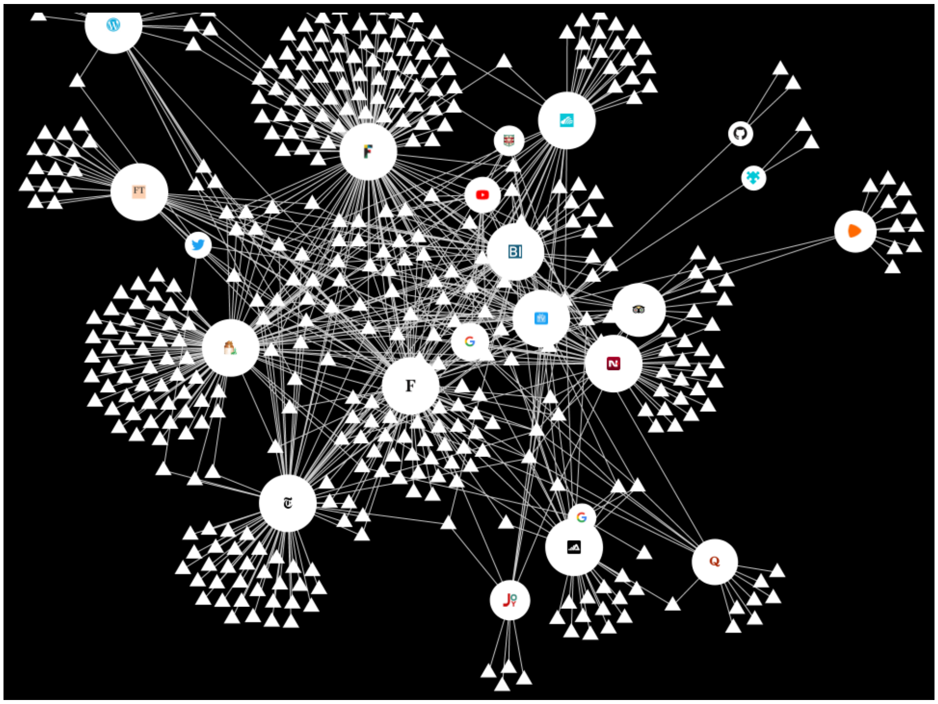
In their simplest form, they are tiny clear images, often the size of a single pixel. They download as an image when the web page is loaded, or the email is opened, making a call to a remote server for the image. The server call alerts the company that their email has just been opened or their web page visited.

This is why you should not display images in emails from senders you do not trust.

Web beacons are also used by online advertisers who embed them into their ads so they can independently track how often their ads are being displayed.

How to track the trackers!

Last year I interned at Mozilla through the Outreachy program. I worked on a tool called Lightbeam. This privacy browser extension helps you discover who's tracking you online while you browse the web.



A visualization from Lightbeam showing first and third party trackers

When you activate Lightbeam and visit a website, the browser extension creates a real time visualization of all the third-party trackers that are active on that page. As you then browse to a second site, it highlights the third-party trackers that are also active there, and shows which third parties have seen you on both sites. The visualization grows with every site you visit, and every request made from your browser.

References and Further Reading

There is much more to web tracking than what is written here. Here are some good places to learn more:

1. The WebTAP project

2. Third-party tracking on the web

3. Mozilla Privacy Basics

Stay safe and make the internet a healthier place!