Session-2: Tracking techniques: (HTTP cookies, third party cookies)

Unlike the telephone system, which has an effective tracking and billing capability based on the need to charge users of its services on a per-call basis, the Internet has no standard provisions for tracking or tracing the behavior of its users. The original Internet model never envisioned billing on the basis of single host-to-host connection or some fine-grained unit of delivery of service (such as a small charge for each email message sent or received). Originally, access to the Internet was sponsored by ARPA and was free to its end users. Current Internet charges are based on large-grained service provisions, such as monthly connectivity, connection speed, and storage capacity, all of which do not require Internet service providers (ISPs) to track the fine-grained behavior of their customers in order to bill them.

One of the consequences of Internet protocols being designed under the assumption of a benign and trustworthy user community was the lack of any provision for cryptographic authentication of the information contained in IP packets. Therefore, an advanced user can readily modify any information in an IP packet and, in particular, can forge the source address of a packet, effectively hiding its true origin. For a one-way communication, an attacker needs only to insert a false address in the source field, a relatively straightforward task. For a two-way communication, forging the source address is more complex, but the techniques are well known to the attacker community. Another way to obscure the true origin of an attack on a desired target host is to first compromise a number of intermediate hosts and to then use them as stepping stones on the way to the final target. It's not uncommon for a large number of stepping stones to be used, and from the target's perspective, the attack will appear to have originated from the last stepping stone (i.e., the compromised host that transmitted the attack packets directly to the target).14 This "packet laundering" technique is quite effective at thwarting traceback attempts, particularly if there is a significant time delay between any attacker activities15 involving the stepping stones and the attack on the final target. Moreover, an attacker may arrange things16 so that the packets transmitted from stepping stone A to stepping stone B may be quite different in nature from the packets transmitted from stepping stone B to stepping stone C on the way to the final target T, making traceback even more difficult because attempts to trace the attack by correlating similar packets will fail.

If you use a web browser like Chrome, Firefox, Internet Explorer, Edge, or Safari, then you've probably picked up a few cookies along the way. Cookies are used to remember things about websites: your login information, what you have in your shopping cart, what language you prefer. They are created by websites and sit in your browser until they expire.

Some cookies are harmless, but others remain active even on websites that they didn't originate from, gathering information about your behavior and what you click on. These are called third-party persistent cookies or, more colloquially, tracking cookies.

Tracking cookies can be so invasive that many antivirus programs classify them as spyware. Despite their bad reputation, they have become so ubiquitous that it's nearly impossible to avoid

them. In this article, we'll go into detail and explain how tracking cookies record your web activity, why they're so popular, and how to stop them.

Types of cookies explained

First, let's briefly cover the main types of cookies: session cookies and persistent cookies. Whenever you go into your browser settings and clear your cookies, you're deleting the persistent cookies.

Session cookies

The most basic type of cookie is a session cookie. Session cookies only exist in temporary memory and are deleted when you close the browser. Any cookie created without an expiration date is automatically a session cookie. A common use for session cookies includes remembering what's in your shopping cart on an ecommerce site (although most modern ecommerce sites now store this info in a database on their servers).

First-party persistent cookies

Persistent cookies are written onto your device's memory and come with an expiration date. They are only used by the website that created them, and can last however long the website dictates. They remain on your device even after you close your web browser. Your web browser uses first-party persistent cookies for many quality-of-life enhancements, like remembering that you're signed in so you don't need to log in every time you visit the same site.

Third-party persistent cookies

Third-party persistent cookies, also known as tracking cookies, are the main focus of this article. Like their first-party brethren, these cookies are stored in your device's memory and have a set expiration date. Unlike the first-party variety, however, third-party persistent cookies are accessed on websites that didn't create them. This allows the cookie's creator to collect and receive data any time the user visits a page with a resource belonging to them.

Where do tracking cookies come from?

Websites today are rarely made up solely of code and content created by the website owner or administrator. Instead, they use resources from other sites to build and add functionality to their web pages. These resources are often useful and even essential for a website to compete. Unfortunately, those same resources are often the biggest perpetrators of online tracking. Some of the most common resources that use tracking cookies include:

- Advertisements
- Social media widgets (Like and Share buttons, comments sections, etc)
- Web analytics

You don't even need to click on an ad or social media sharing button for a tracking cookie's information about you to be transmitted back to a server owned by the person or company who created it. As soon as you load the page, the cookie is sent to the server where it originated. If no cookie exists yet, the resource can create one.

Let's say I write a blog post an include an image that's hosted on another website. The other website can create a cookie or send and existing one to its server, even though I'm not actually on that website; I'm just loading a resource from it. Similarly, most ads and widgets aren't hosted by the websites they reside on. They are just resources pulled from third-parties, and they all use cookies.

According to The Guardian, some of the biggest companies using tracking cookies include:

- AddThis
- Adnxs
- Doubleclick
- Facebook
- Google
- Quantserve
- Scorecard Research
- Twitter
- Yieldmanager

What do track cookies know about me?

Tracking cookies are usually used for advertising purposes, retargeting in particular. Retargeting is a tactic that often relies on tracking cookies to show ads to people who have previously visited a specific site or shown interest in a particular product. If you've ever bought or even looked at a product on Amazon and then started seeing ads for similar products on other websites, you've been retargeted.

Here's a simplified step-by-step explanation of how retargeting works:

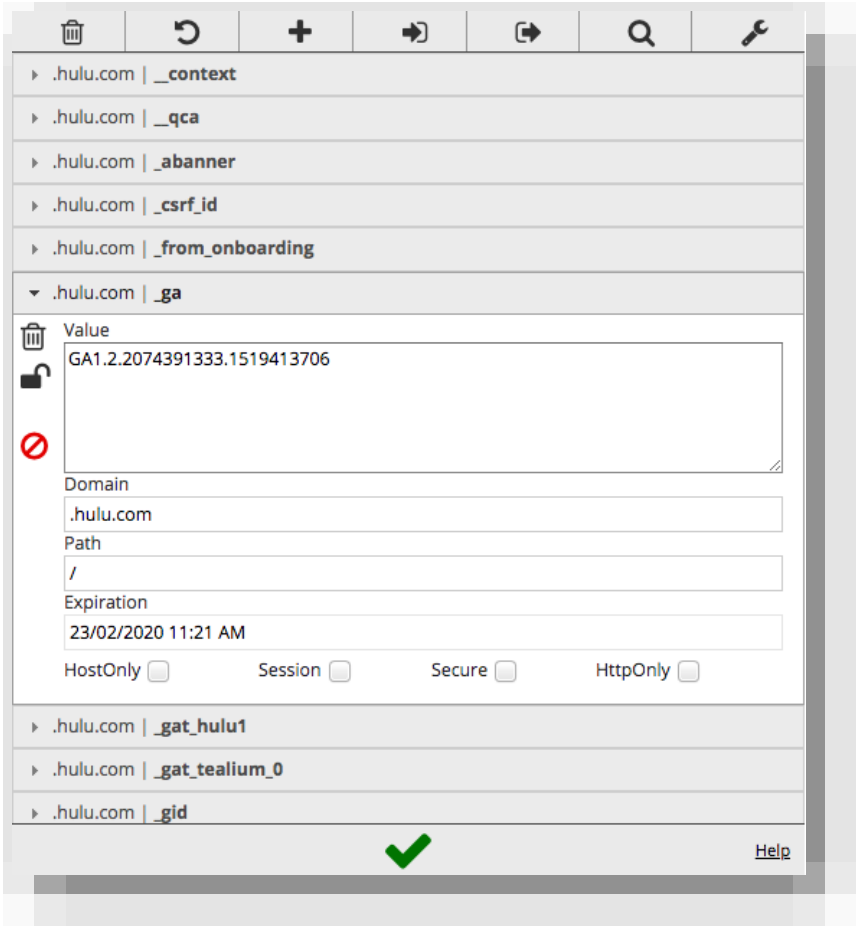
1. You pick up a tracking cookie on your favorite blog or shopping site. That cookie contains a unique ID that doesn't identify you personally, but does identify your web browser.
2. The owner of the shopping site signs up and pays for an advertising platform like Google.
3. Google's ads aren't static; when you visit other websites that use Google ads to make money, the website sees the cookie and sends it to Google through the ad. Google sees the unique ID stored in the cookie and recognizes that it came from your favorite shopping site.
4. Google then shows an ad for the shopping site accordingly.

Likewise, other advertisers on Google's ad network can use that cookie, too, if your advertising profile meets their criteria of the target audience. It doesn't only benefit the site where you

picked up the cookie. This might seem harmless at first, but those tracking cookies can start racking up a lot of information about how you browse the web. Google's ads are everywhere, and while it's the largest online advertising company in the world, there are many, many others. Because of this, advertising companies can cobble together a history of what websites you visit, in what sequence, and for how long. When cookies are sent back to their servers, they often include information about the previous site that a user visited, called a referrer URL. Browsing history is just the start. Tracking cookies can record all kinds of information: search queries, purchases, device information, location, when and where you saw previous advertisements, how many times you've seen an ad, and what links you click on. All of this and more is collected, often without your consent or knowledge. In the UK and EU, websites are required to notify users if they use tracking cookies. In the US and other countries, however, all of this data is hoarded in the background.

What makes a cookie?

If you examine the actual contents of a cookie file, none of this is obvious. Cookies only consist of three components: name, value, and attributes. Using the Chrome extension EditThisCookie, we can see what makes up a cookie:



Name is used by websites and advertisers to identify cookies and what they're used for.

The value component is where your unique advertising ID stored so that the tracker's creator can identify you when you visit other websites. It usually appears as a seemingly random string of numbers and digits, but in some cases it's not random and can contain coded information as laid out above.

Attributes include characteristics of the cookie like:

- When the cookie expires. If no expiration date is set, the cookie ends when the browser is closed. Tracking cookies always have expiration dates.
- If the cookie can be used by other domains.
- Whether the cookie can be sent over an insecure connection or not. Essentially, it checks for HTTPS.
- Whether the cookie can be accessed through JavaScript. Disabling this prevents cross-site scripting (XSS) attacks that can be used to steal login credentials and modify cookies for nefarious purposes.

How to stop tracking cookies

The first step toward preventing tracking cookies from monitoring your behavior is to delete the ones you already have. You can clear your cookies in your browser settings. If you're not sure how, check out this guide on clearing cookies for all the major browsers and operating systems.

Your browser doesn't distinguish between persistent cookies that perform useful tasks like keeping you logged into a website from those that invade your privacy and track you around the web. When you clear cookies in your browser, all of them are deleted.

Do not track

Somewhere in your browser settings you'll find an option to toggle on Do Not Track. Enabling this feature will send a request for the website you're currently on to disable its cross-site user tracking of individual users. This includes tracking cookies.

While some sites honor your choice to opt out with Do Not Track, many will not. Do Not Track does not add any technical limitations and there's no enforcement from any authority. That means there's no consequences for websites that ignore you opt-out request and use tracking cookies anyway.

You should definitely enable Do Not Track in your browser, but you'll need to go a step further if you want to put a halt to tracking cookies.

Keeping track of where all the cookies in your browser come from and whether they track you would be a very tedious task. Instead, you can install an anti-tracking browser extension to stymie tracking cookies for you.

Privacy Badger and Disconnect are two good options. An ad blocker like Adblock Plus can help, too. All of these not only make the web more private, they also speed up page load times by blocking third-party elements.

Created, by the Electronic Frontier Foundation, Privacy Badger is a plugin for Firefox and Chrome that automatically blocks advertisers that use tracking cookies from loading any more content in your browser. It does this by keeping track of third-party domains that embed images, scripts, and advertising into pages that you visit. The extension doesn't use a blacklist of known tracking sites. Instead, it observes the behavior of third-party domains on web pages and blocks them if they collect unique identifiers.

Privacy Badger also protects against canvas fingerprinting and super cookies, which we'll explain a bit later.