Agenda

- Review: Http Cookies
- Web Tracking
In computing, the same-origin policy is an important concept in the web application security model. Under the policy, a web browser permits scripts contained in a first web page to access data in a second web page, but only if both web pages have the same origin. The same origin means two pages have the same (choose all that apply.)

- A. protocol
- B. host
- C. port
- D. ip address
An http cookie may have several attributes, such as domain, path, httponly, hostonly, secure. What does the "secure" cookie attribute mean?
The Icon
The AdChoices Icon

- AdChoices is a self-regulatory program for online interest-based advertising that exists in the United States, Canada and across Europe.
- The program calls for advertising companies to establish and enforce responsible privacy practices for interest-based advertising, aimed to give consumers enhanced transparency and control.
- "Interest-based advertising" selectively displays ads based on browsing history, primarily using cookies, to users most likely to identify with and respond to the ad’s specific content. The AdChoices icon is shown automatically by companies part of the self-regulatory program, and is meant to indicate to consumers when interest-based advertising data is being collected or used. By clicking on the icon, a consumer can learn more about the ad or a website’s collection practices, and be provided with the ability to opt-out of such targeting.
Terms

- **First-party cookie**: cookies that are associated with the host domain (the domain name that appears in the browser address bar);
- **Third-party cookie**: cookies from any other domain;
- For example, suppose that you visit www.example1.com by typing the URL in the address bar, and sample.example1.com, www.example2.com, and www.example1.net have banner ads on this page. If these sites all set cookies, the cookies from www.example1.com and sample.example1.com are in a first-party context, and the cookies from www.example2.com and www.example1.net are in a third-party context.
Web Tracking

▶ a.k.a. third-party web tracking, or tracking.
▶ as you browse the web, your browsing behavior may be observed and aggregated by third-party websites ("trackers") that you don’t visit directly.
▶ web-tracking, the practice by which websites identify, and collect information about user - generally in the form of some subset of web browsing history.
▶ has become a topic of increased public debate.
Privacy Problems of Web Tracking

- A first party sells the user’s identity
- A first party unintentionally provides identity - Viewing a local ad on the Home Depot website sent the user’s first name and email address to 13 companies.
- A hacker breaks into a tracking company and publishes its tracking information, causing some embarrassing fact about the consumer become known and inflicting emotional distress (harm).
Observe trackers: Using Burp Suite

- or httptrace (for chrome), or httpfox (for firefox)
- visit the following websites and observe the major difference, in terms of how many websites you would actually access.
  - http://cs.boisestate.edu/~jxiao/cs331
  - http://www.cnn.com
  - http://www.boisestate.edu
  - http://www.maroon5.com
Observe trackers: Lightbeam

- an add-on for Firefox that displays third party tracking cookies placed on the user’s computer while visiting various websites.
- it displays a graph of the interactions and connections of sites visited and the tracking sites to which they provide information.
Top Trackers

- google-analytics.com (present on nearly a third of top sites)
- doubleclick.net, acquired by Google in 2008, appeared on over 15% of all the popular sites.
Classification of Web Trackers

- **Analytics:** the tracker serves as a third-party analytics engine for site. It can only track users within sites, e.g., Google Analytics

- **Vanilla:** the tracker uses third-party storage that it can get and set only from a third-party position to track users across sites, e.g., Doubleclick

- **Forced:** the tracker forces user to visit its domain directly via popup or redirect, e.g., InsightExpress

- **Referred:** the tracker relies on another tracker to leak unique identifiers to it, rather than on its own cookies, e.g., Invite Media

- **Personal:** the tracker is visited by the user directly in other contexts, e.g., Facebook, Twitter. (personal trackers commonly appear as social widgets, e.g., "Like" or "tweet" buttons.)
Case Study: Doubleclick
What is Doubleclick?

Doubleclick is a business owned by Google that makes money from online advertisers and publishers. This is done by:

- **Ad-serving:** online publishers use Doubleclick to display adverts on their websites.
- **Ad delivery:** Doubleclick will let advertisers control how often an ad is shown to a browser, how long it is shown for and how often it will appear.
- **Behavioural targeting:**
  - Targeting for one website owner: An online publisher can set a Doubleclick cookie to tell them what sections of their sites you are browsing. Doubleclick will then judge the type of adverts you might like to see from what you’re browsing. For example, if you are on a news website and you visit the sports pages, then adverts for match tickets may be more relevant than makeup. This information belongs to the website owner only.
  - Targeting in advertising networks: Google runs a service called Adsense, in which lots of different publishers pool the information they get on browsers. This helps them build up a better idea of the type of adverts someone might want to see.
Who uses Doubleclick?

Any web publishers can use Doubleclick, as long as they meet the company’s terms and conditions. Google also uses Doubleclick across its own sites.
What information is Doubleclick tracking?

In their privacy policy, Google explains how data is recorded from a generic Doubleclick cookie. It looks like this:

time: 06/Aug/2008 12:01:32
ad_placement_id: 105
ad_id: 1003
userid: 00000000000000001
client_ip: 123.45.67.89
referral_url: "http://youtube.com/categories"

This tells Doubleclick the time and date you saw an advert. It also shows: userid: the unique ID number the cookie has given your browser ad_id: the unique ID of the advert ad_placement_id: the ID of where the advert was seen on the site referral_url: what page you were on when you saw the advert

Because it records your IP address, Doubleclick can also make a good guess of your country and town/city, too.
How is this information used?

- On its own, this data can tell Doubleclick how many times you have seen an ad and, for example, whether you need to see the UK or US version.
- If the cookie is set on an website that is part of AdSense and then you browse another site using AdSense, the same information will be recorded and pooled. Over time, guesses can be made about the interests of the person using that browser – whether they like football or ballet, for example, or if they visit Marks & Spencers more than Tesco.
- If enough browsers display similar patterns, they become part of a "segment" labelled something like "football lovers", "fine food lovers", "current affairs enthusiasts", etc – these become labels that Doubleclick lets advertisers choose from when they select the types of people they want to see their ads.
- To check what segments Doubleclick has you in, go to Google’s ad preferences manager: https://www.google.com/settings/ads/onweb
How tracking genuinely benefits the people being tracked?

- Relevancy (Before cookies the web was much worse at distributing adverts that had absolutely no relevance or interest to the people browsing. It still isn’t an exact science but they say it’s hard to argue with the evidence that ads that use targeting are more likely to be clicked on.)

- Frequency capping (Cookies can control how often a browser sees the same advert, so browsers don’t get pushed the same advert continuously.)
A large portion of the material is adapted from:


- DoubleClick (Google): What is it and what does it do?: https://www.theguardian.com/technology/2012/apr/23/doubleclick-tracking-trackers-cookies-web-monitoring

- Lightbeam: https://en.wikipedia.org/wiki/Lightbeam_(software)
Backup Slides
Other cookies used by Doubleclick/Google

Types of cookies used by Google

https://www.google.com/policies/technologies/types/
Other techniques to track you

A website may be able to learn properties about the browser that, taken together, form a unique or nearly unique identifier. Some properties require active discovery through a script or plug-in, other properties can be passively learned from network traffic.

▶ Active fingerprinting: operating system, CPU type, user agent, time zone, display settings, installed fonts, installed plugins, enabled plugins.

▶ Passive fingerprinting: IP address, operating system, user agent, language, HTTP accept headers.
To protect yourself or your privacy

- Block third-party cookies
- Send a "Do Not Track" request with your browsing traffic, commonly supported by mainstream browsers.
How to Add Google Analytics to Your Website

http://www.wikihow.com/
Add-Google-Analytics-to-Your-Website