

Lecture 9: Online Tracking

CS 181E

Fall 2026

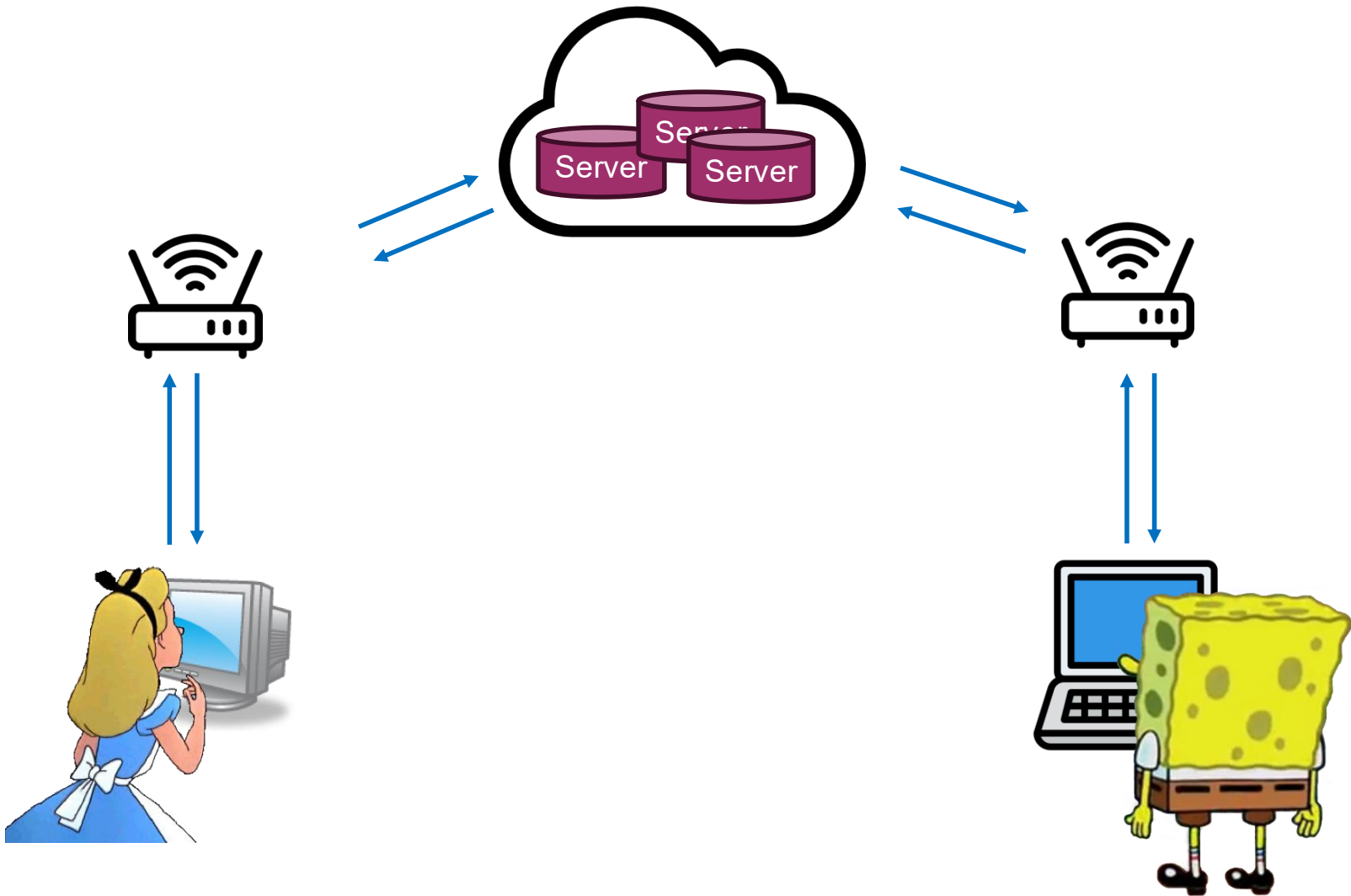
Online Tracking



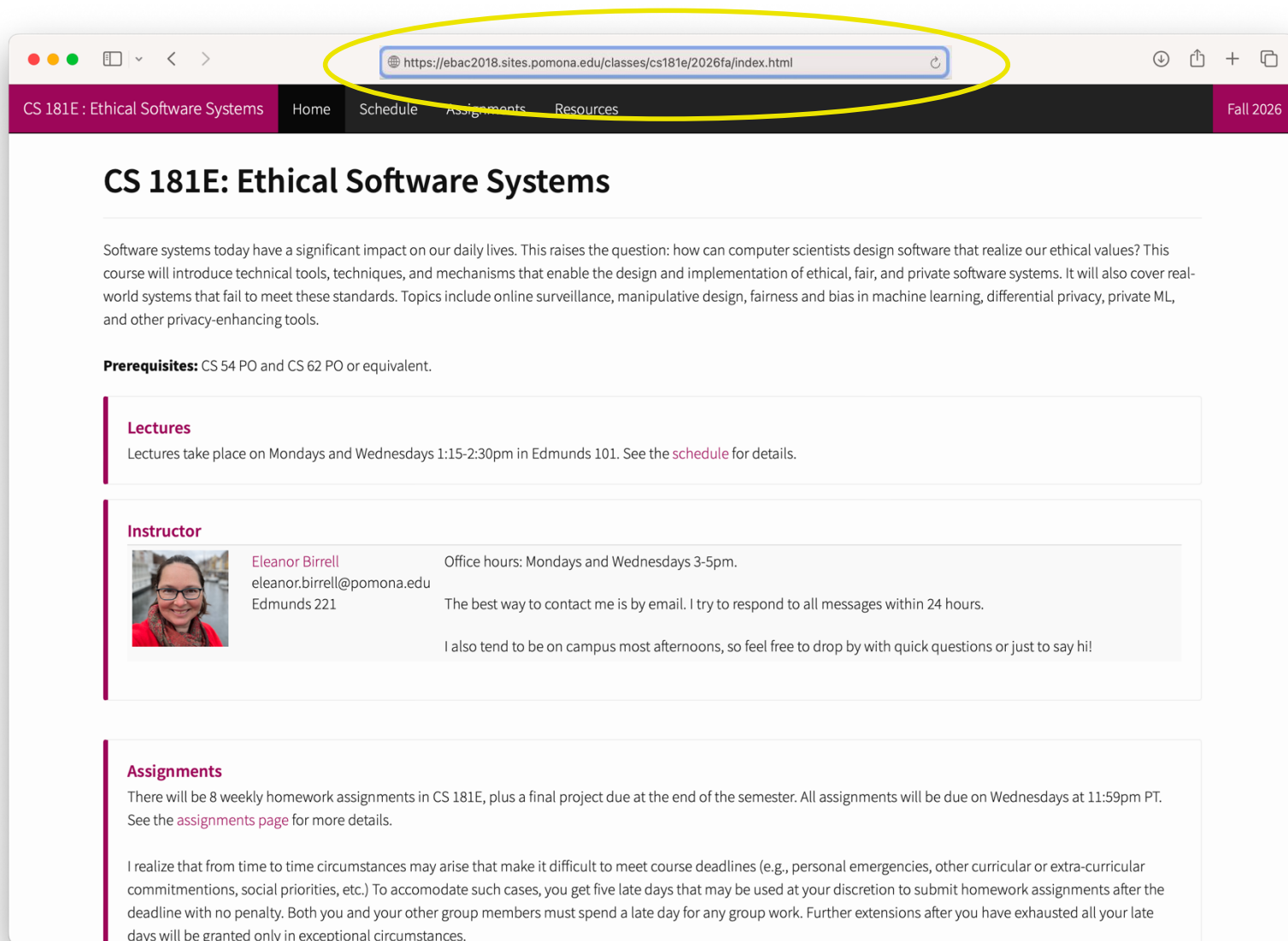
Online Tracking



Review: Networked Systems



The web



URLs

- Unique name for a file: URL (Universal Resource Locator)
- Example URL:

<http://www.cs.pomona.edu/classes/cs181e/index.html>

prefix

suffix

- Clients use *prefix* (`http://www.cs.pomona.edu`) to infer:
 - What kind (protocol) of server to contact (HTTP)
 - Where the server is (`www.cs.pomona.edu`)
- Servers use *suffix* (`/classes/cs181e/index.html`) to:
 - Specify what content they want

HTTP Requests

- HTTP request is a **request line**, followed by zero or more **request headers**
- **Request line:** `<method> <uri> <version>`
 - `<method>` is one of GET, POST, OPTIONS, HEAD, PUT, DELETE, or TRACE
 - `<uri>` is typically URL for proxies, URL suffix for servers
 - A URL is a type of URI (Uniform Resource Identifier)
 - See <http://www.ietf.org/rfc/rfc2396.txt>
 - `<version>` is HTTP version of request (HTTP/1.1 or HTTP/2.0)
- **Request headers:** `<header name>: <header data>`
 - Provide additional information to the server

HTTP Example

- HTTP Request: Request Method Path Protocol Version

1 GET / HTTP/1.1

2 Host: developer.mozilla.org

3 Accept-Language: fr

Headers (optional)

HTTP Responses

- HTTP response is a **response line** followed by zero or more **response headers**, possibly followed by **content**
 - a blank line (“\r\n”) separates headers from content.
- **Response line:** `<version> <status code> <status msg>`
 - `<version>` is HTTP version of the response
 - `<status code>` is numeric status
 - `<status msg>` is corresponding English text
 - 200 OK Request was handled without error
 - 301 Moved Provide alternate URL
 - 404 Not found Server couldn't find the file
- **Response headers:** `<header name>: <header data>`
 - Provide additional information about response
 - `Content-Type:` MIME type of content in response body
 - `Content-Length:` Length of content in response body

HTTP Example

- HTTP Request:

```
1 GET / HTTP/1.1
2 Host: developer.mozilla.org
3 Accept-Language: fr
```

- HTTP Response:

Protocol Version

Status Code

Status Msg

```
1 HTTP/1.1 200 OK
2 Date: Sat, 09 Oct 2010 14:28:02 GMT
3 Server: Apache
4 Last-Modified: Tue, 01 Dec 2009 20:18:22 GMT
5 ETag: "51142bc1-7449-479b075b2891b"
6 Accept-Ranges: bytes
7 Content-Length: 29769
8 Content-Type: text/html
9
10 <!DOCTYPE html... (here comes the 29769 bytes of the request)
```

Headers (optional)

Body

Web Content

- Web servers return content to clients
 - *content*: a sequence of bytes with an associated MIME (Multipurpose Internet Mail Extensions) type
- Example MIME types
 - `text/html` HTML document
 - `text/plain` Unformatted text
 - `image/gif` Binary image encoded in GIF format
 - `image/png` Binary image encoded in PNG format
 - `image/jpeg` Binary image encoded in JPEG format

You can find the complete list of MIME types at:

<http://www.iana.org/assignments/media-types/media-types.xhtml>

HTML

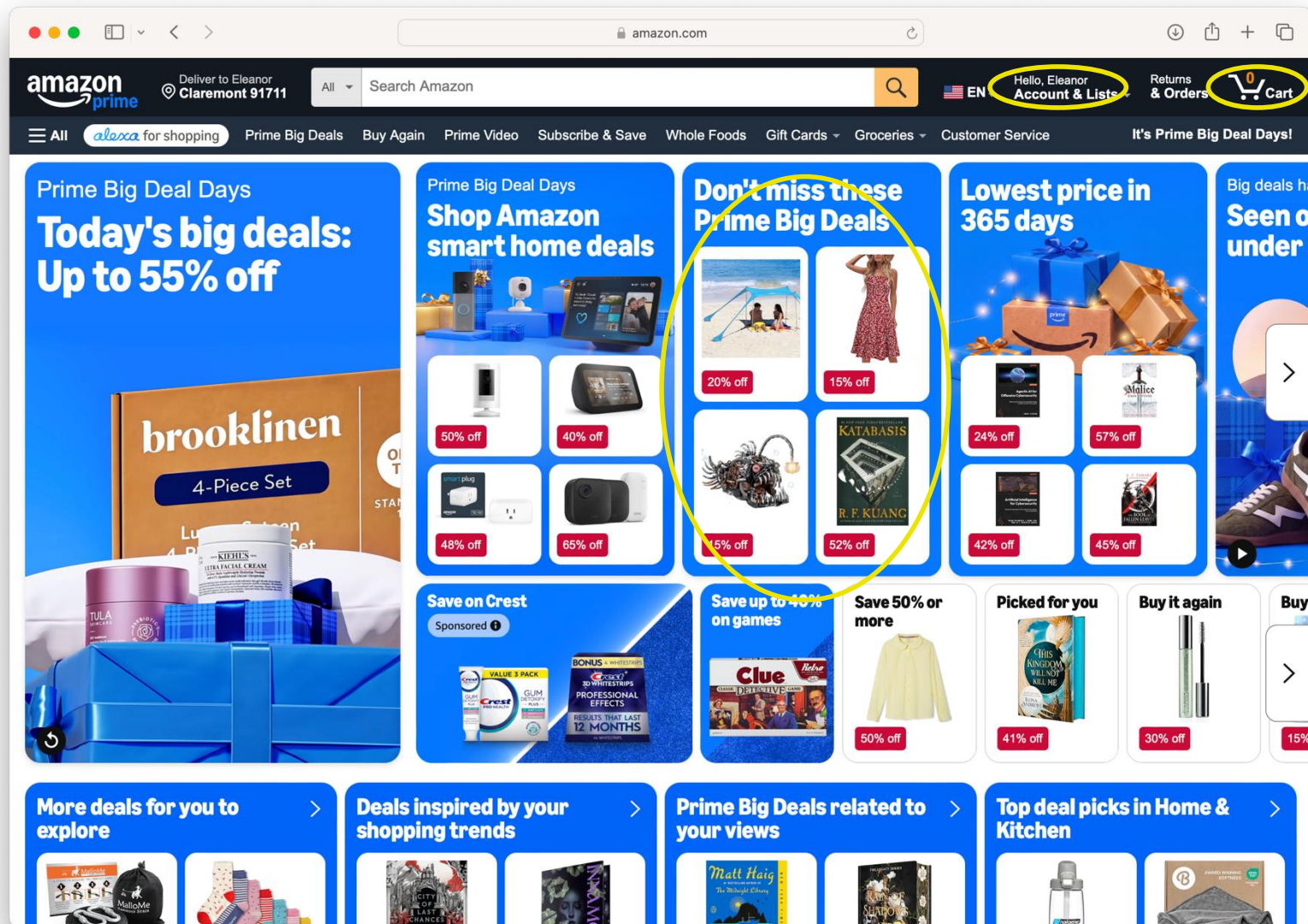
Key Observation: All of this is deterministic

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8">

    <title>CS 181E - Fall 2026</title>
    <link href='https://fonts.googleapis.com/css?family=Source+Sans+Pro:300,300i,600,700,700i' rel='stylesheet' type='text/css'>
    <link href='https://fonts.googleapis.com/css?family=Inconsolata:400,700,700i' rel='stylesheet' type='text/css'>

    <link href="resources/css/bootstrap.min.css" rel="stylesheet">
    <link rel="stylesheet" href="resources/css/main.css">
  </head>
  <body>
    <header class="site-header">
  <div class="navbar navbar-inverse navbar-fixed-top">
    <div class="container-fluid">
      <div class="navbar-header">
        <button type="button" class="navbar-toggle" data-toggle="collapse" data-target=".navbar-collapse">
          <span class="sr-only">Toggle navigation</span>
          <span class="icon-bar"></span>
          <span class="icon-bar"></span>
          <span class="icon-bar"></span>
        </button>
        <a class="navbar-brand"> CS 181E
          <span class="hidden-xs hidden-sm">: Computer Systems</span>
          <span class="hidden-md hidden-lg"> - Fall 2026</span>
        </a>
      </div>
    </div>
  </div>
```

Dynamically-Generated Content



Static and Dynamic Content

- The content returned in HTTP responses can be either **static** or **dynamic**
 - *Static content*: content stored in files and retrieved in response to an HTTP request
 - Examples: HTML files, images, audio clips
 - Request identifies which content file
 - *Dynamic content*: content produced on-the-fly in response to an HTTP request
 - Example: content produced by a program executed by the server on behalf of the client
 - Request identifies file containing executable code
- Bottom line: *Web content is associated with a file that is managed by the server*

Dynamic Web Pages

Server-Side

- PHP
- Ruby
- Python
- Java
- Go

Client-Side

- Javascript

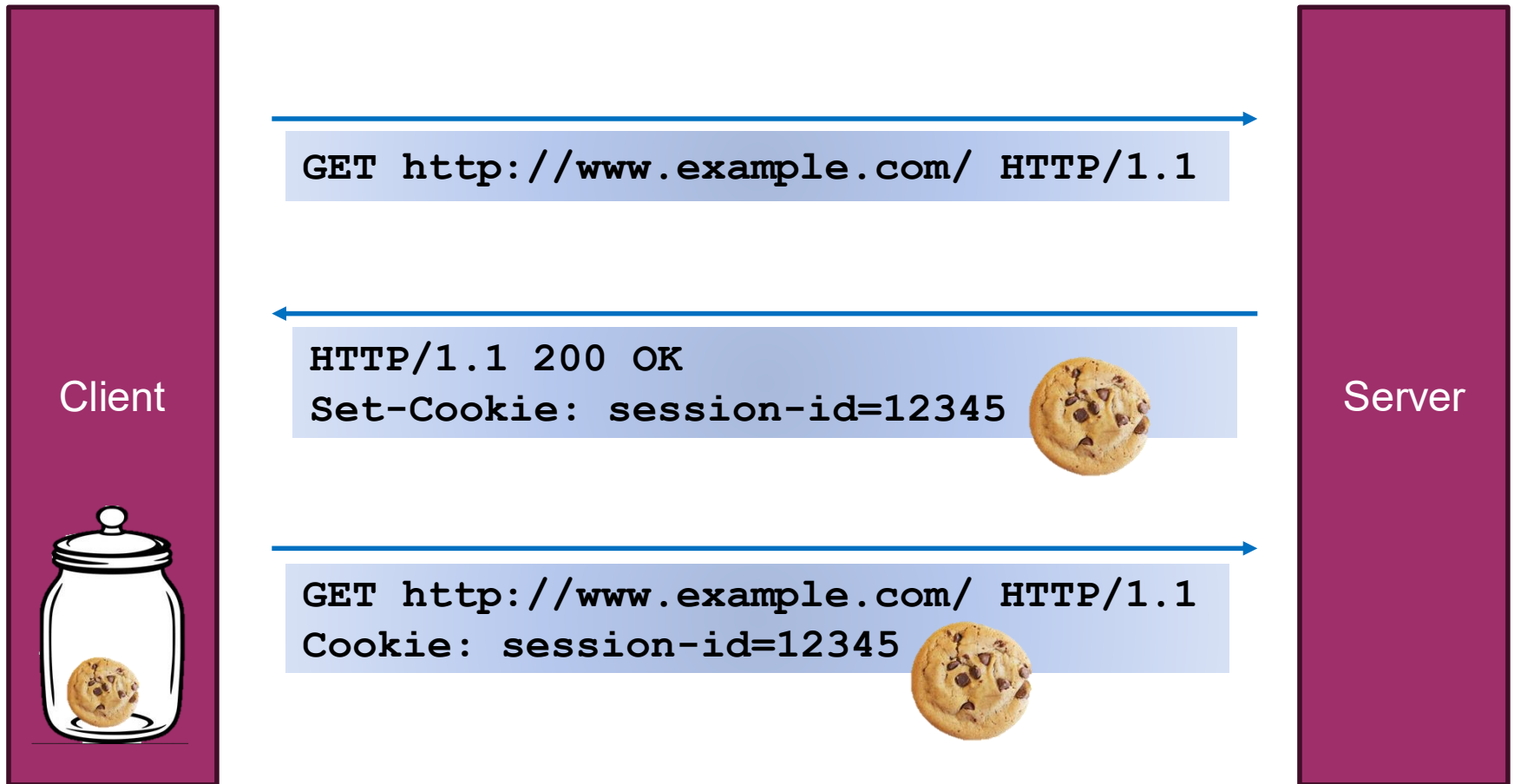
Key Observation: All of this is still stateless

Cookies

- Cookies are small blocks of data stored locally by the web browser
- Cookie is sent with every request to that domain
- Can be used to keep track of whether a user has authenticated (as which user)
- And also other things...
- Can be set by third parties

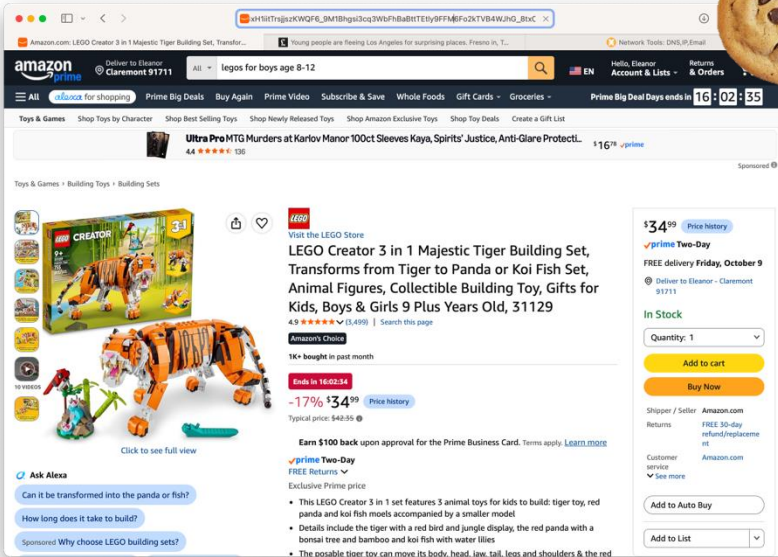


Cookies

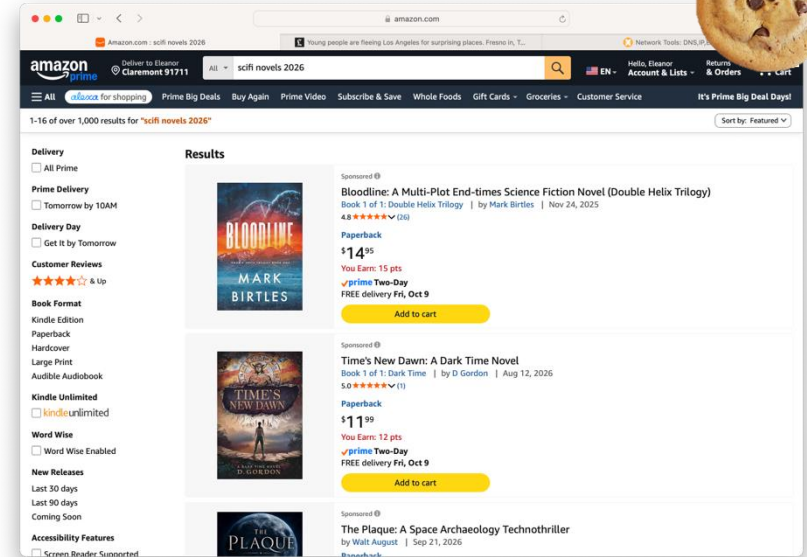


First party tracking

- uses cookies, javascript, and url parameters to track behavior on site



https://www.amazon.com/LEGO-Majestic-Building-Featuring-Imaginat.../dp/B09FM57L7B/ref=sr_1_1_sspa?crd=28QO8CV6XVUJO&dib=e...yJ2ljoIMSJ9.CgmS_iNoAEFAxySR6sPRkK0HLeFtyw9AMIqzNQCK8...xH1iitTrsjjszKWQF6_9M1Bhgsi3cq3WbFhBaBtTetly9FFM6Fo2kTVB...4WJhG_8txOOexcH89mgxN6yB1l8jg0rByrljMtk-JpW1P5GVsEUQ2II-jX...l6ZSPk01uwAjpg-JJ2smfK0pgay8DV-6xafyDF0vHKQI4eD1lwzquEix5...GpkZopJrTvZml56pqgrN59fSZZ_kDT1-xle6m0VJ81KbBEeNXX3afC7Y...r-X2Z_Ke0UZA9p92bfn7QFVAZ8UAcIA.KgOBGNSmCjKO_6yt5cXP8T...P8H0ty6SXgkSkxFWuL6l&dib_tag=se&keywords=legos+for+boys+a...ge+8-12&qid=1791384961&sprefix=legos+for+%2Caps%2C221&sr=8-1-spons&sp_csd=d2lkZ2V0TmFtZT1zcF9hdGY&psc=1



HTML DOM

```
<!DOCTYPE html>
<html>
```

```
<head>
  <title>CS 181E - Fall 2026</title>
```

```
  <link href='https://fonts.googleapis.com/css?family=Source+Sans+Pro:300,300i,600,700,700i'
    rel='stylesheet' type='text/css'>
  <link href='https://fonts.googleapis.com/css?family=Inconsolata:400,700,700i' rel='stylesheet'
    type='text/css'>
```

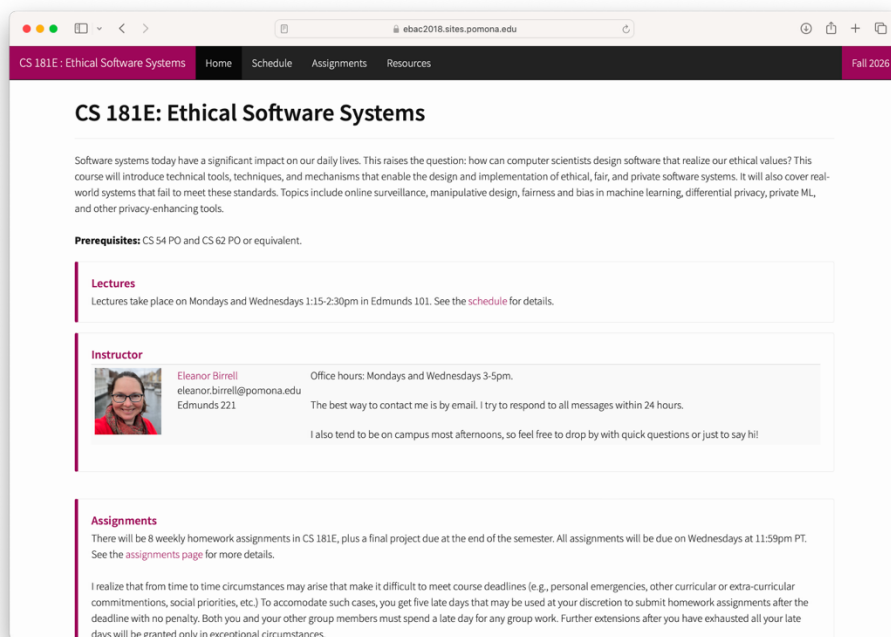
```
  <link href="resources/css/bootstrap.min.css" rel="stylesheet">
  <link rel="stylesheet" href="resources/css/main.css">
</head>
```

```
<body>
  <header class="site-header">
    <div class="navbar navbar-inverse navbar-fixed-top">
      <div class="container-fluid">
        <div class="navbar-header">
          <a class="navbar-brand" href="">CS 181E <span class="hidden-xs hidden-sm">: Ethical Software
            Systems</span> <span class="hidden-md hidden-lg"> - Fall 2026</span> </a>
        </div>
        <div class="collapse navbar-collapse">
          <ul class="nav navbar-nav">
            <li class="active"><a href="index.html">Home</a></li>
            <li><a href="schedule.html">Schedule</a></li>
            <li><a href="assignments.html">Assignments</a></li>
            <li><a href="resources.html">Resources</a></li>
          </ul>
          <ul class="nav navbar-nav navbar-right hidden-xs hidden-sm"><li><a>Fall 2026</a></li></ul>
        </div><!-- /.navbar-collapse -->
      </div><!-- /.container -->
    </div><!-- /.navbar -->
  </header>
```

```
  <div class="container">
    <div class="page-header"><h1>CS 181E: Ethical Software Systems</h1></div>
    <p>Software systems today have a significant impact on our daily lives. This raises the question: how can
      computer scientists design software that realize our ethical values? This course will introduce technical
      tools, techniques, and mechanisms that enable the design and implementation of ethical, fair, and private
      software systems. It will also cover real-world systems that fail to meet these standards. Topics include
      online surveillance, manipulative design, fairness and bias in machine learning, differential privacy, private
      ML, and other privacy-enhancing tools. <br/><br/><b>Prerequisites:</b> CS 54 PO and CS 62 PO or
      equivalent. </p>
```

```
  <div class="bs-callout bs-callout-class"><h4>Lectures</h4>
    <div><p>Lectures take place on Mondays and Wednesdays 1:15-2:30pm in Edmunds 101. See the <a
      href="schedule.html">schedule</a> for details. </p></div>
  </div>
```

```
<div class="bs-callout bs-callout-class"><h4>Instructor</h4><div>
  <table><tbody>
    <tr class="row1">
      <td></td>
      <td valign="top" width="200">
        <a href="http://www.cs.pomona.edu/~ebirrell">Eleanor Birrell</a><br/>
        eleanor.birrell@pomona.edu<br/>
        Edmunds 221 <br/>
      </td>
      <td valign="top">Office hours: Mondays and Wednesdays 3-5pm.
    </td>
  </tr>
  </tbody>
</table>
</div>
```



Third-Party Tracking Cookies

The screenshot shows the Los Angeles Times website in a browser. The page features a blue header with the site name, a navigation menu, and a search bar. Below the header is a large blue advertisement for Amazon Prime. The main content area displays a headline about young people fleeing Los Angeles, accompanied by a photo of a man. To the right, there is a section titled 'Subscribers are Reading' with a link to a review.

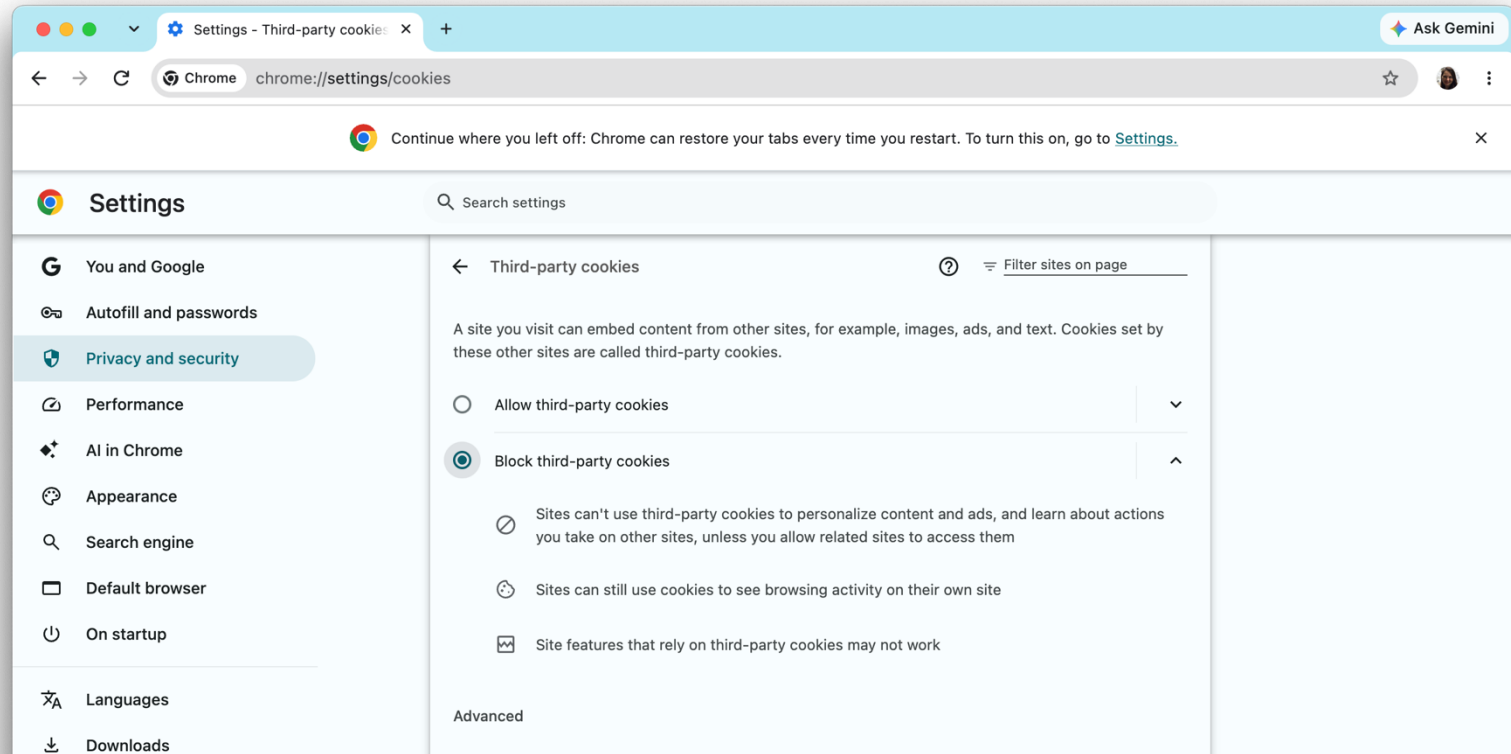
Below the main content, a network inspector tool is open, displaying a list of network requests. The 'Name' column shows various requests, including 'young-people-are-fleeing-los-angeles...', 'admi', 'about:blank', 'about:srcdoc', 'activeview', 'event', and '3pEvent'. The 'Domain' column shows the corresponding domains, such as 'www.latimes.com', 'web-cf-event-cell8', and 'pagead2.googlesyn'. The 'Type' column indicates the request type, mostly 'document' and 'fetch'. The 'Initiator' column shows the initiator, such as 'ujs_web_disp...' and 'Bootstrap.js:2...'. The 'Trans...' column shows the transfer size, and the 'Time' column shows the time taken for the request.

A yellow circle highlights the '340' in the bottom status bar, which represents the number of cookies stored by the browser. The status bar also shows the total size of the cookies (2.73 MB) and the number of cookies (1).

Exercise: Tracking Cookies

- Go to Chrome and turn on developer tools (find a partner with a laptop + Chrome installed if you don't)
- Pick a website that you use regularly
- Try to identify what requests it sends with cookies. Look up what these are.
- How many known companies did you find? How many other trackers/advertisers did you find?

Disabling 3rd-party cookies



Most modern browsers (except Chrome) disable third-party cookies by default (Safari, Firefox, Brave, DuckDuckGo, etc.)

Increasing rates of user rejection (decline cookie in banner, enable GPC, use ad blocker, etc.)

CNAME Cloaking

Domain Name Service (DNS)

- Maps domain name (e.g., www.cs.pomona.edu) to IP address
- DNS A record
- IP address is then used at the network level to send request to the right server

CNAME

- alternate type of DNS record
- redirects from domain name to another domain
- can be used to make 3rd party domains look like 1st party (e.g., redirect from tracker.pomona.edu to tracker.com)

URL Parameters

- Unique name for a file: URL (Universal Resource Locator)
- Example URL:

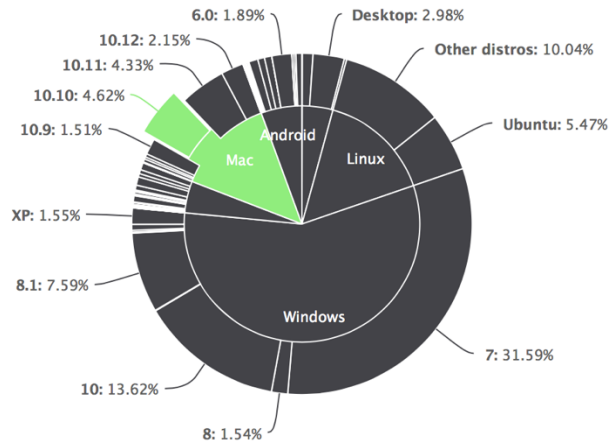
<code>http://www.cs.pomona.edu</code>	<code>/classes/cs181e/index.html</code>	<code>?id=1234&a=True</code>
prefix	suffix	parameters

- Clients use *prefix* (`http://www.cs.pomona.edu`) to infer:
 - What kind (protocol) of server to contact (HTTP)
 - Where the server is (`www.cs.pomona.edu`)
- Servers use *suffix* (`/classes/cs105/index.html`) to:
 - Specify what content they want
- Parameters are key-value pairs passed from client to server
 - usually used for internal information
 - can also pass tracking ID to third-parties

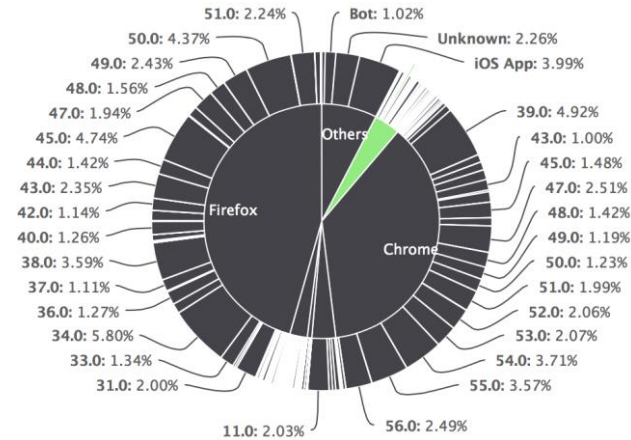
Browser Fingerprinting

- Key Idea: many features of the browser that are relatively unique to your machine
 - Fonts
 - GPU model anti-aliasing (Canvas fingerprinting)
 - User-agent string
- Use combination of device features as an identifier

Browser Fingerprinting



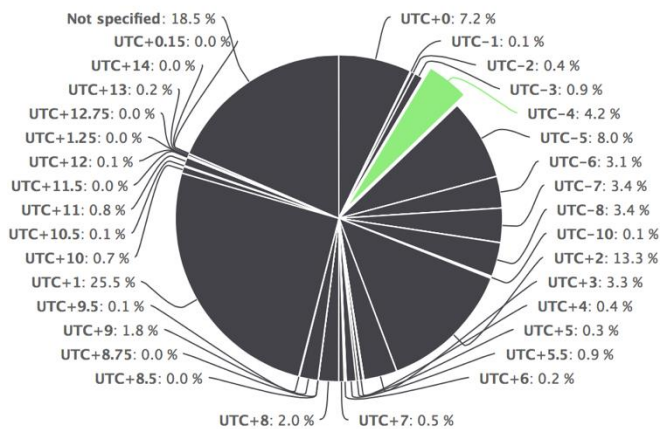
Operating System



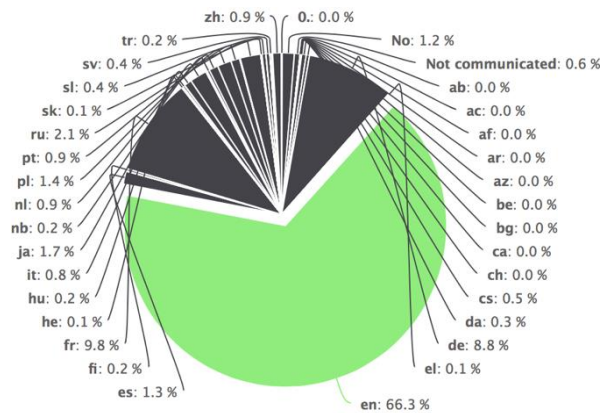
Browser

Cookies enabled ⓘ	77.92%	"yes"
Do Not Track ⓘ	30.91%	"yes"
Timezone ⓘ	4.23%	"240"
Screen resolution ⓘ	3.63%	"1280x800x24"
Use of local storage ⓘ	74.40%	"yes"
Use of session storage ⓘ	74.40%	"yes"

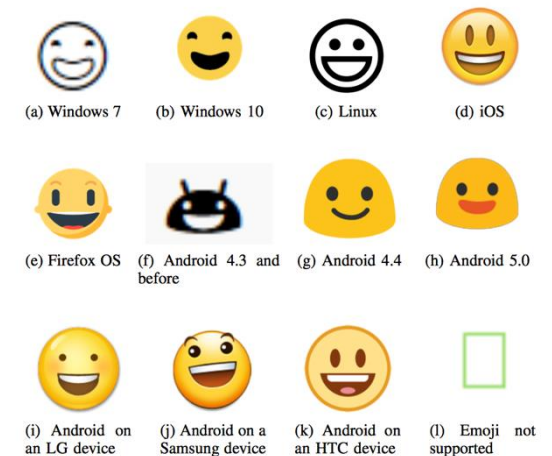
Settings



Time Zone



Language



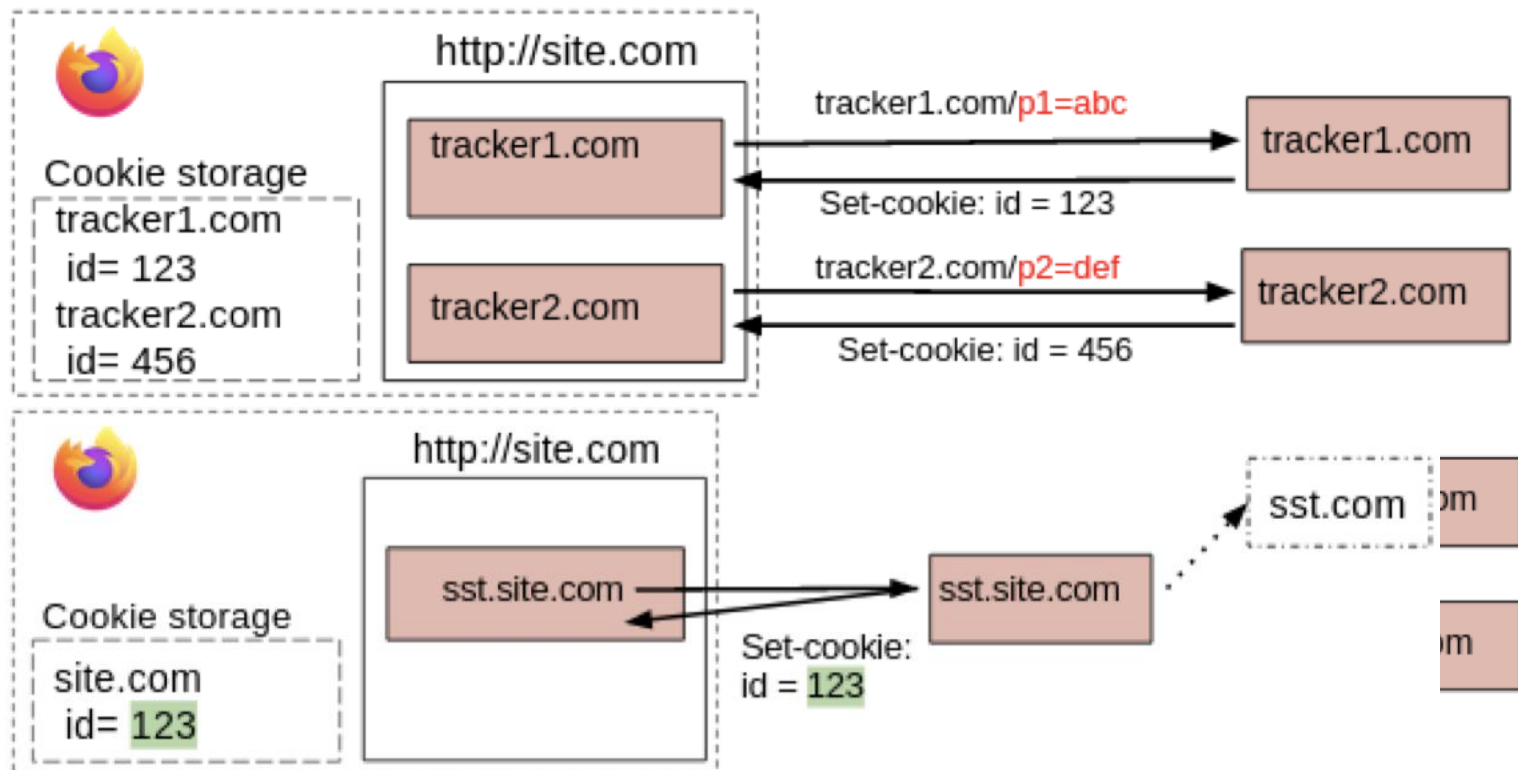
Canvas

Exercise: Browser Fingerprinting

- The EFF runs a project that lets you test how unique your browser configuration is: <https://coveryourtracks.eff.org/>
- Go there, test your browser, and read through the results
- How unique is your browser?
- What types of info were used to fingerprint you? Are any of them surprising?

Server-side Tracking

- Key idea: instead of browser sending info to third-parties, company collects data and then shares with third-parties on the backend/server side
- introduced by Google in 2020



Trackers and Advertisers

- trackers observe user behavior across one or (usually) many sites
- use ML to make inferences about user demographics and interests
- advertisers serve ads on first or third party sites
- get paid by companies that place ads
- can choose which ads to place based on context or user attributes

Some trackers are also advertisers (e.g., Google, Meta)

Some trackers are **data brokers** who sell collected and inferred information to other third parties

California Data Broker Registry

The screenshot displays the California Privacy Protection Agency's Data Broker Registry. The page includes a navigation bar with links to Announcements, Meetings & Events, Laws & Regulations, FAQ, Resources, Career Opportunities, File a Complaint, and Data Broker Registry. A search bar is located at the top right. Below the navigation bar, the page indicates that 603 data brokers are registered and provides a link to download registration information. Two filter sections are present: 'DATA COLLECTION INCLUDES' and 'DATA SOLD / SHARED TO'. The 'DATA COLLECTION INCLUDES' filter includes options like Minors' Data, Account Logins, IDs / SSN, Citizenship, Union Membership, Sexual Orientation, Gender Identity, Biometric Data, Precise Geolocation, and Reproductive Health. The 'DATA SOLD / SHARED TO' filter includes options like Foreign Actor, Federal Government, State Governments, Law Enforcement, and GenAI Developer. A search bar is provided for searching by data broker name or website. Below the filters, the page shows a table of selected data brokers, including 01Advertising Inc., 33 Mile Radius LLC, 33Across, Inc., 5x5 US, LLC, 6sense Insights, Inc., and A Direct Marketing Inc.

California Privacy Protection Agency

Search this website

Announcements Meetings & Events Laws & Regulations FAQ Resources Career Opportunities File a Complaint Data Broker Registry

Last Updated: 7/29/2026

603 data brokers registered

Download Registration Information

DATA COLLECTION INCLUDES Clear

Minors' Data Account Logins IDs / SSN Citizenship

Union Membership Sexual Orientation Gender Identity Biometric Data

Precise Geolocation Reproductive Health

DATA SOLD / SHARED TO Clear

Foreign Actor Federal Government State Governments Law Enforcement

GenAI Developer

Search by data broker name or website...

Selected all 603 data brokers

DATA BROKER NAME	WEBSITE	DATA COLLECTION INCLUDES	DATA SOLD/SHARED TO
01Advertising Inc.	01advertising.com	Sexual Orientation Gender Identity Precise Geolocation	— View
33 Mile Radius LLC	33mileradius.com	—	— View
33Across, Inc.	33across.com	—	— View
5x5 US, LLC	5x5data.com	Precise Geolocation	— View
6sense Insights, Inc.	6sense.com	—	— View
A Direct Marketing Inc			