11. Web Privacy

Blase Ur and Grant Ho
February 12\textsuperscript{th}, 2024
CMSC 23200
Additional Web Security Topics
Processing Data on the Server

- JavaScript is client-side
- Server-side you find Perl (CGI), PHP, Python (Django)
- Process data on the server
- What happens if this code crashes?
Storing Data on the Server

- Run a database on the server
- MySQL, SQLite, MongoDB, Redis, etc.
- You probably don’t want to allow access from anything other than localhost
- You definitely don’t want human-memorable passwords
- To emphasize the last lecture: use prepared statements and otherwise do worry about code injection!
CMS (Content Management System)

- WordPress (PHP + MySQL), Drupal
CMS Defaults / Vulnerabilities

• WordPress attempted logins:

```
root@super:/var/log/apache2# cat_error* | grep "wp-
[Fri Feb 18 09:05:49.042574 2022] [php7:error] [pid 3789616] [client 103.109.96.11:60066] script '/var/www/html/eusec20/wp-
/login.php' not found or unable to stat
[Thu Feb 17 08:23:31.605082 2022] [php7:error] [pid 3630350] [client 102.165.48.97:40892] script '/var/www/html/wp-login.php' not found or unable to stat
[Thu Feb 17 08:23:31.951171 2022] [php7:error] [pid 3631784] [client 102.165.48.97:40894] script '/var/www/html/eusec20/wp-
/login.php' not found or unable to stat
[Thu Feb 17 08:23:31.978838 2022] [php7:error] [pid 3632298] [client 102.165.48.97:40896] script '/var/www/html/eusec/wp-
login.php' not found or unable to stat
[Thu Feb 17 10:03:18.958818 2022] [php7:error] [pid 3641153] [client 47.104.66.61:58626] script '/var/www/html/interestsre-
search/wp-login.php' not found or unable to stat, referer: http://interestsresearch.io/wp-login.php
[Thu Feb 17 11:04:27.068009 2022] [php7:error] [pid 3646525] [client 80.251.219.111:60460] script '/var/www/html/computers-
ecurityclass.com/wp-login.php' not found or unable to stat, referer: http://computersecurityclass.com/wp-login.php
[Thu Feb 17 11:35:43.470994 2022] [php7:error] [pid 3649892] [client 107.173.165.214:34454] script '/var/www/html/aifairne-
sstech/wp-login.php' not found or unable to stat, referer: http://aifairness.tech/wp-login.php
```
Online Tracking
Online Tracking

- Advertisers want to show you advertisements targeted to your interests and demographics.
Data-Driven Inferences

You might like dogs!
Online Tracking

• First party = the site you are visiting (whose address is in the URL bar)
• Third party = other sites (i.e., origins) contacted as a result of your visit to the first party
• First-party tracking (on search engines, shopping sites)
• Third-party tracking (ads on lots of sites)
Mechanics of First-Party Online Tracking

- Use cookies, JavaScript, URL parameters to track
Mechanics of First-Party Online Tracking

https://www.amazon.com/Plant-Seeds-Succulent-Garden-Bonsai/dp/B08RRQR6RF/ref=sr_1_9?dchild=1&keywords=rare+plants&qid=1621969916&sr=8-9
Mechanics of Third-Party Online Tracking
Details of What’s Happening in HTTP (Request)

GET / HTTP/2
Host: www.uchicago.edu
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefox
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate, br
DNT: 1
Connection: keep-alive
Cookie: uchicago-prod_last_visit=1306604446; uchicago-prod_last_activity=1621964446;
Upgrade-Insecure-Requests: 1
If-Modified-Since: Tue, 25 May 2021 17:40:36 GMT
TE: Trailers
Details of What’s Happening in HTTP (Request)

Request Headers (735 B)

GET / HTTP/2
Host: www.uchicago.edu

User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefox
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate, br
DNT: 1
Connection: keep-alive
Cookie: uchicago-prod_last_visit=1306604446; uchicago-prod_last_activity=1621964446;
Upgrade-Insecure-Requests: 1
If-Modified-Since: Tue, 25 May 2021 17:40:36 GMT
TE: Trailers
Details of What’s Happening in HTTP (Cookies)

Image taken from https://networkencyclopedia.com/http-cookie/
Details of What's Happening in HTTP (Request)

Request Headers (735 B)

GET / HTTP/2
Host: www.uchicago.edu
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefox
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate, br
DNT: 1
Connection: keep-alive
Cookie: uchicago-prod_last_visit=1306604446; uchicago-prod_last_activity=1621964466;
Upgrade-Insecure-Requests: 1
If-Modified-Since: Tue, 25 May 2021 17:40:36 GMT
TE: Trailers
Details of What’s Happening in HTTP (Response)

HTTP/2 200 OK
date: Tue, 25 May 2021 18:00:35 GMT
content-type: text/html; charset=UTF-8
server: Apache
x-frame-options: SAMEORIGIN
expires: Mon, 26 Jul 1997 05:00:00 GMT
pragma: no-cache
vary: Accept-Encoding
set-cookie: uchicago-prod_last_visit=1306605629; expires=Wed, 25-May-2022 18:00:29 GMT
set-cookie: uchicago-prod_last_activity=1621965629; expires=Wed, 25-May-2022 18:00:2
set-cookie: uchicago-prod_tracker=%7B%22%22%22%22index%22%2C%22token%22%22%22%22%2226944a
set-cookie: uchicago-prod_CSRF_token_deleted; expires=Thu, 01-Jan-1970 00:00:01 GMT;
set-cookie: uchicago-prod_CSRF_token=86d47d8690aa7646e1628dd095cd5b464db16bd3; expires=Thu, 01-Jan-1970 00:00:01 GMT;
last-modified: Tue, 25 May 2021 18:00:29 GMT
content-encoding: gzip
x-varnish: 10696657 9201444
age: 5
via: 1.1 varnish (Varnish/5.2)
accept-ranges: bytes
X-Firefox-Spdy: h2
Details of What’s Happening in HTTP (Response)

Response Headers (1.078 KB)

HTTP/2 200 OK
date: Tue, 25 May 2021 18:00:35 GMT
content-type: text/html; charset=UTF-8
server: Apache
x-frame-options: SAMEORIGIN
expires: Mon, 26 Jul 1997 05:00:00 GMT
pragma: no-cache
vary: Accept-Encoding

set-cookie: uchicago-prod_last_visit=1306605629; expires=Wed, 25-May-2022 18:00:29 GMT
set-cookie: uchicago-prod_last_activity=1621965629; expires=Wed, 25-May-2022 18:00:29 GMT
set-cookie: uchicago-prod_tracker=%7B%22%22%3A%22index%22%2C%22token%22%3A%2226944a
set-cookie: uchicago-prod_csrf_token=deleted; expires=Thu, 01-Jan-1970 00:00:01 GMT;
set-cookie: uchicago-prod_csrf_token=86d47d8690aa7646e1628dd095cd5b464db16bd3; expires=

last-modified: Tue, 25 May 2021 18:00:29 GMT
content-encoding: gzip
x-varnish: 10696657 9201444
age: 5
via: 1.1 varnish (Varnish/5.2)
accept-ranges: bytes
X-Firefox-Spdy: h2
## HTTP Headers (uchicago.edu → youtube.com)

<table>
<thead>
<tr>
<th>Method</th>
<th>Domain</th>
<th>File</th>
<th>Initiator</th>
<th>Type</th>
<th>Transferred</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET</td>
<td><a href="http://www.uchicago.edu">www.uchicago.edu</a></td>
<td>/</td>
<td></td>
<td>html</td>
<td>114.1 KB</td>
<td>39 KB</td>
</tr>
<tr>
<td>POST</td>
<td><a href="http://www.youtube.com">www.youtube.com</a></td>
<td>/</td>
<td></td>
<td>html</td>
<td>604.8 B</td>
<td>0 B</td>
</tr>
<tr>
<td>GET</td>
<td><a href="http://www.youtube.com">www.youtube.com</a></td>
<td>P-xofi-782U/autohide=1&amp;fs=1&amp;autoplay: subdocument</td>
<td>html</td>
<td>21.81 KB</td>
<td>51 B</td>
<td></td>
</tr>
<tr>
<td>GET</td>
<td>cdn.hype.mar</td>
<td>uchicago.www/width=1169&amp;paginate=true</td>
<td>a5B5e5.js3</td>
<td>128.06 KB</td>
<td>12 KB</td>
<td></td>
</tr>
<tr>
<td>GET</td>
<td>cdn.hype.mar</td>
<td>popUpModeEndpoint</td>
<td>a5B5e5.js3</td>
<td>10.95 KB</td>
<td>10 KB</td>
<td></td>
</tr>
</tbody>
</table>

### Response Headers (642 B)

- **HTTP/3 200 OK**
  - content-type: text/html; charset=utf-8
  - x-content-type-options: nosniff
  - cache-control: no-cache, no-store, max-age=0, must-revalidate
  - pragma: no-cache
  - expires: Mon, 01 Jan 1990 00:00:00 GMT
  - date: Tue, 25 May 2021 18:36:36 GMT
  - strict-transport-security: max-age=31536000
  - permissions-policy: ch-ua-full-version*, ch-ua-platform=*, ch-ua-platform-version=*
  - content-encoding: br
  - server: ESF
  - x-xss-protection: 0
  - alt-svc: h2-29;":443"; ma=2592000,h3-29;":443"; ma=2592000,h3-7051;":443"; ma=2592000,h3-9050;":443"; ma=2592000,h3-9051;":443"

### Request Headers (621 B)

- GET /embed/p-xofi-782U/autohide=1&fs=1&autoplay=0&rel=0&modestbranding=1&showinfo=0&hd=1&me Host: www.youtube.com
  - User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefox/88.0
  - Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
  - Accept-Language: en-US,en;q=0.5
  - Accept-Encoding: gzip, deflate
  - DNT: 1
  - Alt-Used: www.youtube.com
  - Connection: keep-alive
  - Referer: https://www.uchicago.edu/
<table>
<thead>
<tr>
<th>Status</th>
<th>Method</th>
<th>Domain</th>
<th>File</th>
<th>Initiator</th>
<th>Type</th>
<th>Transferred</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>GET</td>
<td><a href="http://www.uchicago.edu">www.uchicago.edu</a></td>
<td>/</td>
<td>document</td>
<td>html</td>
<td>11.41 KB</td>
<td>39...</td>
</tr>
<tr>
<td>204</td>
<td>POST</td>
<td><a href="http://www.youtube.com">www.youtube.com</a></td>
<td>atr?ns=yt&amp;el=embedded&amp;cpn=ho5PKBh-</td>
<td>base.js:1023</td>
<td>html</td>
<td>604 B</td>
<td>0 B</td>
</tr>
<tr>
<td>200</td>
<td>GET</td>
<td><a href="http://www.youtube.com">www.youtube.com</a></td>
<td>P-xlixF7B2U?autohide=1&amp;fs=1&amp;autoplay=</td>
<td>subdocument</td>
<td>html</td>
<td>21.81 KB</td>
<td>51...</td>
</tr>
<tr>
<td>200</td>
<td>GET</td>
<td>cdn.hypemar.com</td>
<td>uchicagowww?width=1169&amp;paginate=true</td>
<td>a5b5e5.js:3</td>
<td>html</td>
<td>128.06 KB</td>
<td>12...</td>
</tr>
<tr>
<td>200</td>
<td>GET</td>
<td>cdn.hypemar.com</td>
<td>popUpModalEndpoint</td>
<td>a5b5e5.js:3</td>
<td>html</td>
<td>10.99 KB</td>
<td>10...</td>
</tr>
</tbody>
</table>
HTTP Headers (uchicago.edu → youtube.com)

Request Headers (621 B)

GET /embed/P-lixirF7B2U?autohide=1&fs=1&autoplay=0&rel=0&modestbranding=1&showinfo=0&hd=1&
Host: www.youtube.com
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefox/88.0
Accept: text/html, application/xhtml+xml, application/xml; q=0.9, image/webp, */*; q=0.8
Accept-Language: en-US, en; q=0.5
Accept-Encoding: gzip, deflate, br
DNT: 1
Alt-Used: www.youtube.com
Connection: keep-alive
Referer: https://www.uchicago.edu/
Cookie: VISITOR_INFO1_LIVE=dACtKpaJViQ; PREF=tz=America.Chicago&f4=40000000; YSC=p2jSvxCMel
Upgrade-Insecure-Requests: 1
TE: Trailers
HTTP Headers (uchicago.edu → youtube.com)

GET /embed/P-xlixF7B2U?autoplay=1&fs=1&autoplay=0&rel=0&modestbranding=1&showinfo=0&hd=1&
Host: www.youtube.com

User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefox/88.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate, br
DNT: 1
Alt-Used: www.youtube.com
Connection: keep-alive
Referer: https://www.uchicago.edu/
Cookie: VISITOR_INFO1_LIVE=dACtKPaJViQ; PREF=tz=America.Chicago&f4=4000000; YSC=p2jSvxCMel
Upgrade-Insecure-Requests: 1
TE: Trailers
HTTP Headers (uchicago.edu → youtube.com)

Request Headers (621 B)

GET /embed/P-xlixF7B2U?autoplay=1&fs=1&autoplay=0&rel=0&modestbranding=1&showinfo=0&hd=1&
Host: www.youtube.com
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefox/88.0
Accept: text/html, application/xhtml+xml, application/xml;q=0.9, image/webp, */*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate, br
DNT: 1
Alt-Used: www.youtube.com
Connection: keep-alive
Referer: https://www.uchicago.edu/
Cookie: VISITOR_INFO1_LIVE=dACtKPajViQ; PREF=tz=America.Chicago&f4=4000000; YSC=p2jSvxCM ei
Upgrade-Insecure-Requests: 1
TE: Trailers
Putting It Together

• (Unless browser is blocking it) third party gets its cookies
• (Unless browser is blocking it) third party sees “referer” [sic]
• First party can choose to send info to third party via URL parameters (not a violation of Same Origin Policy!)
• Third party sees this information for many first parties
Mechanics of Cookie Syncing

- JavaScript / images from advertising networks loaded as part of your page
  - In iframes
  - Or sometimes not
  - Why does this matter?
- Let’s discuss: what can an ad network learn, and how?
Mechanics of Cookie Syncing

Track Visited Sites

• Subtle side channel!

• (Loophole has since mostly been closed)
Browser Fingerprinting

• Use features of the browser that are relatively unique to your machine
  – Fonts
  – GPU model anti-aliasing (Canvas fingerprinting)
  – User-agent string
  – *(Often not)* IP address *(Why not?)*
Browser Fingerprinting

- Use combination of device features as an identifier
- [https://coveryourtracks.eff.org/](https://coveryourtracks.eff.org/)
Alternatives to Cookies for Tracking / Profiling
Google’s FLoC

• Federated Learning of Cohorts

• Clusters users based on their browsing activity and assigns a cohort ID
  – Uses SimHash for clustering
  – Clusters intended to contain 1,000s of users

• Criticisms include fingerprintability, ability to tie cohort to PII, and collapse of different browsing contexts

• (Abandoned in early 2022)
Google’s FLoC

Selecting Interest-based Ads Using FLoC

1. Browsers use a FLoC service to get the mathematical model, consisting of many calculated ‘cohorts.’ In this model, each cohort corresponds to many web browsers having similar recent browsing histories and contains a unique ID.

2. Using that FLoC Model algorithm, your browser calculates your cohort.

3. Let’s say you visited the site of an advertiser abc.com that sells kitchen appliances. Then that site requests the cohort ID from your browser.

4. If you visited additional pages of the advertiser, like searching kitchen utensils, it would record those interests.

5. Advertisers record these cohort activities periodically and share that information with the ad tech company that helps to deliver advertisements.

6. In the same manner, let’s say you visited a publisher site that sells ad space, it will also request your cohort ID.

7. Then the publisher site requests advertisements relevant to that cohort from the ad tech company.

8. The ad tech company combines the data received from the advertiser company about the cohort’s interests and data from the publishing company.

9. Next, the ad tech company chooses suitable ads according to the interests of the cohort.

10. The publisher site then displays the selected advertisement relevant to the interests of the cohort.
Google’s FLoC

Google’s FLoC Is a Terrible Idea

By BENNETT CYPHERS | MARCH 3, 2021

Update, April 4, 2022: We’ve launched Am I FLoC’d, a new site that will tell you whether your Chrome browser has been turned into a guinea pig for Federated Learning of Cohorts (FLoC), Google’s latest targeted advertising experiment.

The one-party cookie is dying, and Google is trying to create its replacement. But no one should mourn the death of the cookie as we know it. For more than two decades, the third-party cookie has been theynchpin in a shadowy, seedy, multi-billion-dollar advertising-compliance industry on the Web, ushering out.

Join Our Newsletter!
Email updates on news, actions, events in your area, and more.

Email Address
Postal Code (optional)

Anti-spam question: Enter the three-letter abbreviation for Electronic Frontier Foundation.

Image taken from https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea
# Google’s FLoC

<table>
<thead>
<tr>
<th>FLoC service</th>
<th>Browser</th>
<th>Advertiser</th>
<th>Publisher</th>
<th>Adtech</th>
<th>Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create FLoC model</td>
<td>Calculate browser cohort</td>
<td>Observe cohort activity</td>
<td>Observe visitor cohorts</td>
<td>Select ads relevant to cohort</td>
<td>Display relevant ads</td>
</tr>
<tr>
<td>The FLoC service creates a mathematical model with thousands of “cohorts”. Each cohort corresponds to thousands of web browsers with similar recent browsing histories.</td>
<td>A user’s browser, on their device, uses the FLoC model to calculate which cohort corresponds most closely to its own browsing history.</td>
<td>An advertiser observes the activity of cohorts on its site and shares this with its adtech platforms.*</td>
<td>An publisher observes the cohorts of visitors to its site and shares this with its adtech platforms.*</td>
<td>An adtech platform selects ads appropriate for the user’s cohort.</td>
<td>The publisher displays ads that are relevant to the user based on their browser’s cohort.</td>
</tr>
</tbody>
</table>

* The adtech platform may handle observing the cohort information on behalf of the advertiser or publisher.
Google’s FLoC

Selecting Interest-based Ads Using FLoC

1. Browsers use a FLoC service to get the mathematical model, consisting of many calculated “cohorts.” In this model, each cohort corresponds to many web browsers having similar recent browsing histories and contains a unique ID.

2. Using that FLoC Model algorithm, your browser calculates your cohort.

3. Let’s say you visited the site of an advertiser abc.com that sells kitchen appliances. Then that site requests the cohort ID from your browser.

4. If you visited additional pages of the advertiser, like searching kitchen utensils, it would record those interests.

5. Advertisers record those cohort activities periodically and share that information with the ad tech company that helps to deliver advertisements.

6. In the same manner, let’s say you visited a publisher site that sells ad space; it will also request your cohort ID.

7. Then the publisher site requests advertisements relevant to that cohort from the ad tech company.

8. The ad tech company combines the data received from the advertiser company about the cohort’s interests and data from the publishing company.

9. Next, the ad tech company chooses suitable ads according to the interests of the cohort.

10. The publisher site then displays the selected advertisement relevant to the interests of the cohort.

Image taken from https://www.privacyaffairs.com/google-floc/
Google's new concept assigns users five interests per week based on web activity.

By Emma Roth | Updated Jan 25, 2022, 2:45pm EST
Google’s Topics API

Your browser will store these topics for three weeks before deleting them. Google says that these categories “are selected entirely on your device” and don’t involve “any external servers, including Google servers.” When you visit a website, Topics will show the site and its advertising partners just three of your interests, consisting of “one topic from each of the past three weeks.”

As noted on the [Topics API GitHub](https://www.thewerge.com/2022/1/25/22900567/google-floc-abandon-topcs-api-cookies-tracking) page, there are currently about 350 available topics in its advertising taxonomy (although Google plans on adding anywhere from “a few hundred” to “a few thousand” eventually). Google says Topics won’t include any “sensitive categories” like race or gender. And if you’re using Chrome, the company is building tools to let you view and delete topics, as well as turn off the feature.