

11. Web Privacy



Blase Ur and Grant Ho
February 12th, 2024
CMSC 23200



THE UNIVERSITY OF
CHICAGO

Additional Web Security Topics

Processing Data on the Server

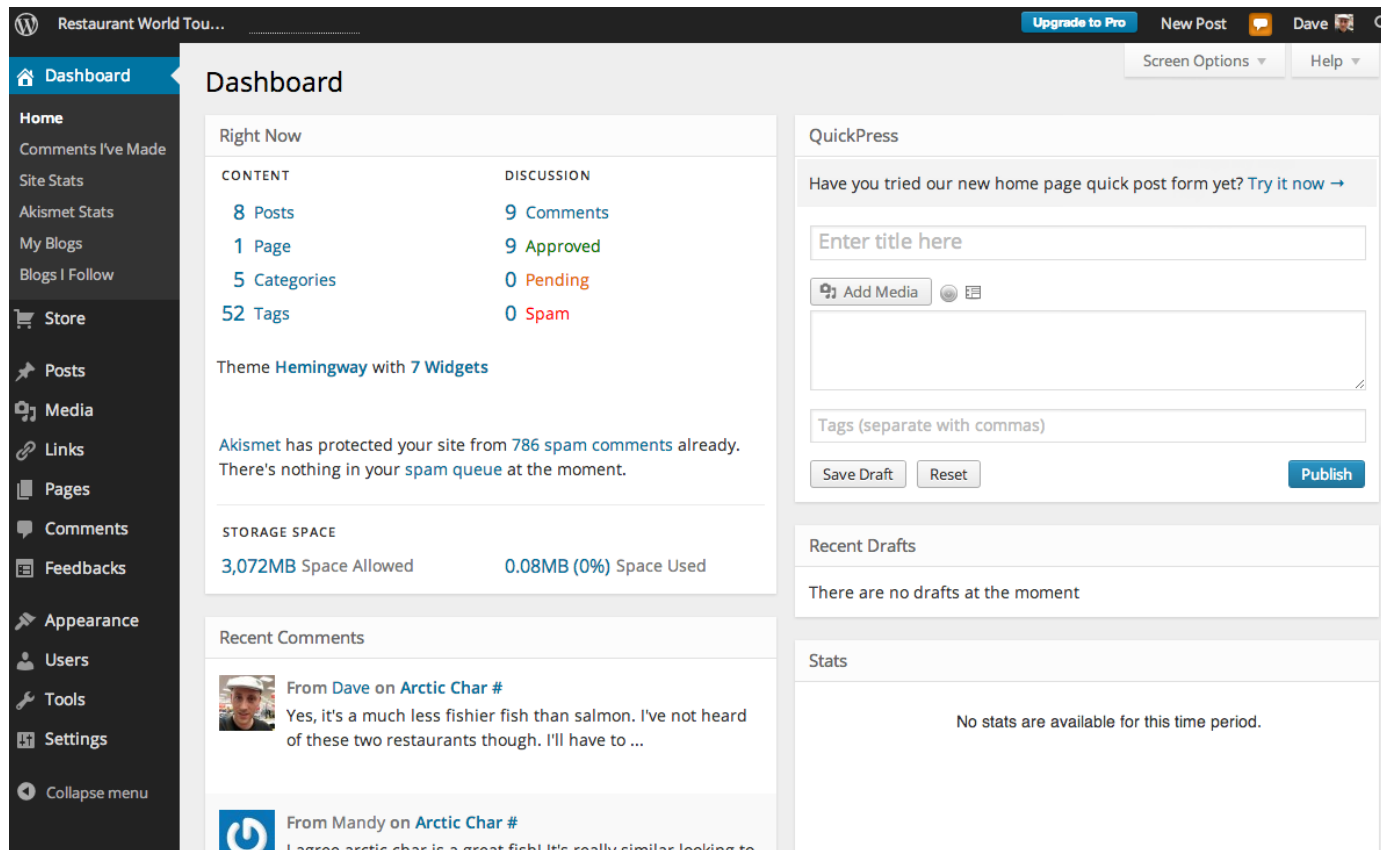
- JavaScript is client-side
- Server-side you find Perl (CGI), PHP, Python (Django)
- Process data on the server
- What happens if this code crashes?

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.p
hp' 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-l
ogin.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/computerse
curityclasscom/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

Ads Preferences

† Ads on Search and Gmail

† **Ads on the web**

Opt out

How your ads are personalized

Ads are based on personal info you've added to your Google Account, data from advertisers that partner with Google, and Google's estimation of your interests. Choose any factor to learn more or update your preferences. [Learn more](#)

Accounting & Finance Jobs	Action & Adventure Films
Action & Platform Games	Adventure Games
Android OS	Autos & Vehicles
Banking	Bars, Clubs & Nightlife
Beaches & Islands	Blues
Bollywood & South Asian Film	Books & Literature
Business & Productivity Software	Business News

Ads on the web

Make the ads you see on the web more interesting

Many websites, such as news sites and blogs, partner with us to show ads to their visitors. To see ads that are more related to you and your interests, edit the categories below, which are based on sites you have recently visited. [Learn More](#)

Your interests are associated with an advertising cookie that's stored in your browser. If you don't want us to store your interests, you can opt out below. Your ads preferences only apply in this browser on this computer. They are reset if you delete your browser's cookies.

† [Watch a video: Ads Preferences on GDN explained](#)

Your categories

Below you can review the interests and inferred demographics that Google has associated with your cookie. You can [remove](#) or [edit](#) these at any time.

Arts & Entertainment

Computers & Electronics

Computers & Electronics - Consumer Electronics - Gadgets & Portable Electronics - PDAs & Handhelds

Internet & Telecom

Internet & Telecom - Mobile & Wireless - Mobile Phones - Smart Phones

Law & Government

Science

Your demographics

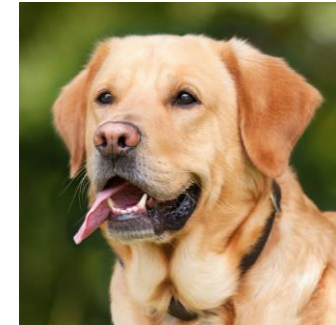
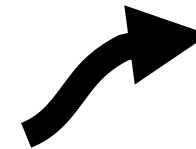
We infer your age and gender based on the websites you've visited. You can [remove](#) or [edit](#) these at any time.

Age: 35-44

Gender: Male

Google

Data-Driven Inferences



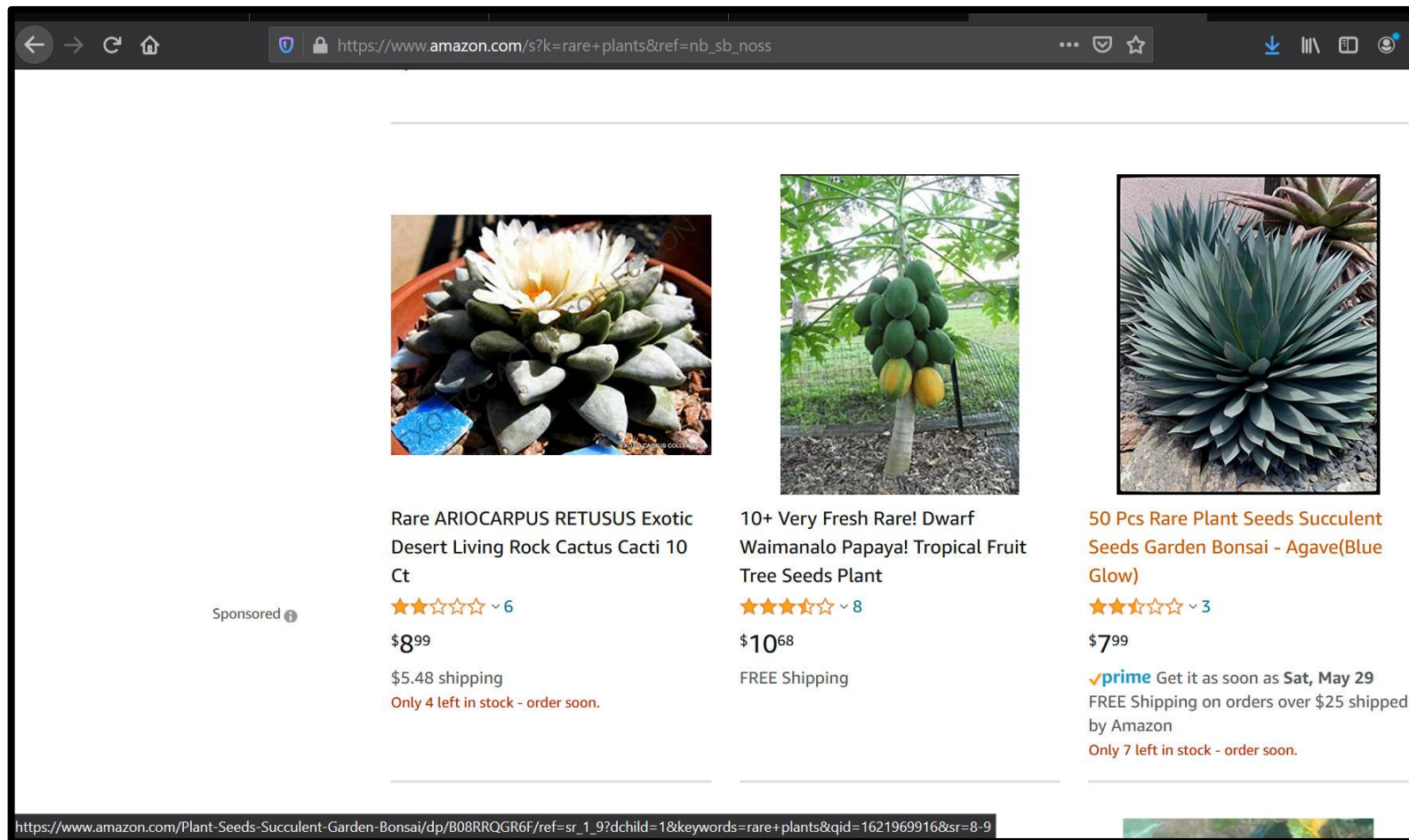
You might like dogs!

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

<p>Sponsored ⓘ</p>	<p>Desert Living Rock Cactus Cacti 10 Ct</p> <p>★★★★☆ ∨ 6</p> <p>\$8⁹⁹</p> <p>\$5.48 shipping</p> <p>Only 4 left in stock - order soon.</p>	<p>Waimanalo Papaya! Tropical Fruit Tree Seeds Plant</p> <p>★★★★☆ ∨ 8</p> <p>\$10⁶⁸</p> <p>FREE Shipping</p>	<p>Seed Glow</p> <p>★★</p> <p>\$7⁹⁹</p> <p>✓prime</p> <p>FREE by An</p> <p>Only 7</p>
--------------------	--	---	--

https://www.amazon.com/Plant-Seeds-Succulent-Garden-Bonsai/dp/B08RRQGR6F/ref=sr_1_9?dchild=1&keywords=rare+plants&qid=1621969916&sr=8-9

Mechanics of Third-Party Online Tracking

The screenshot displays the UChicago News website. At the top, the browser address bar shows 'https://www.uchicago.edu' with a 67% zoom level. The main heading is 'UChicago News', followed by a sub-header: '- Visit the **UChicago Forward website**, for the University's COVID-19 health protocols, campus guidelines, and other Spring Quarter information.'

Below the header, there are three featured news items, each with a thumbnail image and a title:

- What Americans are thinking a year after George Floyd's death**: The thumbnail shows a street scene with many people and flowers.
- Big Brains podcast: Why You're Likely Paying An Unfair Share of Property Taxes**: The thumbnail shows a row of small, stylized houses.
- University announces COVID-19 vaccine requirement for all students**: The thumbnail shows a modern building with a large tree in the foreground.

To the right of these items is a section titled 'Latest News' with the text: 'Founded at UChicago, nonprofit Climate Vault proposes new solution for carbon reduction'. A 'MORE NEWS >' link is located at the bottom right of this section.

At the bottom of the page, there is a large video player area. The video title is 'Explore Chicago: Discover the Global City UChicago...'. The video player includes a 'Watch later' button and a 'Share' button. The video content shows a street scene with a large building and the text 'A WORLD-CLASS UNIVERSITY' overlaid. Below the video player, there is a 'Watch on YouTube' button. To the right of the video player is a section titled 'Explore Chicago' with the text: 'Discover the global city UChicago calls home—filled with inspiration, innovation, and countless opportunities to explore.' A 'VISIT UCHICAGO >' link is located at the bottom right of this section.

Details of What's Happening in HTTP (Request)

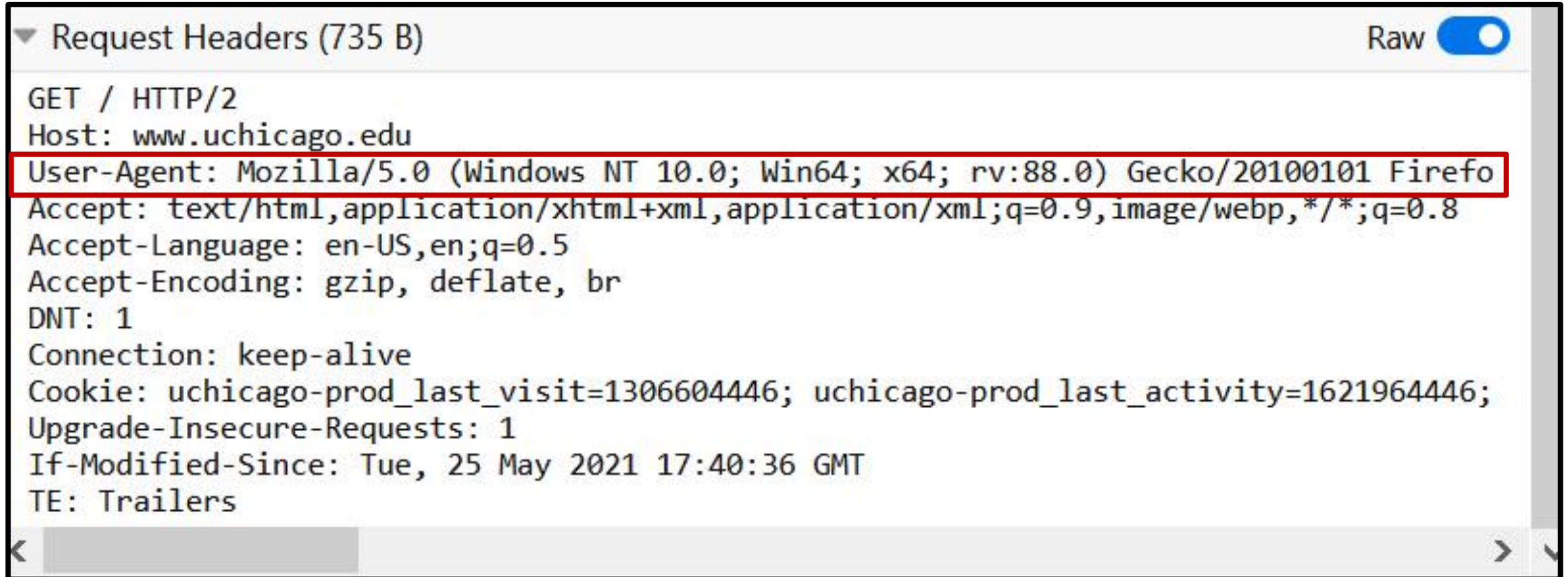
▼ Request Headers (735 B) Raw ☒

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

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
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)

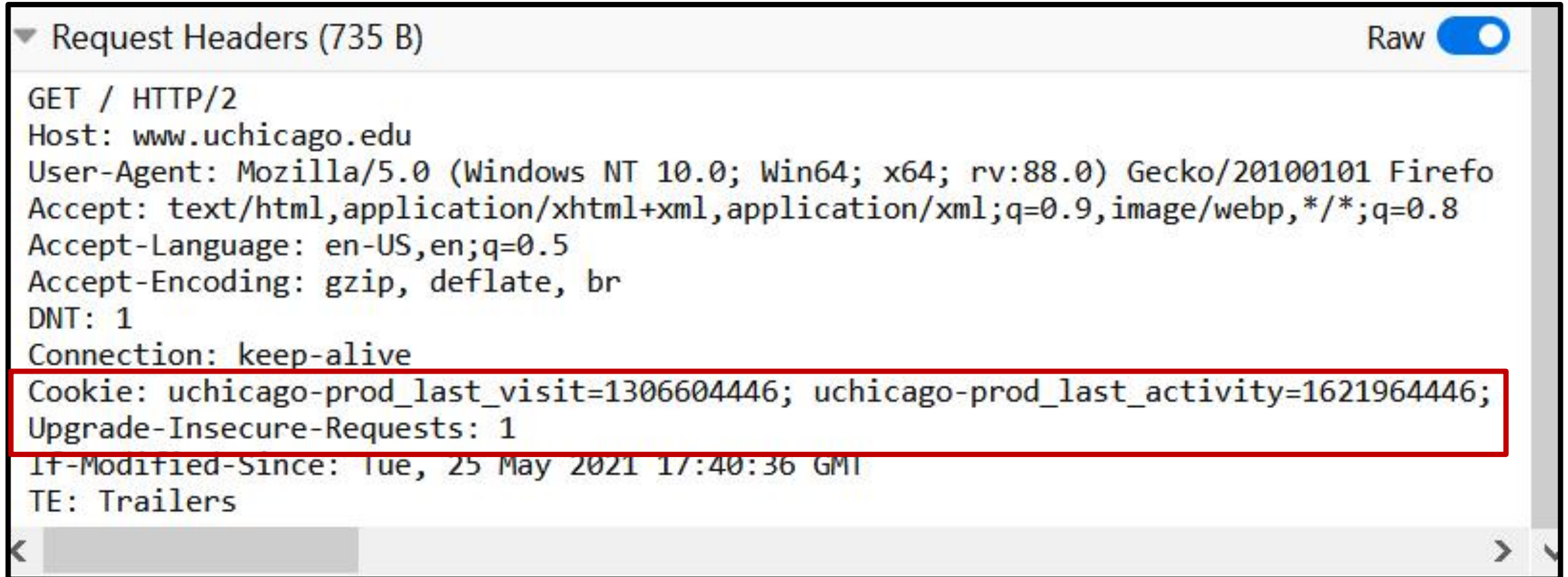


```
▼ Request Headers (735 B) Raw ☒  
GET / HTTP/2  
Host: www.uchicago.edu  
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefo  
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8  
Accept-Language: en-US,en;q=0.5  
Accept-Encoding: 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)



Details of What's Happening in HTTP (Request)



```
▼ Request Headers (735 B) Raw  
GET / HTTP/2  
Host: www.uchicago.edu  
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:88.0) Gecko/20100101 Firefo  
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 (Response)



```
▼ Response Headers (1.078 KB) Raw  
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 G  
set-cookie: uchicago-prod_last_activity=1621965629; expires=Wed, 25-May-2022 18:00:2  
set-cookie: uchicago-prod_tracker=%7B%220%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; expir  
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) Raw ☒

```
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 G
set-cookie: uchicago-prod_last_activity=1621965629; expires=Wed, 25-May-2022 18:00:2
set-cookie: uchicago-prod_tracker=%7B%220%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; expir
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)

Inspector
Console
Debugger
Style Editor
Performance
Memory
Network
Storage
Accessibility
Application
13

Filter URLs
All
HTML
CSS
JS
XHR
Fonts
Images
Media
WS
Other
Disable Cache
No Throttling

Status	Met...	Domain	File	Initiator	Type	Transferred	Size	Headers	Cookies	Request	Response	Timings	Security
200	GET	www.uchica...	/	document	html	11.41 KB	39...						
204	POST	www.youtub...	atr?ns=yt&el=embedded&cpn=ho5PKBh-	base.js:1023 (...)	html	604 B	0 B						
200	GET	www.youtub...	P-xlixF7B2U?autohide=1&fs=1&autoplay=	subdocument	html	21.81 KB	51...						
200	GET	cdn.hypemar...	uchicagowww?width=1169&paginate=tru	a5b5e5.js:3 (s...	html	128.06 KB	12...						
200	GET	cdn.hypemar...	popUpModalEndpoint	a5b5e5.js:3 (s...	html	10.99 KB	10...						

Filter Headers

Status: 200 OK
Version: HTTP/3
Transferred: 21.81 KB (51.50 KB size)
Referrer Policy: strict-origin-when-cross-origin

Response Headers (642 B)
Raw

```

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 18:00:36 GMT
strict-transport-security: max-age=31536000
permissions-policy: ch-ua-full-version=*, ch-ua-platform=*, ch-ua-platform-version=*, ch-ua
content-encoding: br
server: ESF
x-xss-protection: 0
alt-svc: h3-29=":443"; ma=2592000,h3-T051=":443"; ma=2592000,h3-Q050=":443"; ma=2592000,h3

```

Request Headers (621 B)
Raw

```

GET /embed/P-xlixF7B2U?autohide=1&fs=1&autoplay=0&rel=0&modestbranding=1&showinfo=0&hd=1&e
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_INFOF1_LIVE=dActKPajViQ; PREF=tz=America.Chicago&f4=4000000; YSC=p2jSvxCMeI
Upgrade-Insecure-Requests: 1
TE: Trailers

```


5 requests
229.65 KB / 172.86 KB transferred
Finish: 4.03 min
DOMContentLoaded: 451 ms
load: 1.70 s

HTTP Headers (uchicago.edu → youtube.com)

Inspector Console Debugger Style Editor Performance Memory Net							
Filter URLs							All
Status	Met...	Domain	File	Initiator	Type	Transferred	Size
200	GET	www.uchica...	/	document	html	11.41 KB	39...
204	POST	www.youtub...	atr?ns=yt&el=embedded&cpn=ho5PKBh-	base.js:1023 (...)	html	604 B	0 B
200	GET	www.youtub...	P-xlixF7B2U?autohide=1&fs=1&autoplay=	subdocument	html	21.81 KB	51...
200	GET	cdn.hypemar...	uchicagowww?width=1169&paginate=tru	a5b5e5.js:3 (s...	html	128.06 KB	12...
200	GET	cdn.hypemar...	popUpModalEndpoint	a5b5e5.js:3 (s...	html	10.99 KB	10...

HTTP Headers (uchicago.edu → youtube.com)

▼ Request Headers (621 B)

Raw 

```
GET /embed/P-xlixF7B2U?autohide=1&fs=1&autoplay=0&rel=0&modestbranding=1&showinfo=0&hd=1&e
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_INFO01_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)

Raw 

```
GET /embed/P-xlixF7B2U?autohide=1&fs=1&autoplay=0&rel=0&modestbranding=1&showinfo=0&hd=1&e
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_INFO01_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)

Raw 

```
GET /embed/P-xlixF7B2U?autohide=1&fs=1&autoplay=0&rel=0&modestbranding=1&showinfo=0&hd=1&e
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_INFO01_LIVE=dACtKPaJViQ; PREF=tz=America.Chicago&f4=4000000; YSC=p2jSvxCMel
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 “referrer” [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



Figure 1: Example of advertiser.com and tracker.com synchronizing their cookieIDs. Interestingly, and without having any code in website3, advertiser.com learns that: (i) cookieIDs userABC==user123 and (ii) userABC has just visited the given website. Finally, both domains can conduct server-to-server user data merges.

From Papadopoulos et al. "Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask," in *Proc. WWW*, 2019.

Track Visited Sites

- Subtle side channel!
- (Loophole has since mostly been closed)

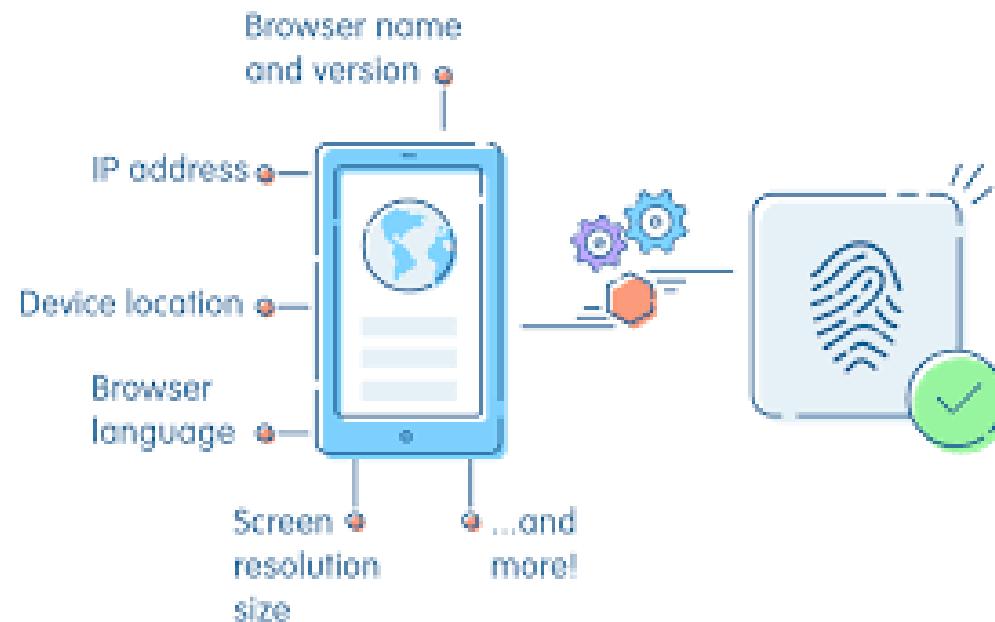
- link one
- second link
- link three (visited)
- fourth link

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/>

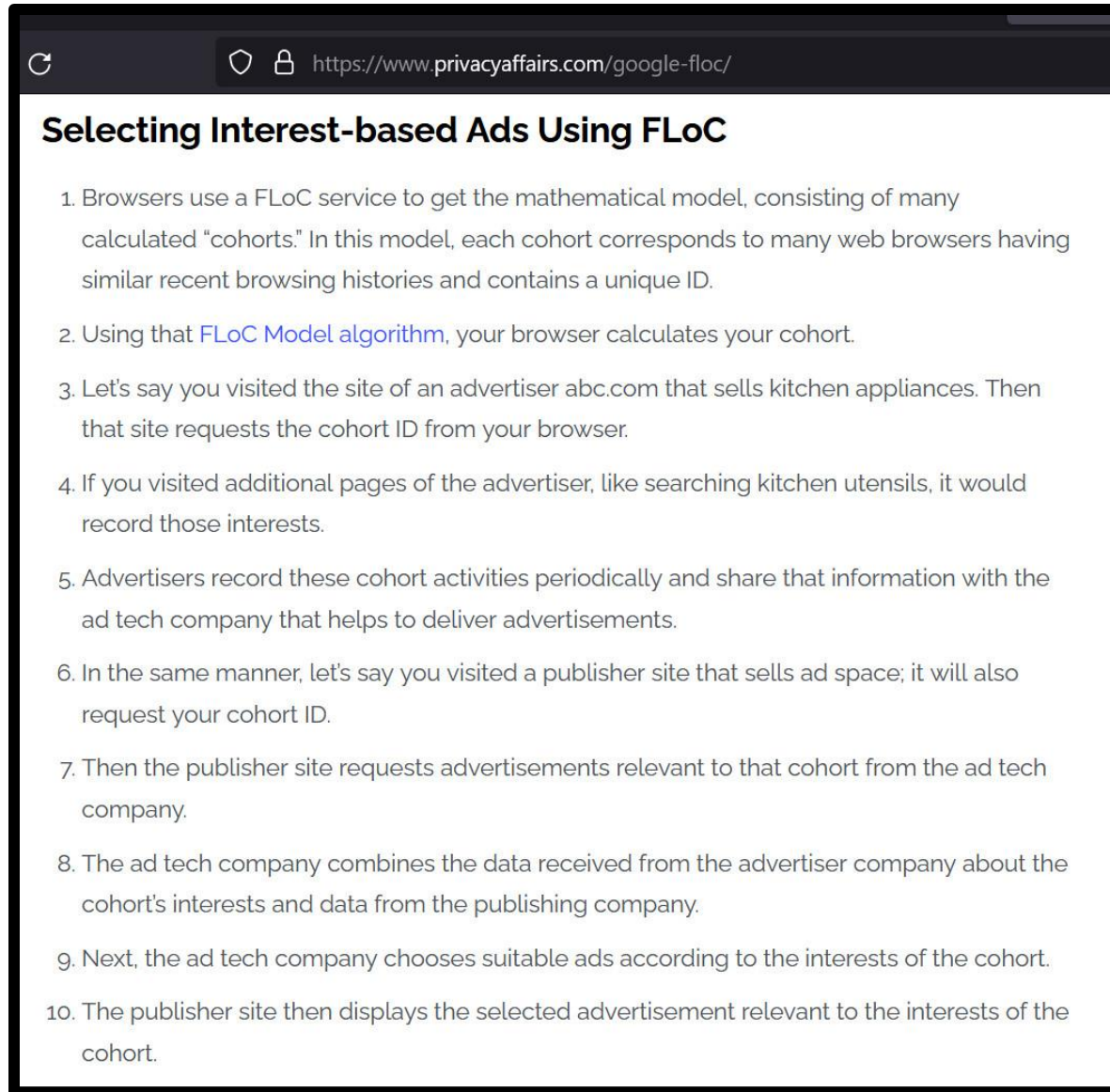


Alternatives to Cookies for Tracking / Profiling

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



The screenshot shows a web browser window with a dark theme. The address bar displays the URL <https://www.privacyaffairs.com/google-floc/>. The main content area has a white background and features the title "Selecting Interest-based Ads Using FLoC" in bold black text. Below the title is a numbered list of 10 steps explaining the FLoC process.

Selecting Interest-based Ads Using FLoC

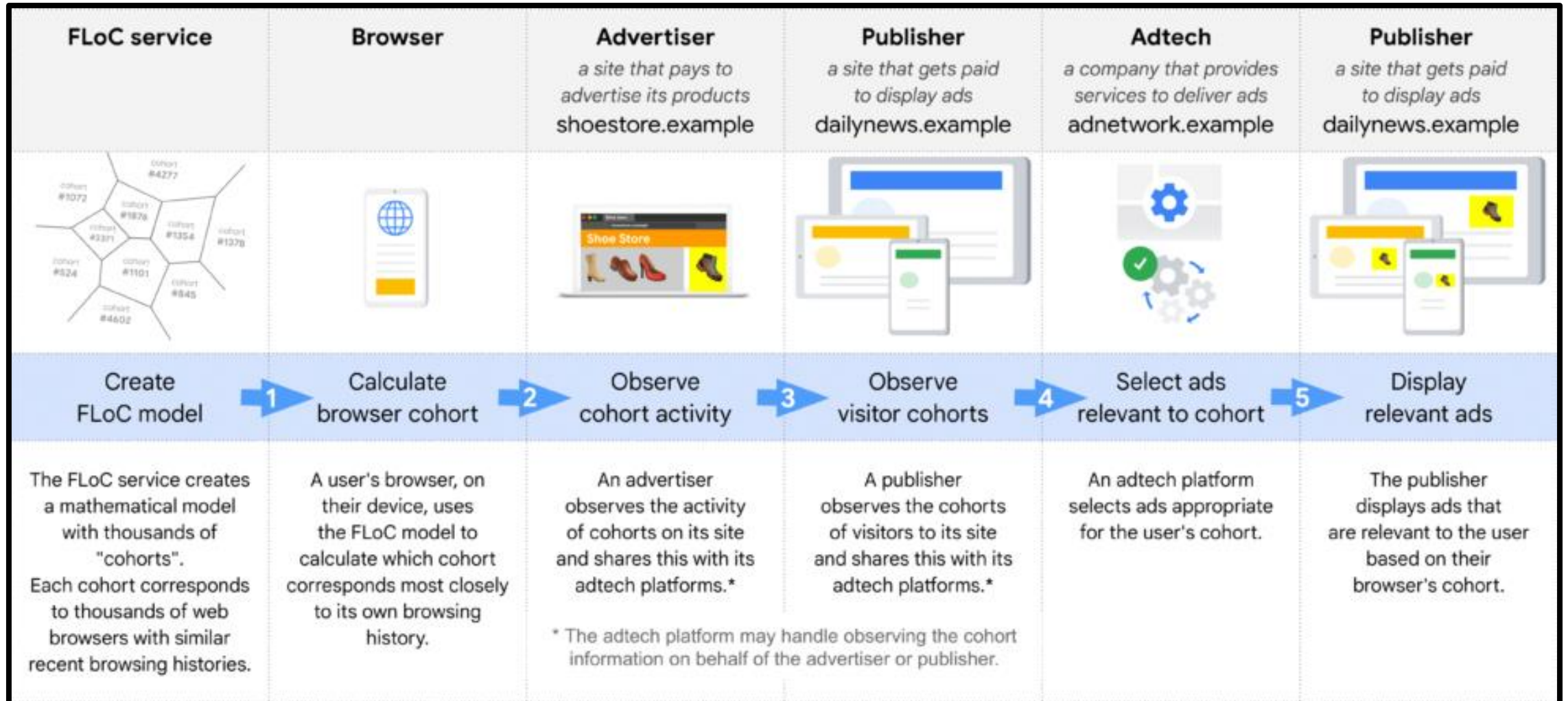
1. Browsers use a FLoC service to get the mathematical model, consisting of many calculated "cohorts." In this model, each cohort corresponds to many web browsers having similar recent browsing histories and contains a unique ID.
2. Using that [FLoC Model algorithm](#), your browser calculates your cohort.
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



Image taken from <https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea>

Google's FLoC



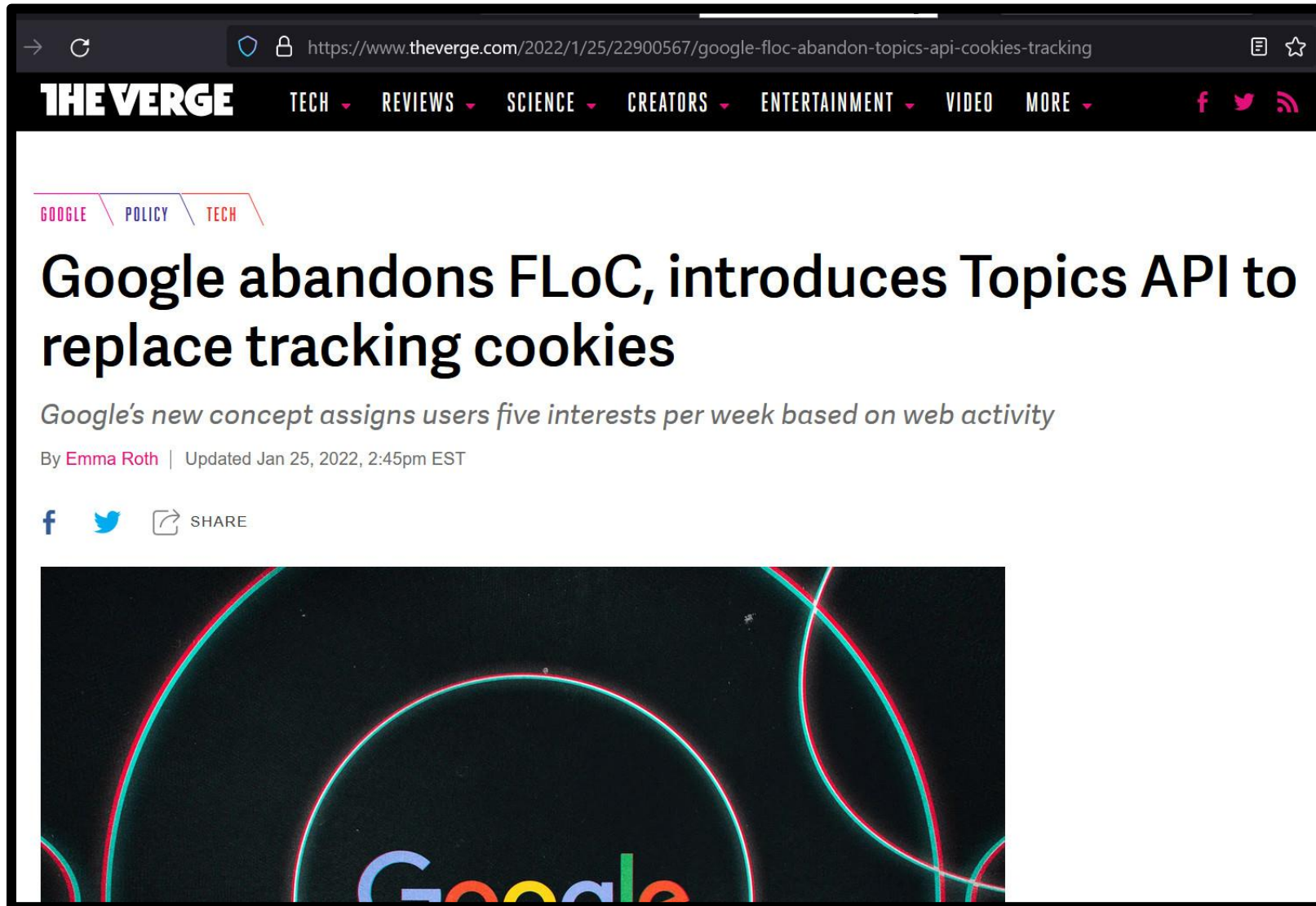
Google's FLoC

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.

Image taken from <https://www.privacyaffairs.com/google-floc/>

Google's Topics API



Google's Topics API

