Online Tracking, What Can Be Done About it, and Who's Doing it

Pete Snyder Senior Privacy Researcher, Director of Privacy pes@brave.com



Hi, I'm Pete 👋

- **Grew up in Chicago** ...actual Chicago
- Law school -> freelance web stuff Started: Anchorage, AK Ended: Judge Judy Show invitation
- University of Illinois at Chicago









Me at Brave

- Research at Braveprivacy, blocking, reliability
- Co-Chair of PING Privacy committee on W3C
- Research <-> Engineering
 Web compat, filter lists, etc.
- Academic <-> Industry Collaborations





Brave in a Slide

- Privacy focused
- Alternative web funding model Fix incentive problems
- Research + Engineering
- Not just a browser
 - search.brave.com
 - talk.brave.com
 - VPN
 - more coming...





Overview

- Why Privacy Matters
 A sloppy manifesto
- Defining Tracking
 Abstracting the problem
- Tracking in Practice
 Methods and defenses
- Privacy Beyond Tracking
 Other issues and concerns

5

Overview

- Why Privacy Matters
 A sloppy manifesto
- Defining Tracking
 Abstracting the problem
- Tracking in Practice
 Methods and defenses
- Privacy Beyond Tracking
 Other issues and concerns



6

Why Does Tracking Exist?



54 Sun Wednesday, April 29, 2015



Graeme Swann THE TALK OF SPORT V

2GM

0

IT WOULD BE A **ROAD TO RUIN** Waitrose

OUR COLUMNISTS ARE

Kid Dan's ton of fun TENAGER Das Lawrence by becoming one of county cricket's youngest worket who was a county worket was to be a co

will return

By JONNY FORDHAM LEE DICKSON believes English rugby can wrestle back control from teams across the Chanel. A French invasion takes place at Twickenham on Saturday as Clermont face Toulon in the European Champions Cap final. Dickson's Premiership leaders Northampton were KO'd by Clermont. But the England scrum-half, 30, said: "We will be back. "In the past we have beaten the blar Freech

beaten the big French teams — and Wasps nearly turned Toulon over."

breath of fresh air around the dressing room. To be captain, he would need to change and become more mature and perhaps even feign gravitas. That simply is not Joe. Of course, there is no vacancy at the moment any-way because Alastair Cook is doing a fine job new he is back concentrating on fest cricket. Cooky won't till be captain in five year' ime but please don't give he job to Joe. Just let Joe Root bat at Vo5 and score hundred after

and score hundred after dred. He can be fand's Steve Waugh. So

<section-header><section-header><section-header><section-header><section-header><text><text><text><text><text><text><text><text><text><text><text><text><text><text><text><text><text><text><text><text>



Grenada bu his fault - th he barbecue right to the air from the

Holder did and mmy was way ground. Holder's brother all in the Statest game. Rooty r a stump to ending it was a

Fulham Parker, Well, Those F

13.855 655

A transfer policy may have to take a bit after chairman Tony Bloom's bet of misforhase. The Seagath supremo placed £13,000 on himself to finish the Brighton Marathon in no more than 3hr 45 mis — only to cross the line four minutes and 38 seconds later.

VIEW TO A DILL **Cough not**

that Good

EVERTON is known as the school of science but deamatics too. Reporters were gathered to interview Gareet Barry when the subject of team-make Ross Barkley instriably surfaced seeing as he is the OMU' inter-esting topic of conversa-tion at Goodison Park. As soon as Barkley's name was mentioned, the club's subtle press officer came out with an outra-groundy staged cough -as in 'be careful here make trickly subject coming up'. Bath corefully man-maty hacks backfired when the entire group fell silent and he was presented with a packet

when the entire group fell silent and he was presented with a packet of throat lozenges at the next press conference.



A 'Parker' can cut it

SOMETHING for the weekend, sir? Traditional barber's shops would always offer extras to geniteman patrons to keep them safe whill

captain

getting a s**g than seeing their team win on Satur-



26M

From ANTONY KASTRINANOS JURGEN KLOPP wrocked Pep Guardiola's Troble bid as ten-man Borussia Dortmund beat Bayern Munich on penaltics in the German Cup semi-finals. Dortmund won the shootout 2-0 at the Alliasz Arena, with the game end-ing 1-1 after extra-time. Robert Lewandowshi's opener against his old club was cancelled out of Pierre-Emerick Automos-ang's 75th-size ieveller for Dorthest. Out and Dortmend boss Klose wanted by a bost of penier League cubs, saw even Kampi sent off for a second booking late on.

Pep gets Klopped Rem ANTONY KASTRINAKS

money, McMe

Wednesday, April 29, 2015 Sun 55



BEEN READING — RECLAIM YOUR MANLINESS— SAUSAGE & BACON SANDWICH

amilleri. The official should have called a ul and awarded Bingham seven sints, but the Rocket was allowed continue his break and won the th frame to lead 3-2. th frame to lead 3-2. Rule 3 (d) ii states a seven-point malty will be awarded if a player ises any object to measure gaps distance".

he ball. exactly what Ronnie has art Bingham was looking d should really be aware. had more knowledge of is pay-p

world champion Ken said: "You are not allowed halk on the table because using it as a tool to



to ref in." 19-year-old O'Sullivan i

is tournment. uped a fine after briefly in his socks during his d win over Craig Stead-came close to snapping in fit of rage against beyens in round two. tently O'Sullivan. 30, has une to boxing and even tring.

s box-office star is the riches earned by tg. He said: "Boxing and if snooker was

After his best season that SHAUN MURPHY leads Anthone has seen titles in Australia and McGill 9-7 after winning the on a roll lunchdow 25 more of last







Have you ever clicked your mouse right HERE? ----

Welcome the The "First" Banner Ad

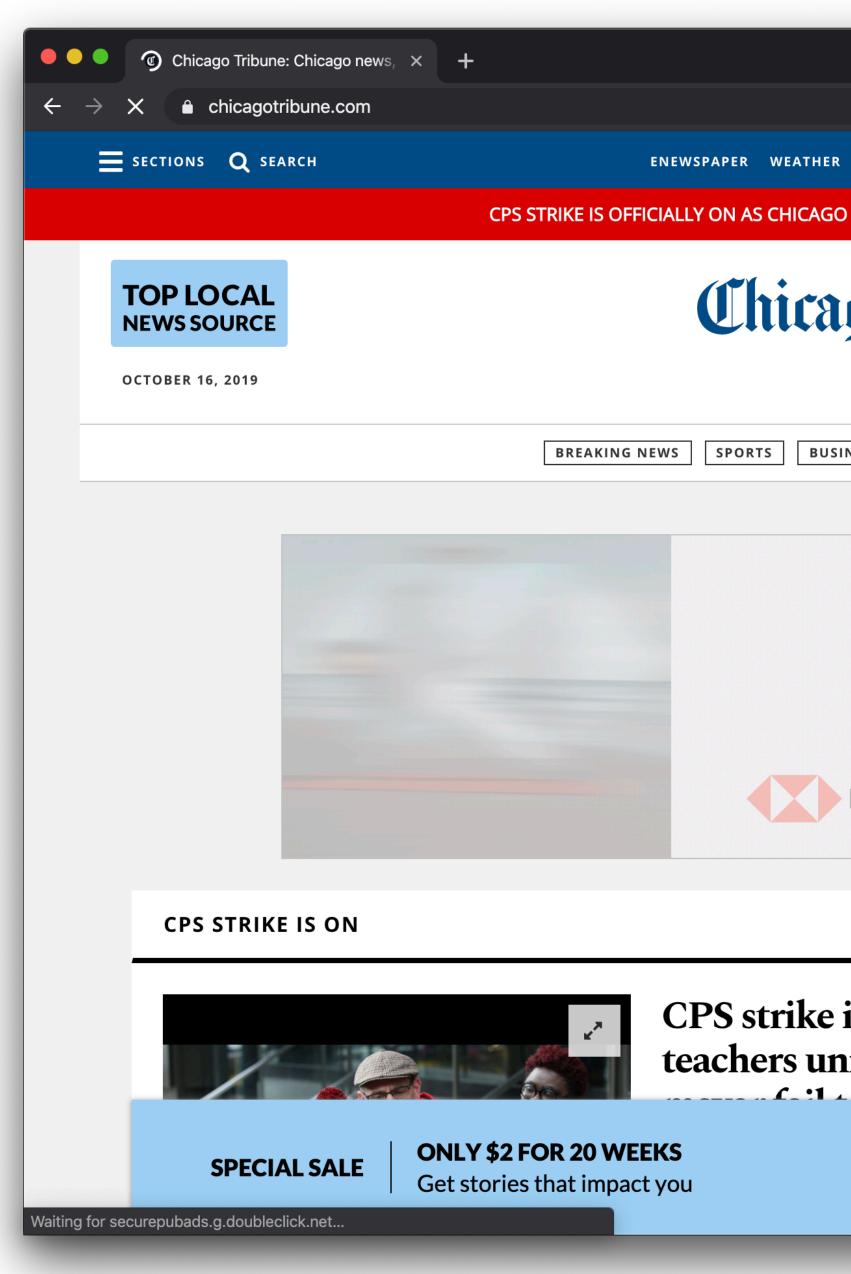
Yes, this site is supposed to look this way. After all, this is what most web pages looked like back on October 27, 1994 -- the day that Wired Magazine flipped the switch on its first website, hotwired.com, starting a revolution in web content and advertising that still reverberates today.

This site is dedicated to showing off one of the ads that ran on that site. No, it wasn't the "first" as there were a handful of other ads that ran on various sections of hotwired.com. This site is also here to tell the story of how that ad came to be, how it succeeded beyond anything we had imagined, and how we tried to set an example for how corporations could communicate with their audiences.

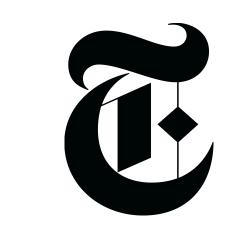
This site launched on October 27, 2014. It is being constantly updated, so please check back again soon for more. In the meantime, get started by clicking your mouse in the banner ad above explore these other options:

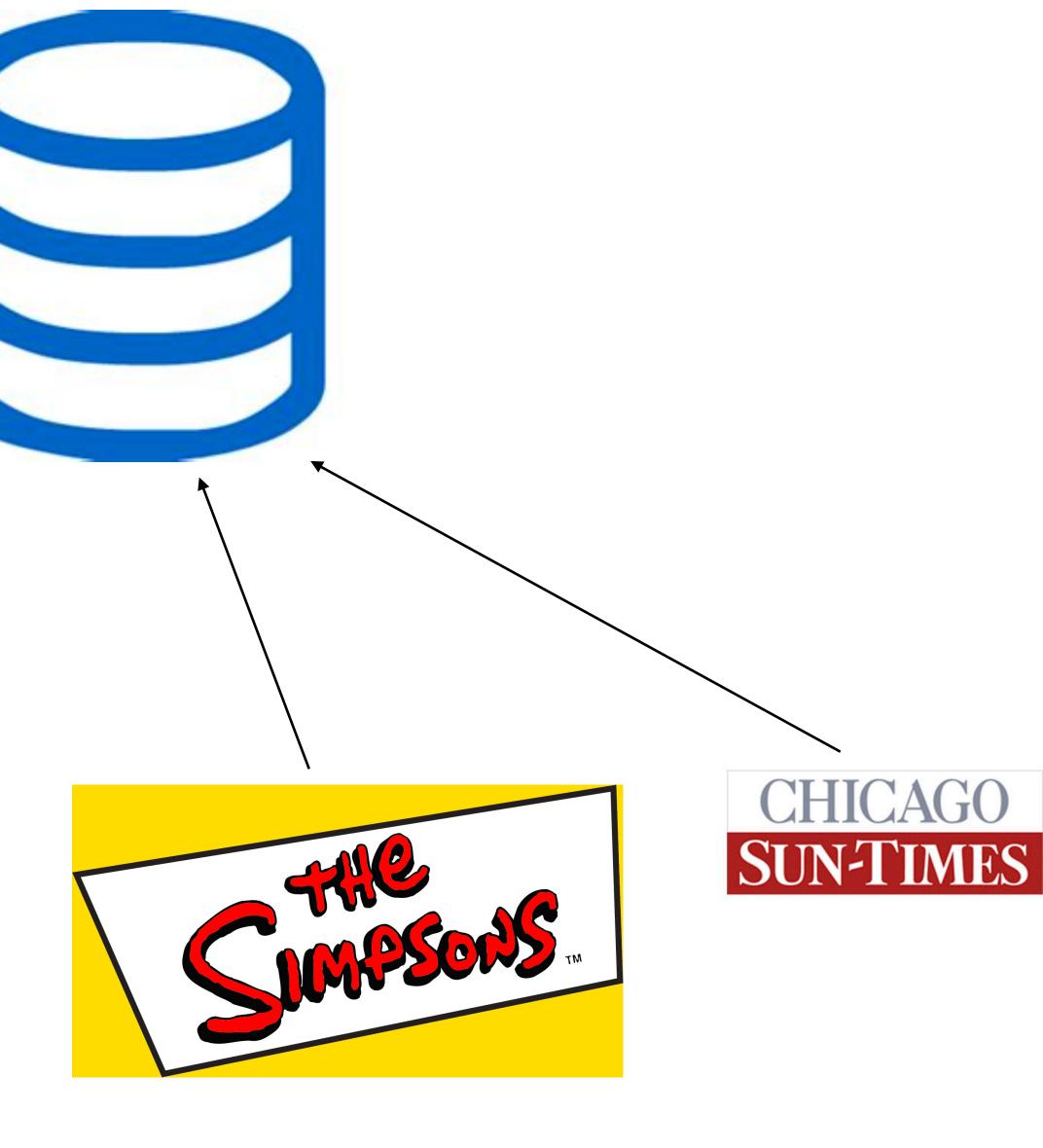






		☆ ▣ €
NEWSLETTERS BEST REVIEWS	\$2 FOR 20 WEEKS SALE ENDS 11/4	➡) LOG IN
EACHERS UNION SAYS THERE IS NO) LAST-MINUTE DEAL	
10 Tribune	20	2 FOR WEEKS DS 11/4
	ජ්	51°F
SS POLITICS OPINION ENTE	ERTAINMENT	
DVERTISEMENT		
BC		
BC		
BC		
SBC	MORE CPS STRIKE COVERAGE	>
officially on as	MORE CPS STRIKE COVERAGE CPS strike live updates: Chicago teache city offer, will walk off job Thursday	
officially on as on, Chicago	CPS strike live updates: Chicago teache city offer, will walk off job Thursday	rs reject
SBC	CPS strike live updates: Chicago teache	rs reject htract
officially on as on, Chicago	CPS strike live updates: Chicago teache city offer, will walk off job Thursday	rs reject





The World's Worst Website Angelfire Build your own FREE website at Angelfire.com Share: del.icio.us | digg | reddit | Twitter | facebook MARKETING IS MORE FUN WHEN YOU SMASH STUFF VIEW THE WORK

Gratuitous use of frames is a common mistake of web designers.

Many older browsers do not support frames. They disrupt the flow of the website and can be difficult to anticipate where a page may appear when a link is clicked. Click here for an example of a frames page which is opening in the wrong window. Use your browser's 'Back' button to escape.

Check out these links to websites whose opinions about frames is self evident:

The "I Hate Frames" Frames Page

Another I Hate Frames Page

The International I Hate Frames Club

Why Frames Suck (Most of the Time)



0



Keep your backgrounds simple. White or light colors usually work best. Your background should not compete with the content of the page for the users attention. If you would like to use a background picture, select a picture that uses muted colors or format your picture as a watermark. Select text colors which will contrast well with the background picture.



me to the World's Wor

This web was designed to graphically demonstrate the most common mistakes made by new Web Page designers.

Where am I and where are the links to other pages?

An easy to use navigation structure is essential to any well designed website! Important information should never be more than 2 clicks away.

As you can see, this text is difficult to read. There needs to more contrast between the background color and the text color. Here's another example of a oor choice of a background/ text color and size.

> Constantly running animations can be distracting when used excessively.





\$\$



¢¢¢



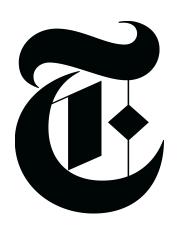


→

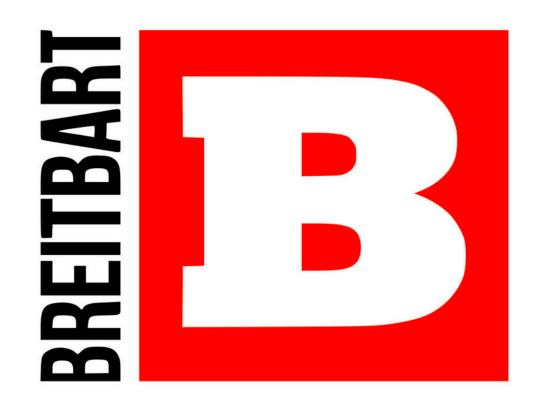
The World's Worst Website





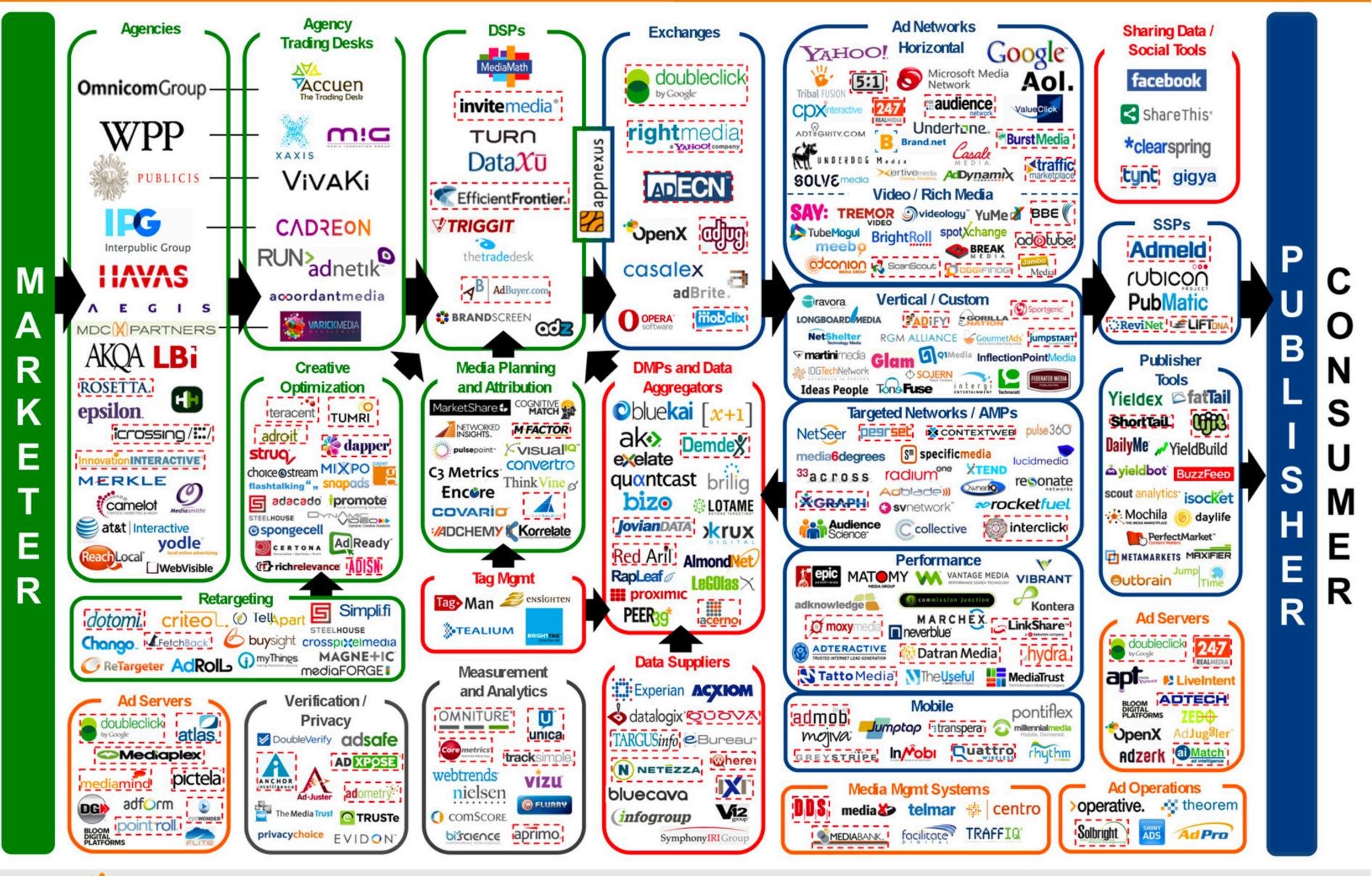


Identify "expensive" people here



Pay a little to advertise to them here

DISPLAY LUMAscape





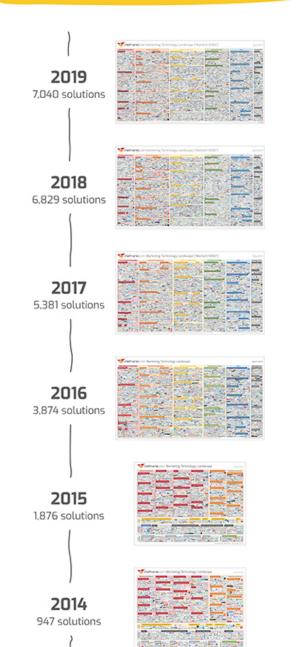


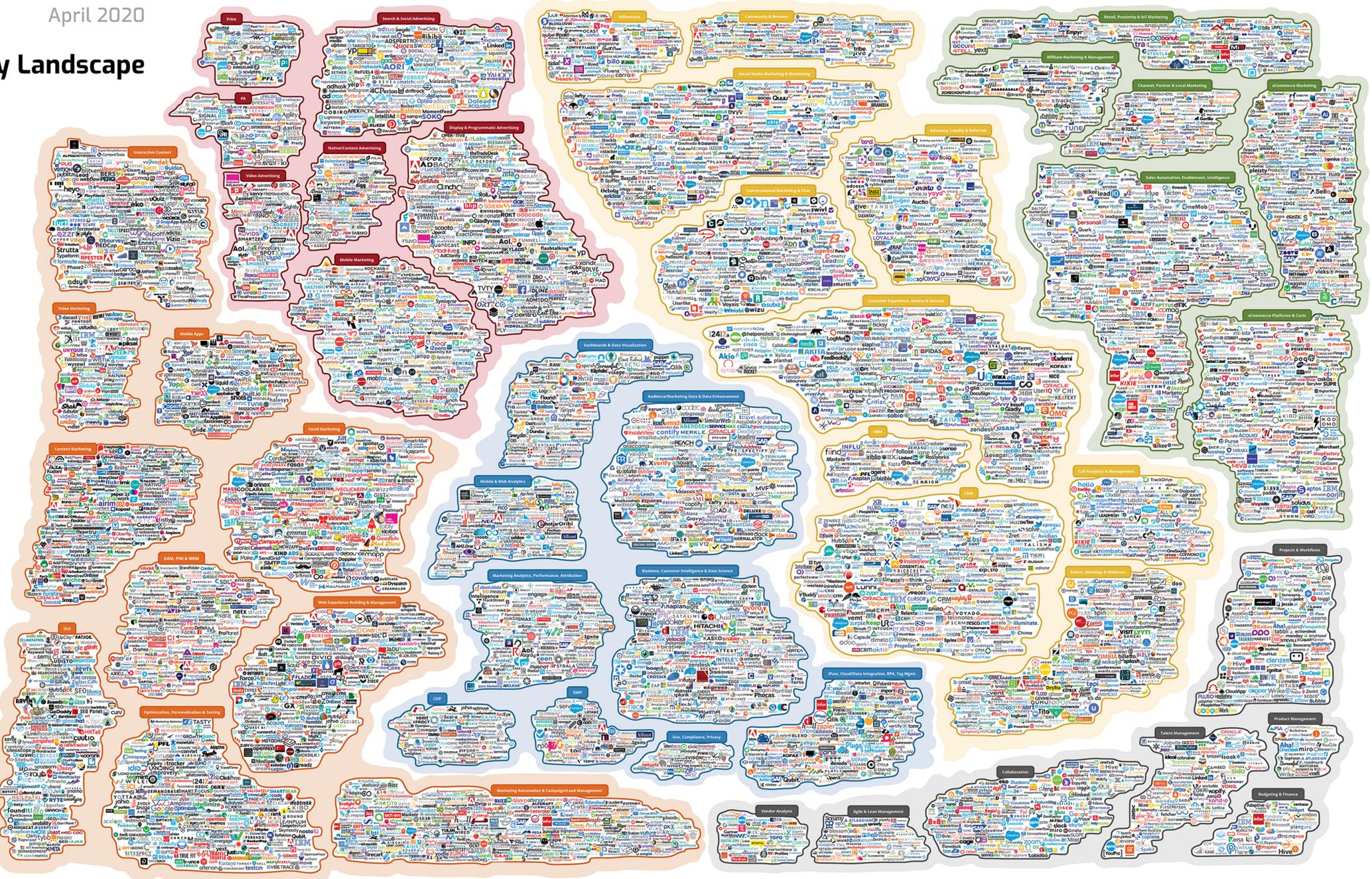
Marketing Technology Landscape

The Martech 5000

Total Solutions	8,000
Advertising & Promotion	922
Content & Experience	1,936
Social & Relationships	1,969
Commerce & Sales	1,314
Data	1,258
Management	601

Access all the data of this landscape & more at **martech5000.com**





Copyright © 2020 Marketing Technology Media, LLC. See https://chiefmartec.com/2020/04/marketing-technology-landscape-2020-martech-5000/ for details and sources.

Summarizing: Why Does Tracking Matter

- Incompatible with dignity
- Power and control
- Transfers wealth from value-creators to attention-attractors



Overview

- Why Privacy Matters
 A sloppy manifesto
- Defining Tracking
 Abstracting the problem
- Tracking in Practice
 Methods and defenses
- Privacy Beyond Tracking
 Other issues and concerns



Definitions

- Website: eTLD+1 (determined by public suffix list)
 e.g., brave.com != mozilla.org
 e.g., talk.brave.com == search.brave.com
 e.g., ted.github.io != betty.github.io
- Origin: The full DNS host name serving a site
- First-party: Site of the top level document
- Third-party: any other site



Definitions (more)

- **DOM Storage:** Explicit storage APIs e.g., cookies, localStorage, IndexDB
- **Network State:** All other storage e.g., caches (v8, DNS, HTTP) e.g., Header instructions (HSTS, ALT-SRV, etc)
- **Online Tracking** Its trickier...



A Rough Definition of Tracking

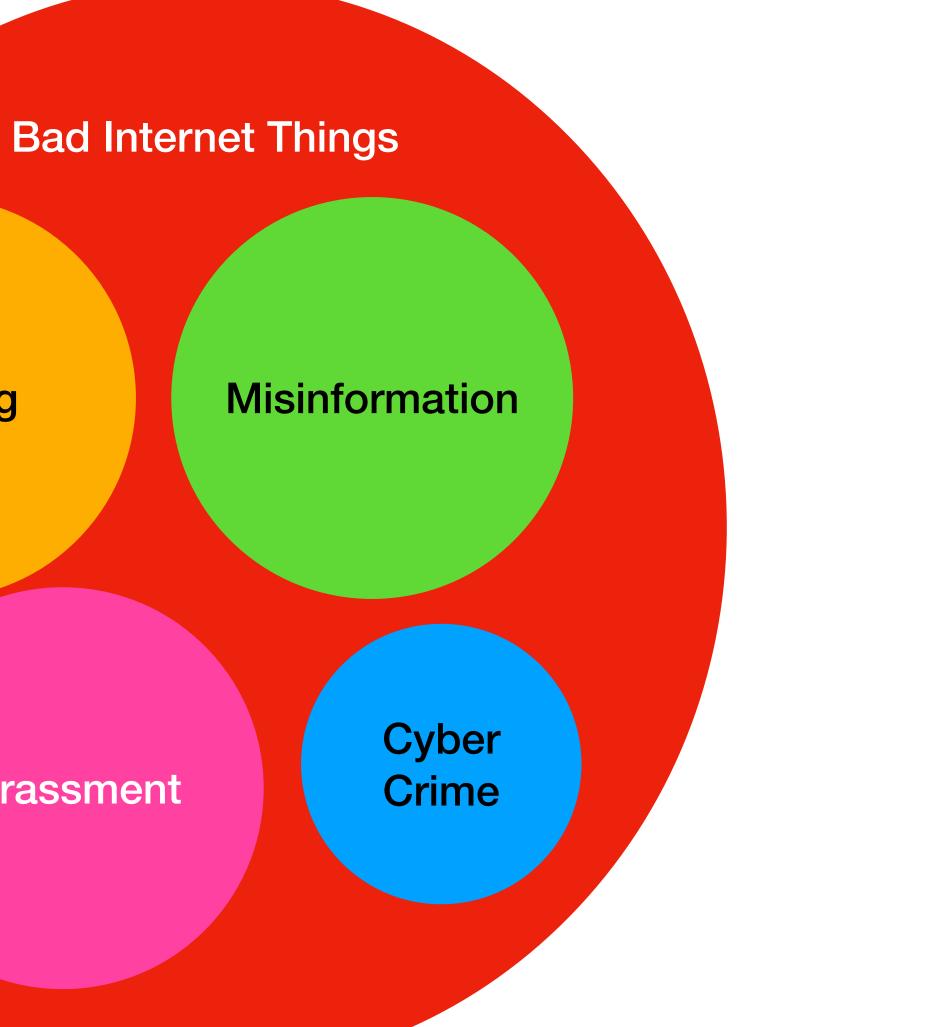
- Linking activities... e.g., being "followed"
- across boundaries... e.g., temporal, geographic, conceptual
- In a way not expected or desired. e.g., ignorance or non-consent



Tracking in Context

Tracking

Hurassment

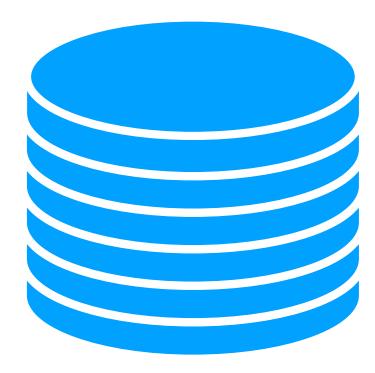


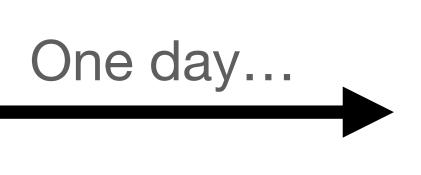
A Rough Definition of Tracking

- Linking activities... e.g., being "followed"
- across boundaries... e.g., temporal, geographic, conceptual
- In a way not expected or desired. e.g., ignorance or non-consent



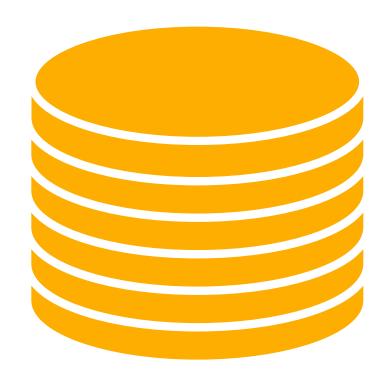
Question One



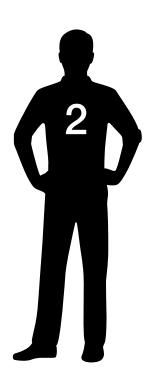


some-site.example



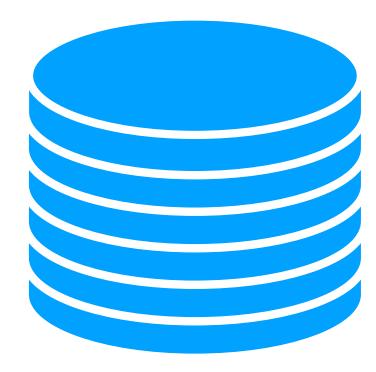


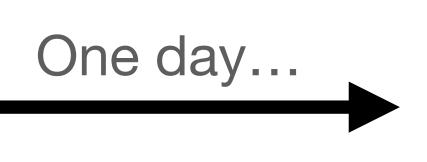
other-site.example



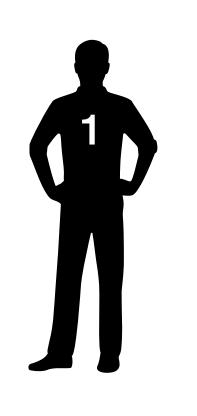


Question One





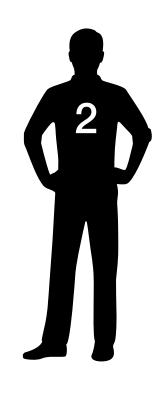
some-site.example





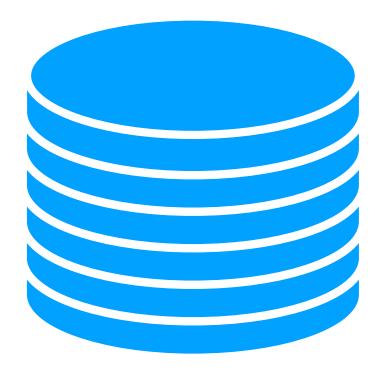
other-site.example

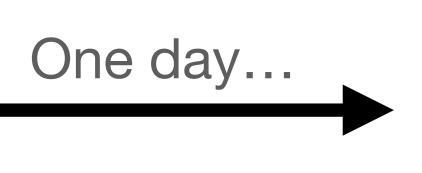






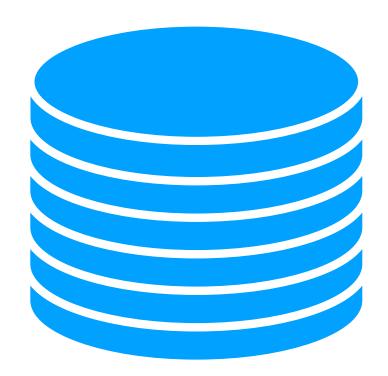
Question Two

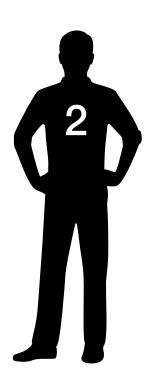




some-site.example

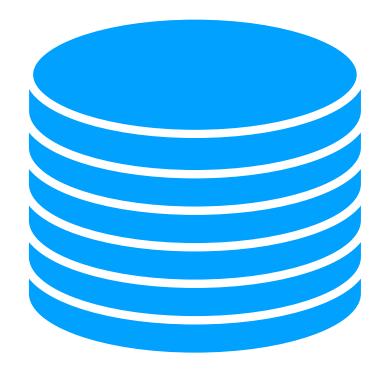


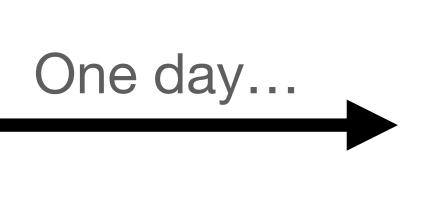




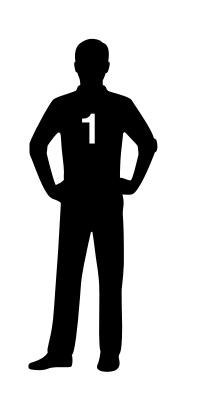


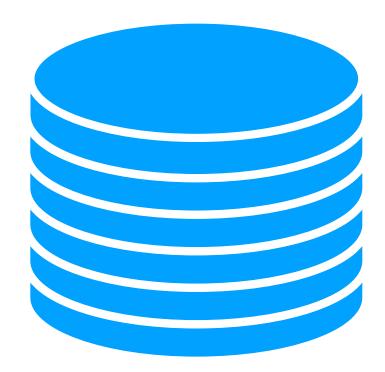
Question Two



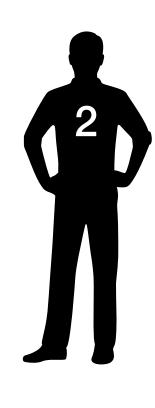


some-site.example



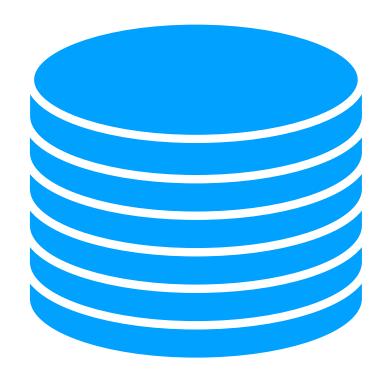


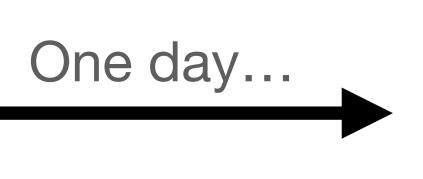






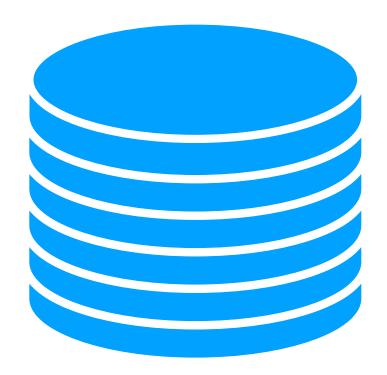
Question Three

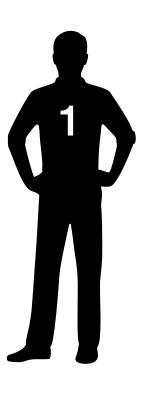




some-site.example

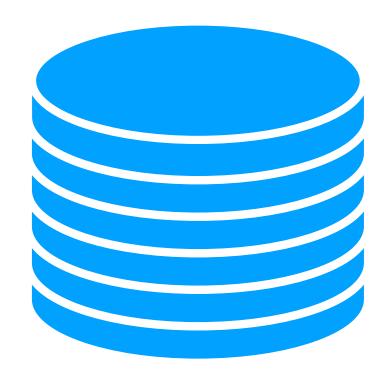






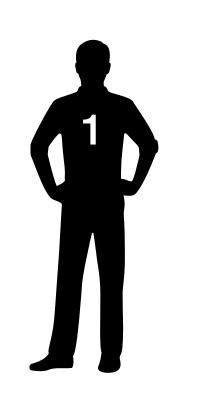


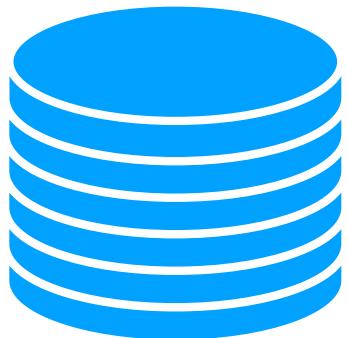
Question Three



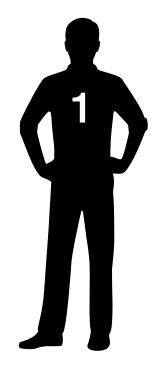


some-site.example







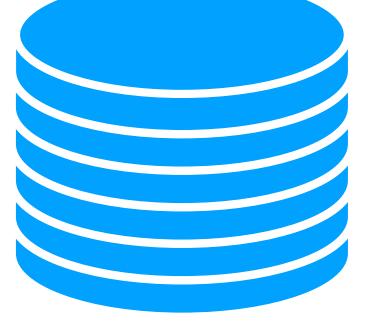




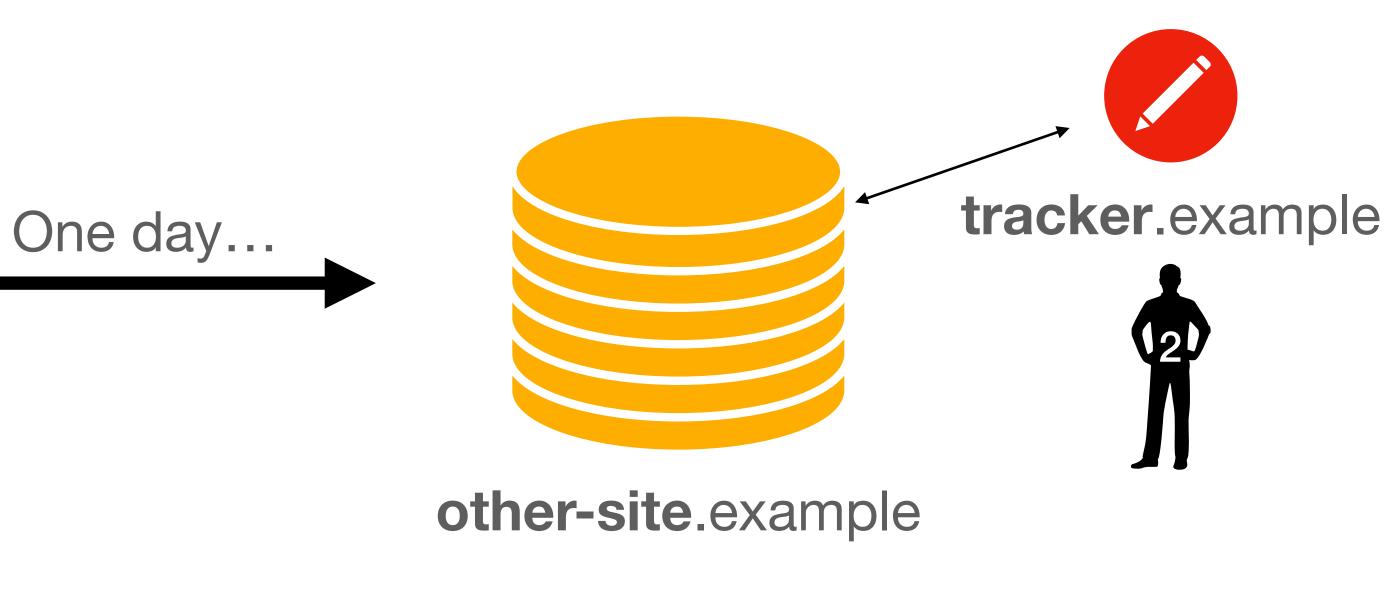
Question Four

tracker.example

Ð









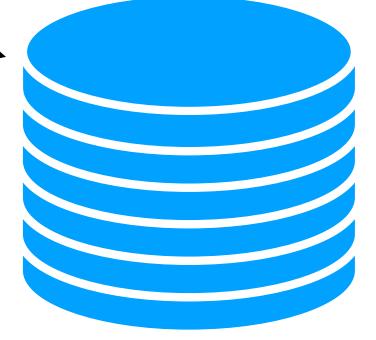




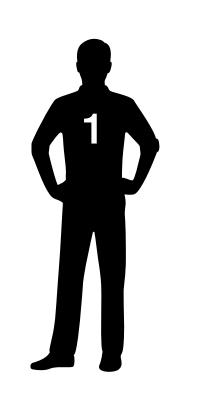
Question Four

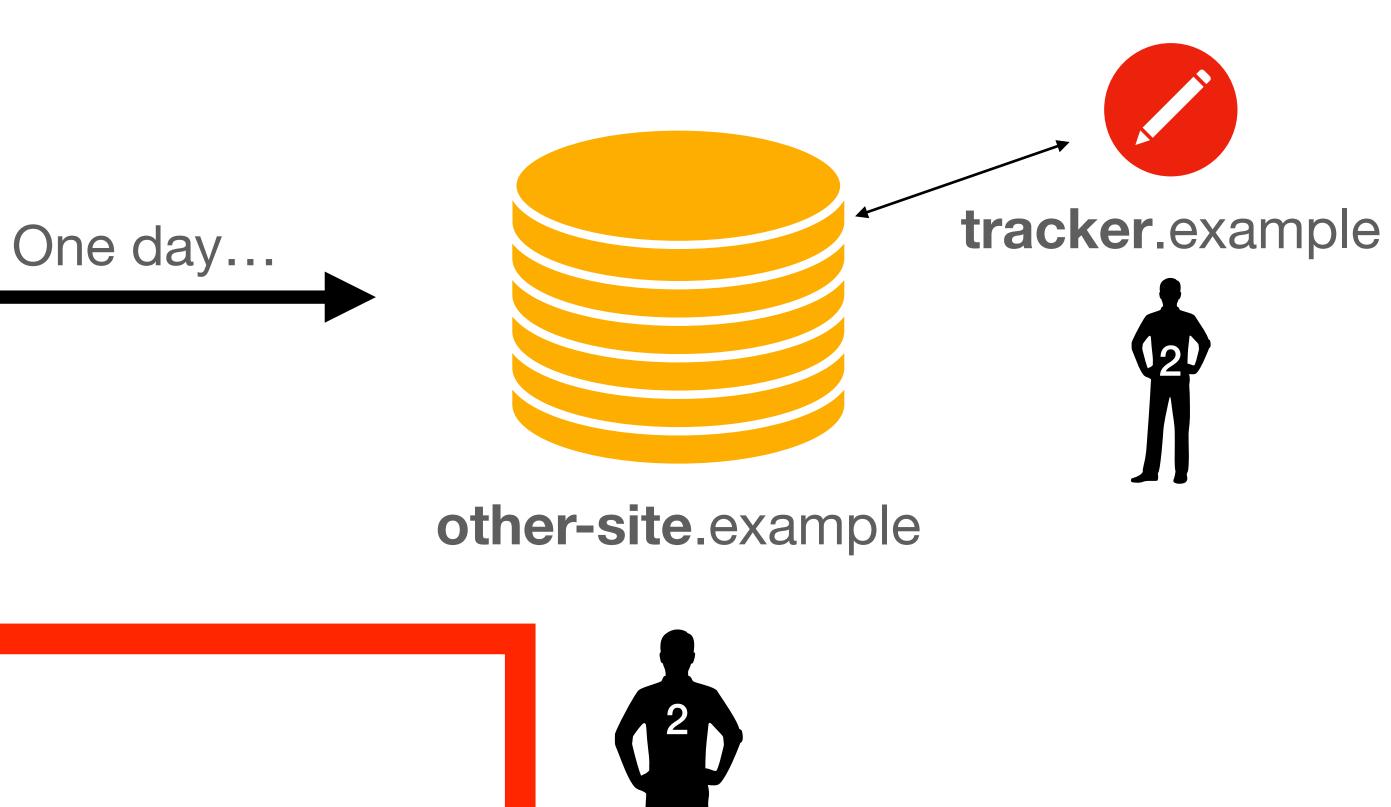
tracker.example

Ð



some-site.example





Not linking



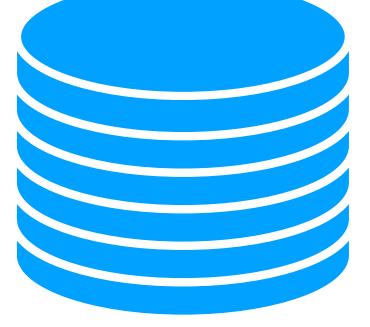


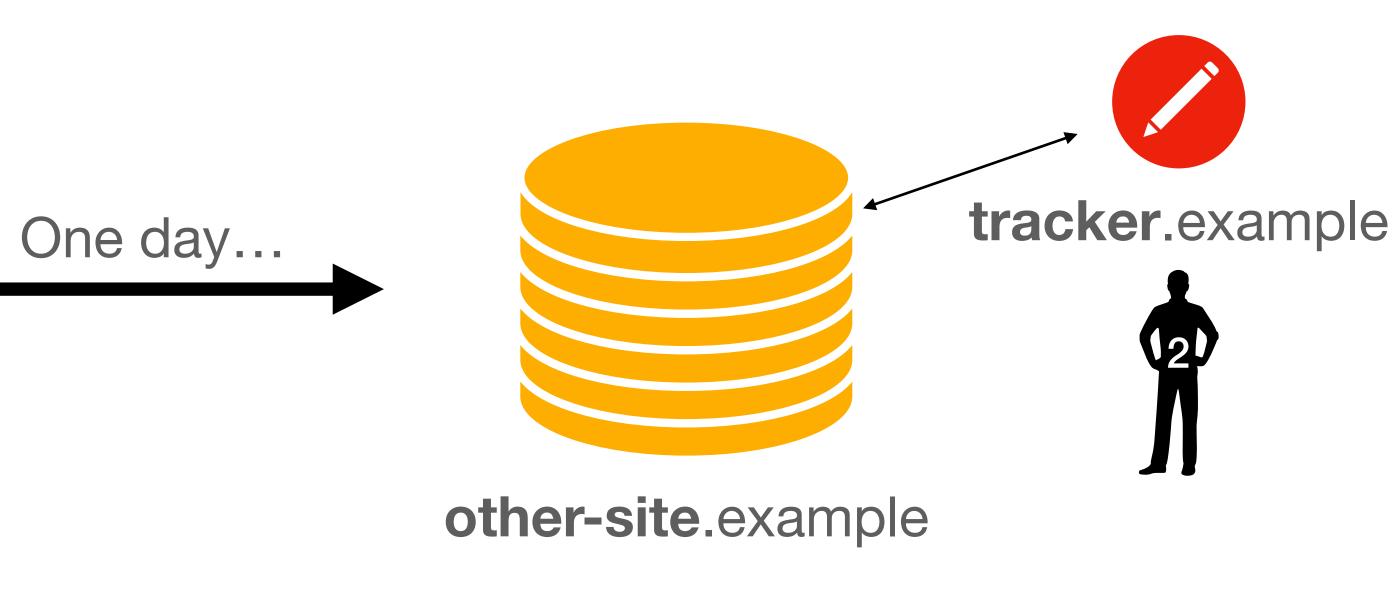


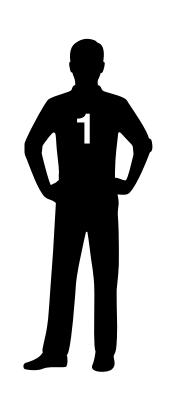
Question Five

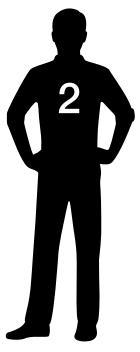
tracker.example

Ð









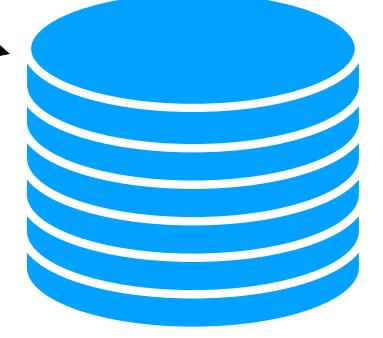


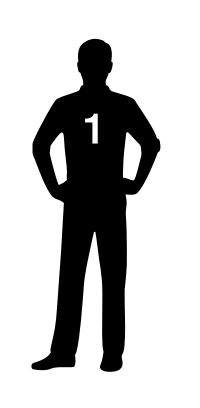


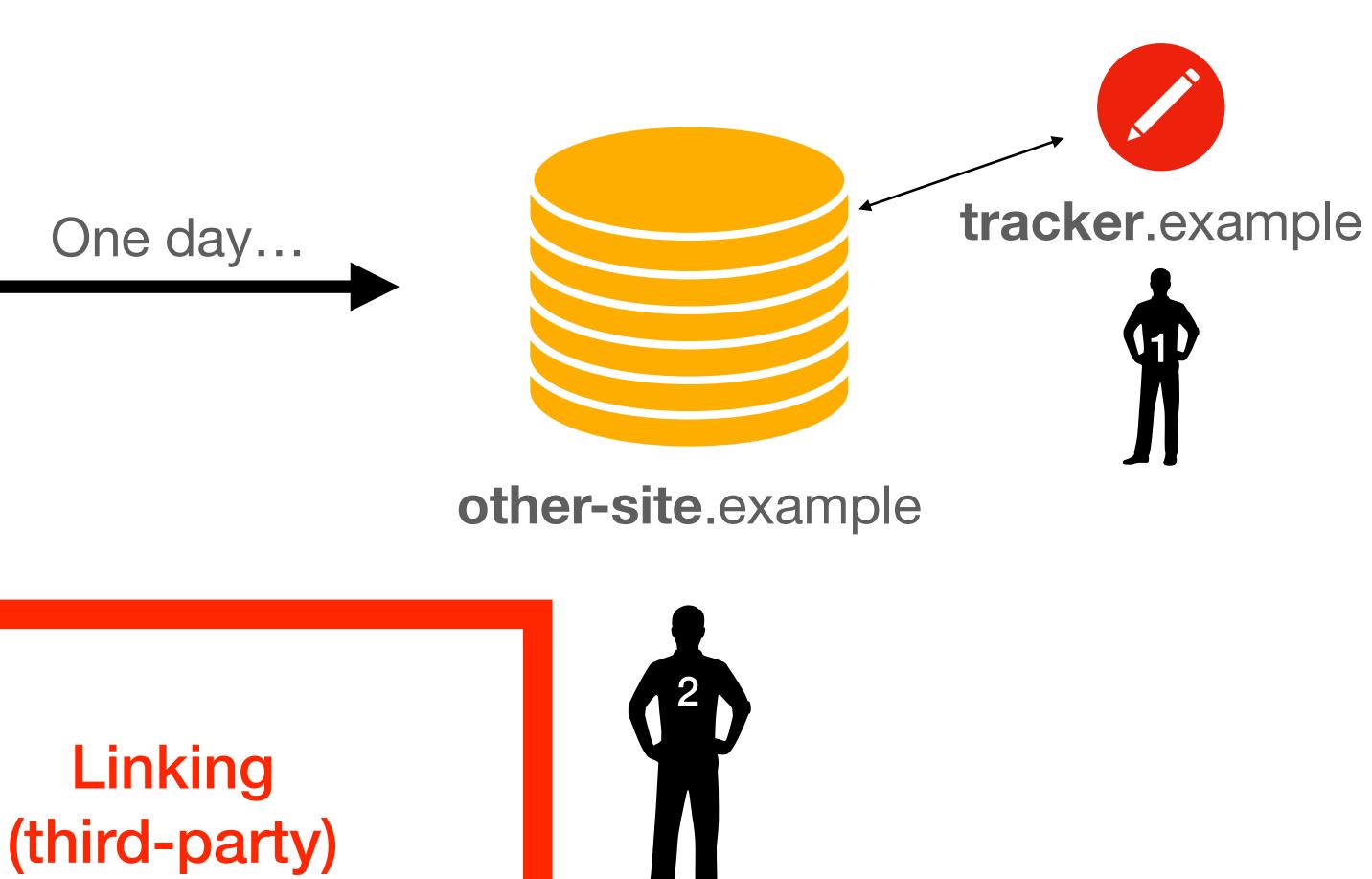
Question Five

tracker.example

Ð









Tracking: Linking...

- Tying behaviors to same identity Could be pseudonymous, or a "real world" identity
- **Probabilistic or deterministic** For some definition of "probable enough"



A Rough Definition of Tracking

- Linking activities... e.g., being "followed"
- across boundaries... e.g., temporal, geographic, conceptual
- In a way not expected or desired. e.g., ignorance or non-consent



Tracking: ...across boundaries...

- **Organizational boundaries** e.g., eTLD+1, origin, "first-party set"
- **Temporal boundaries** e.g., tying something done last year to something done today
- **Profile boundaries** e.g., private browsing, different browsers, accounts



A Rough Definition of Tracking

- Linking activities... e.g., being "followed"
- across boundaries... e.g., temporal, geographic, conceptual
- In a way not expected or desired. e.g., ignorance or non-consent



Tracking: ...expectations

- **Expectations differ across platforms** Facebook inapp browser vs Tor Browser Bundle
- **Expectations differ across people's expertise** e.g., my dad vs **Dworkin**
- **Expectations differ across backgrounds** e.g., outlook.com vs microsoft.com vs github.com
- **Consent is (sometimes) fuzzy** Terms of service < - - - - - - - - - - - Storage Access API



A Rough Definition of Tracking

- Linking activities... e.g., being "followed"
- across boundaries... e.g., temporal, geographic, conceptual
- In a way not expected or desired. e.g., ignorance or non-consent



Overview

- Why Privacy Matters
 A sloppy manifesto
- Defining Tracking
 Abstracting the problem
- Tracking in Practice
 Methods and defenses
- Privacy Beyond Tracking
 Other issues and concerns



Tracking Techniques

- Third-party DOM storage
- Network state
- Bounce tracking
- Browser fingerprinting
- IP address
- Personal identifiers



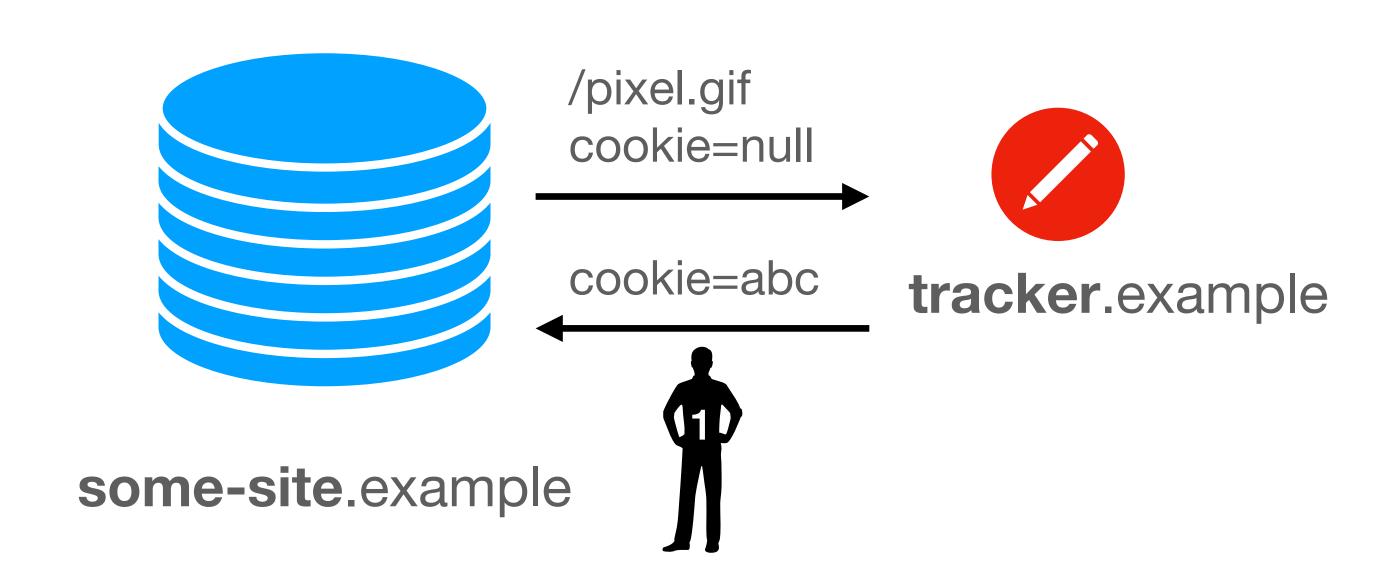
Tracking Techniques

- Third-party DOM storage

- Network state
- Bounce tracking
- Browser fingerprinting
- IP address
- Personal identifiers



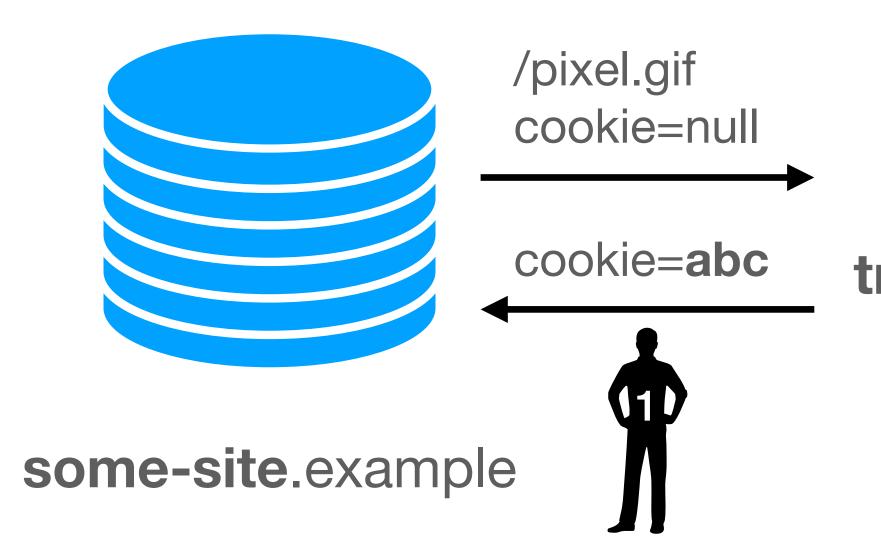
Third-party DOM storage







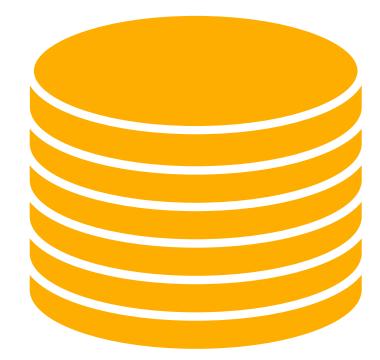
Third-party DOM storage: cookies





/pixel.gif cookie=**abc**

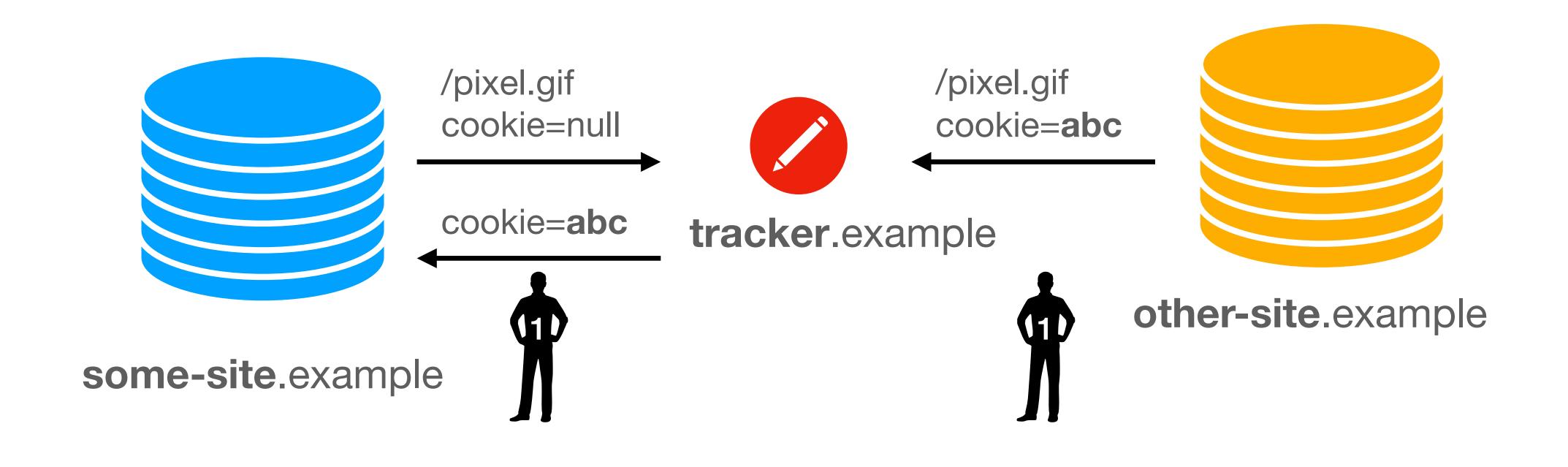
tracker.example



other-site.example

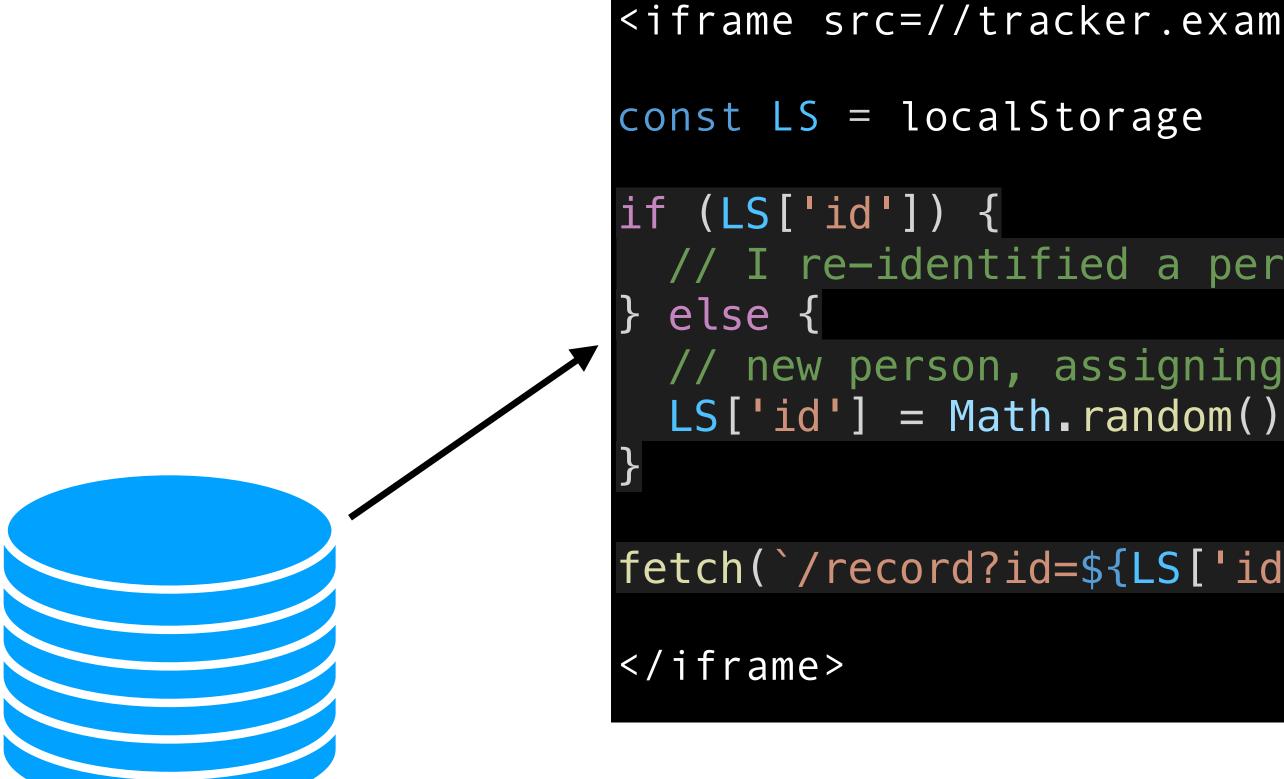


Third-party DOM storage: cookies





Third-party DOM storage: iframe



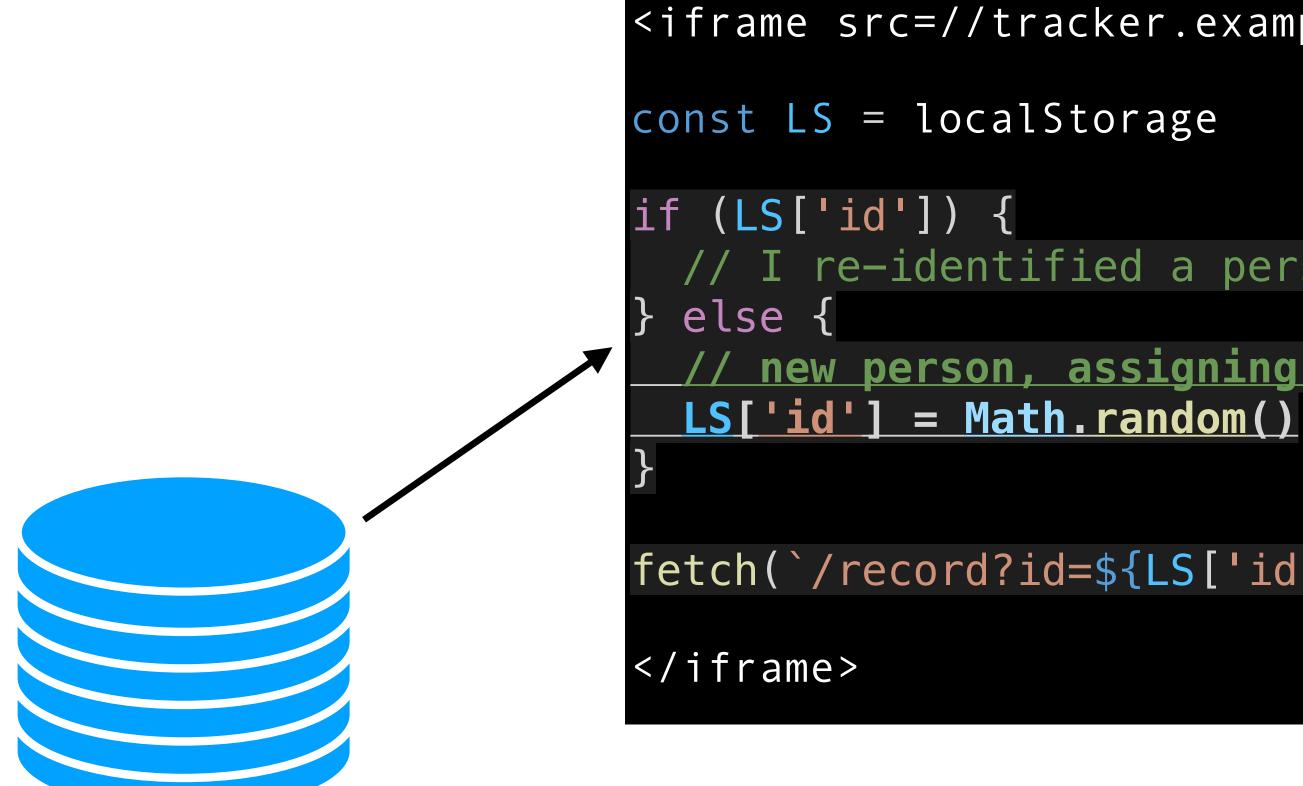
some-site.example

- <iframe src=//tracker.example>

 - // I re-identified a person
 - // new person, assigning ID
- fetch(`/record?id=\${LS['id']}`)



Third-party DOM storage: iframe



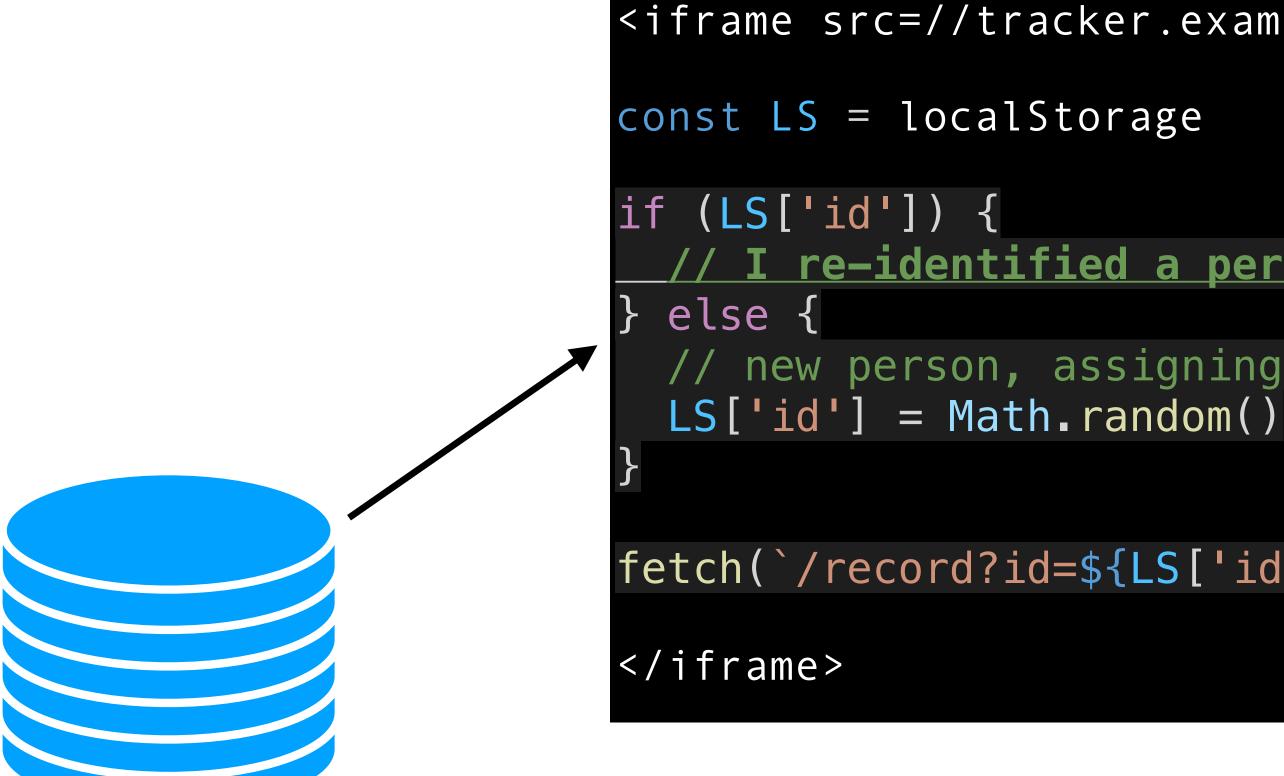
some-site.example

- <iframe src=//tracker.example>

 - // I re-identified a person
 - <u>// new person, assigning ID</u>
- fetch(`/record?id=\${LS['id']}`)



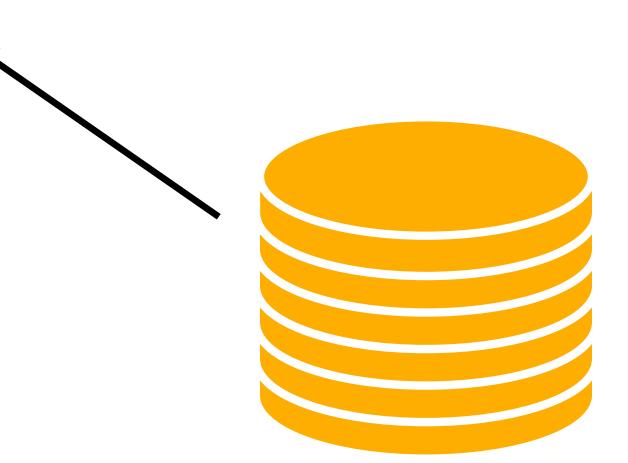
Third-party DOM storage: iframe



some-site.example

- <iframe src=//tracker.example>

 - // I re-identified a person
 - // new person, assigning ID
- fetch(`/record?id=\${LS['id']}`)



other-site.example



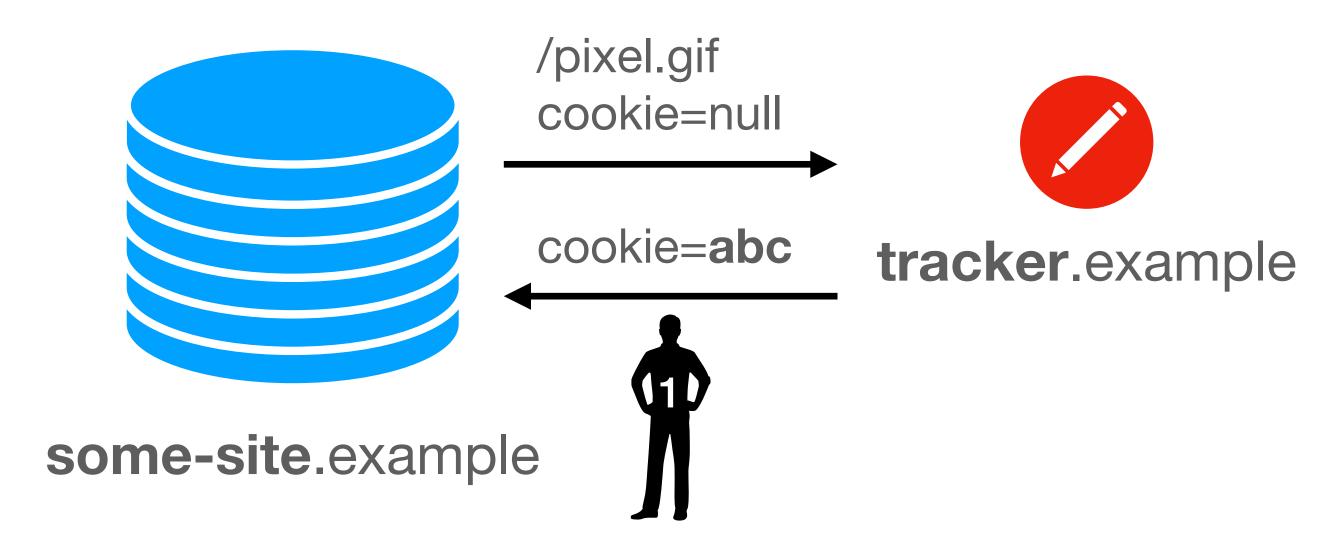
Third-party DOM storage: partitioning

- Third-party storage is not shared across sites
- Sometimes called "dual-keying"
- **Previous:** storage_data = browser_storage[<requested eTLD+1>]
- **Partitioning:**

storage_data = browser_storage[<first-party eTLD+1>][<requested eTLD+1>]

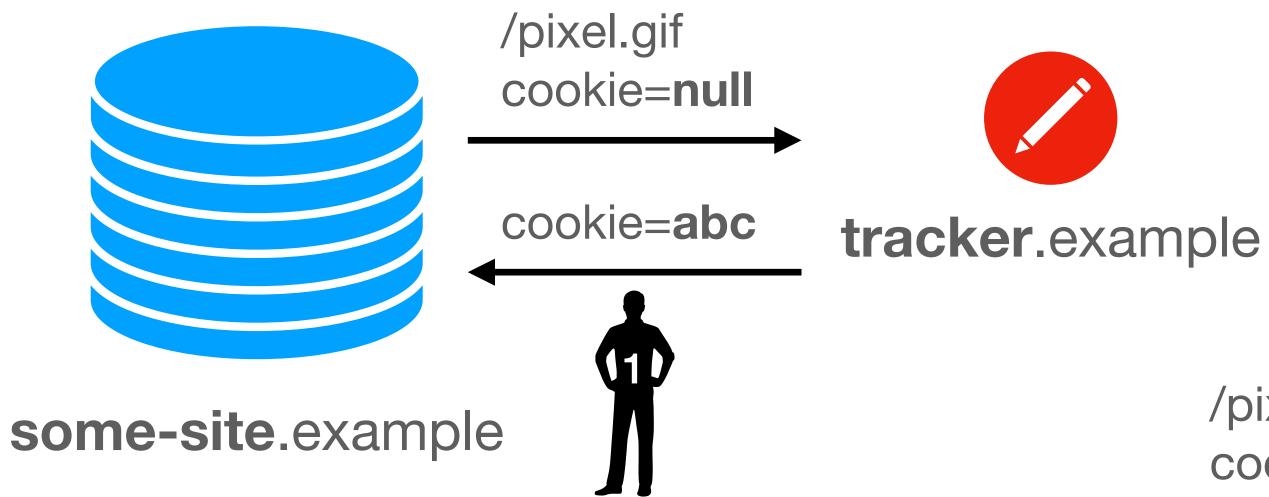


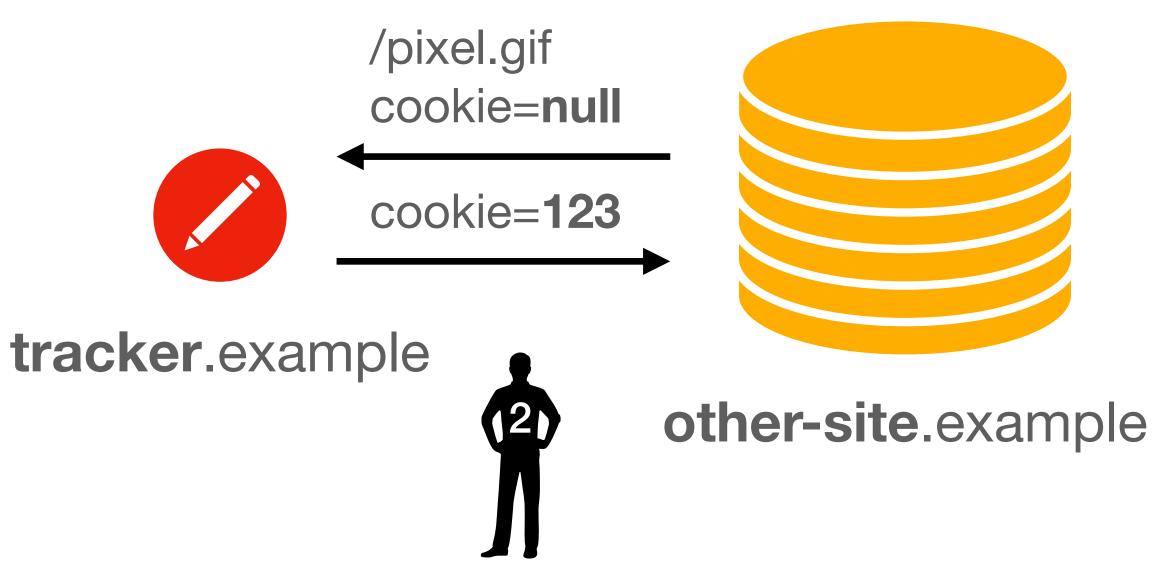
Third-party DOM storage: partitioning





Third-party DOM storage: partitioning







Third-party DOM storage: Defenses

	Chrome	Safari	Edge	Firefox	Brave
Block third-party cookies					
Partition storage					
Ephemeral partitions					
List based defenses					



Tracking Techniques

- Third-party DOM storage
- Network state



- Browser fingerprinting
- IP address
- Personal identifiers

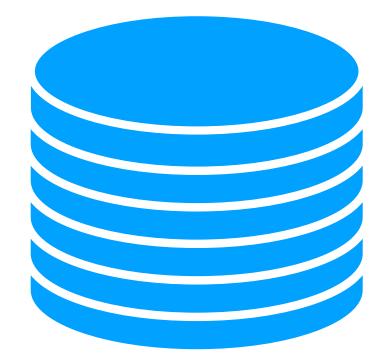


Network State Example: HTTP Cache

- **Browsers cache things for speed** Images, JavaScript, etc.
- **Caches are generally unpartitioned**
- Anything unpartitioned can be a linking key



HTTP Cache Tracking



some-site.example



- /pixel/#.gif?action={set, read}
- action=set: 50% return pixel : 50% 404
- action=read: 100% 404





HTTP Cache Tracking

	/ріх
	50%
	/ріх
	/pix
some-site.example	/pix
	/pix

/pixel/32.gif?action=set

. . .

xel/1.gif?action=set

% a pixel, 50% 404

xel/2.gif?action=set

xel/3.gif?action=set

xel/4.gif?action=set

xel/5.gif?action=set

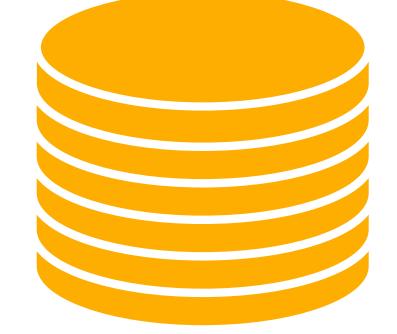


50% a pixel, 50% 404

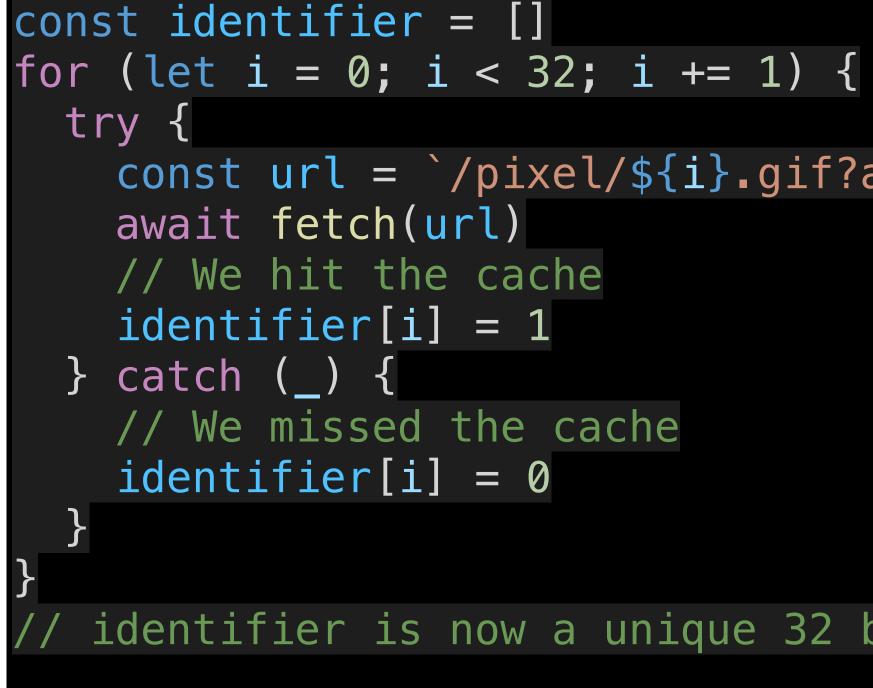




HTTP Cache Tracking



other-site.example



https://tracker.example/script.js

const url = `/pixel/\${i}.gif?action=read`

identifier is now a unique 32 bits







HTTP Strict Transport Security (HSTS)

- Website Says "only HTTPS, forever" e.g. persistent storage
- Automatic Upgrade http://example.org -> https://example.org
- How to leverage?

61

HSTS Tracking

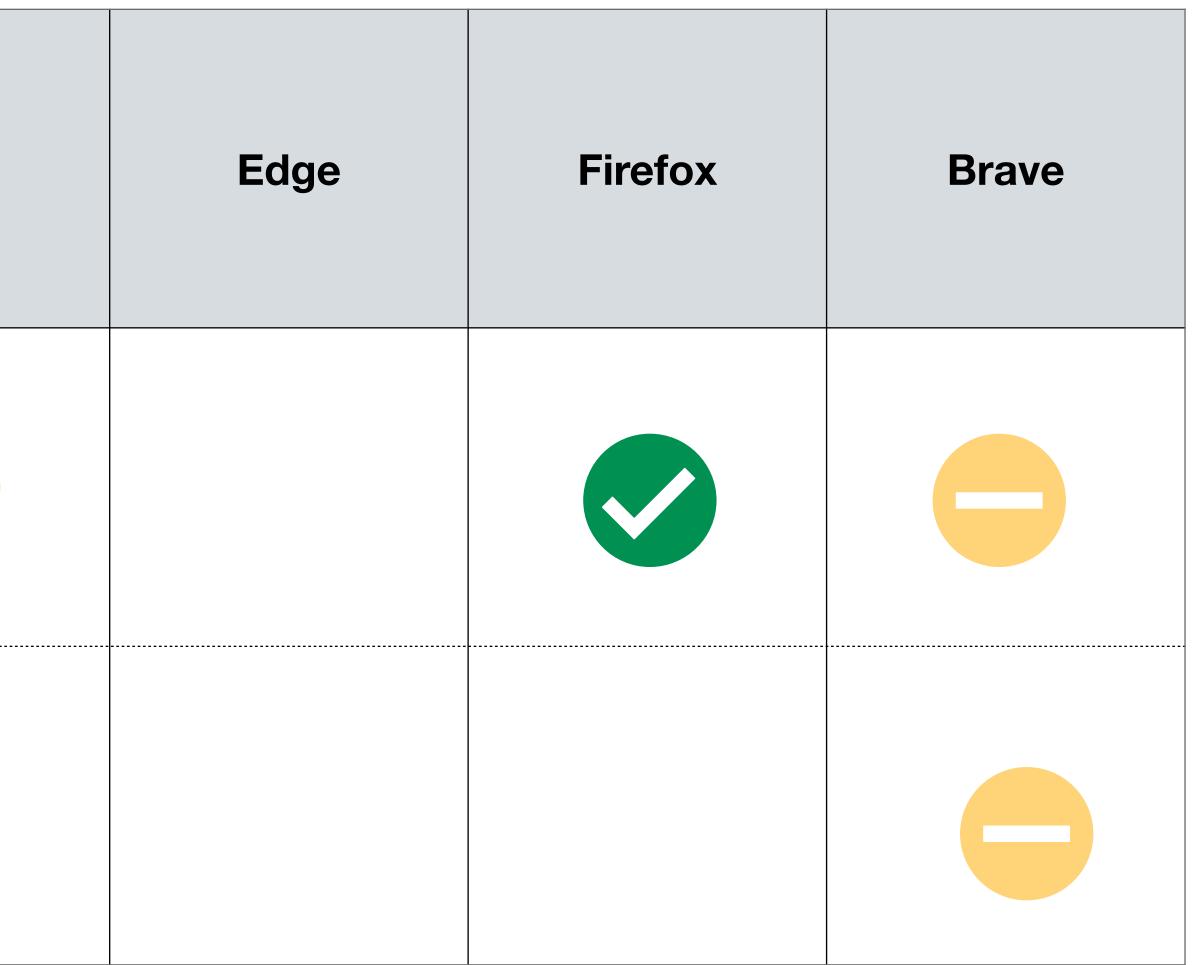
- example.org
- a.example.org
- b.example.org
- a.a.example.org
- b.a.example.org



Network state: Defenses

	Chrome	Safari
Partition network state		
List based defenses		







Tracking Techniques

- Third-party DOM storage
- Network state
- Bounce tracking



- Browser fingerprinting
- IP address
- Personal identifiers

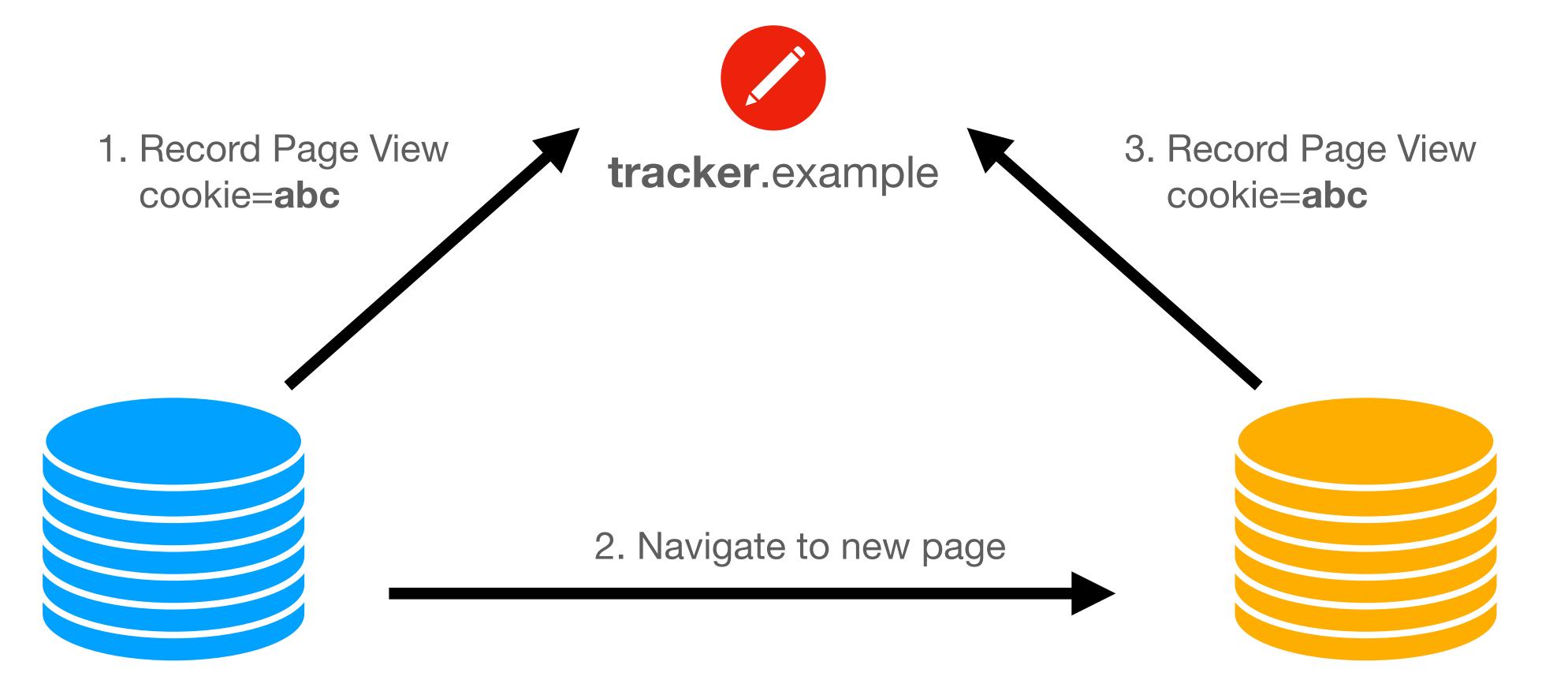


Bounce Tracking

- **Response to partitioning**
- Third parties use first-parties to track
- Growing in importance as partitioning is more common

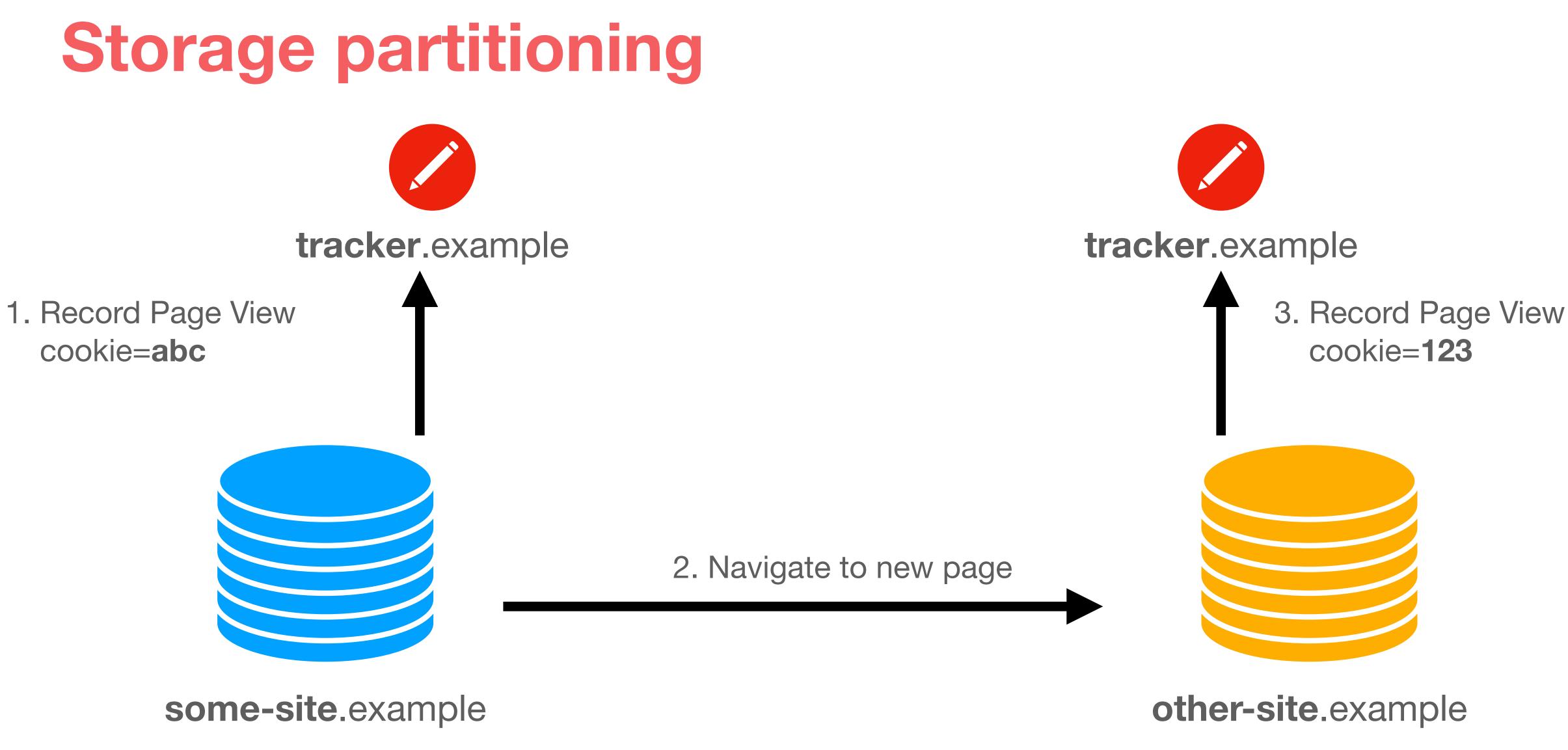


Pre-partitioning



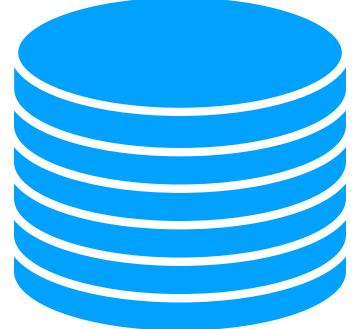
some-site.example

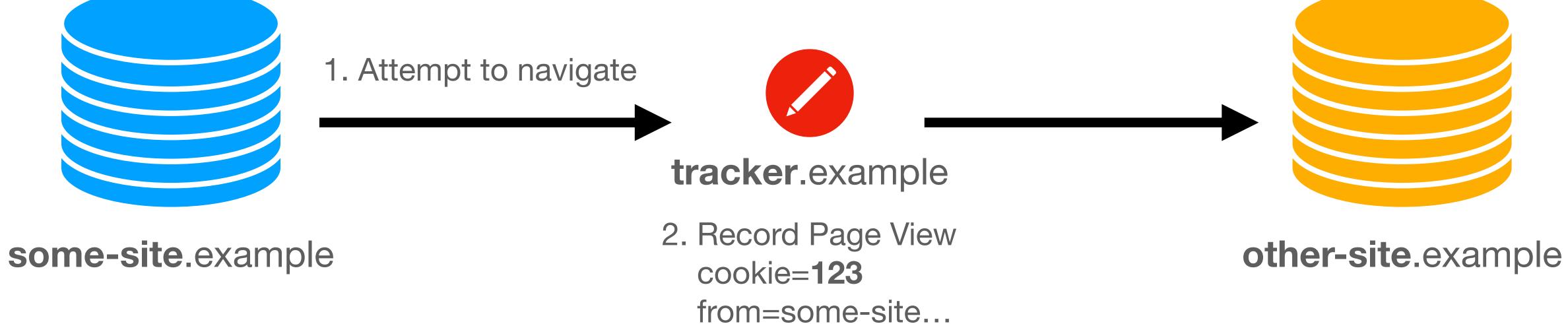
other-site.example





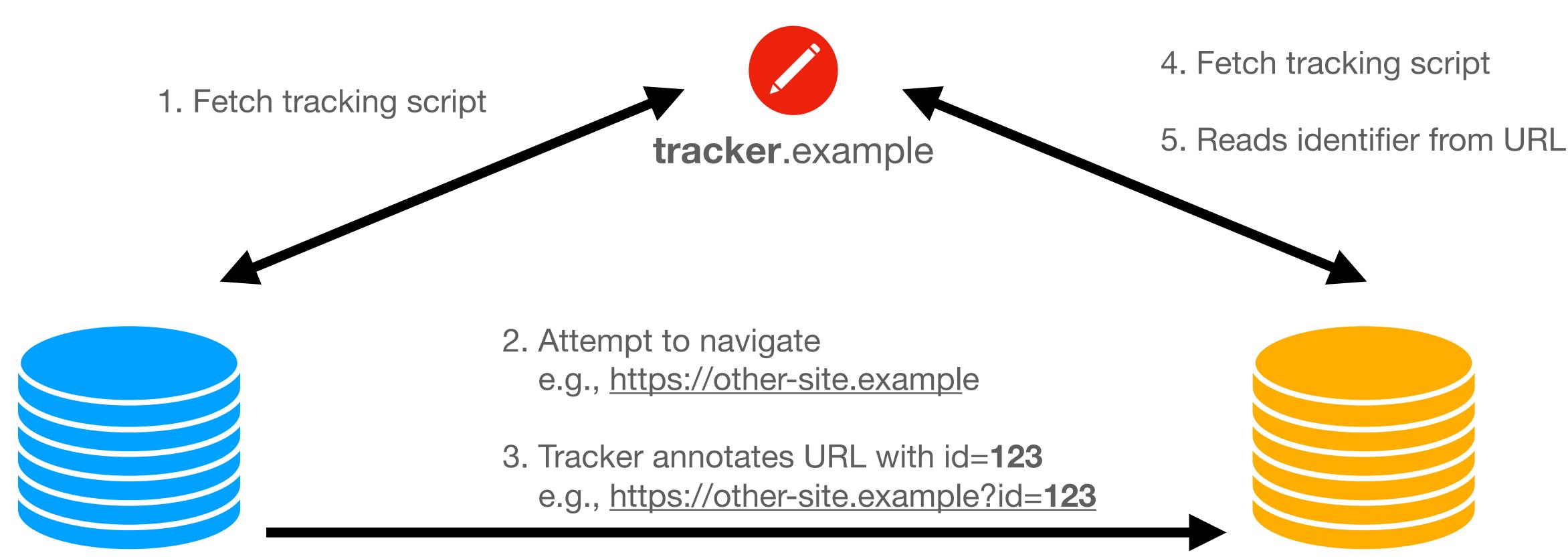
Bounce tracking





to=other-site...

Navigation tracking



some-site.example

other-site.example



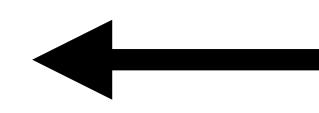
Bounce and Navigation Tracking: Defenses

	Chrome	Safari	Edge	Firefox	Brave
Limit storage		heuristic		List	
"Debounce"					List
Warn user					List



Tracking Techniques

- Third-party DOM storage
- Network state
- Bounce tracking
- Browser fingerprinting



- IP address
- Personal identifiers



Fingerprinting, contrasted

Classic tracking

- Website stores an id on the client
- The client returns the id to the server (cookie or JS)
- The id is what allows re-identification
- "Stateful"
- **Fingerprinting / passive tracking**
 - Website finds things different about each visitor
 - Tracker derives the identifier from minor browser differences
 - "Stateless"

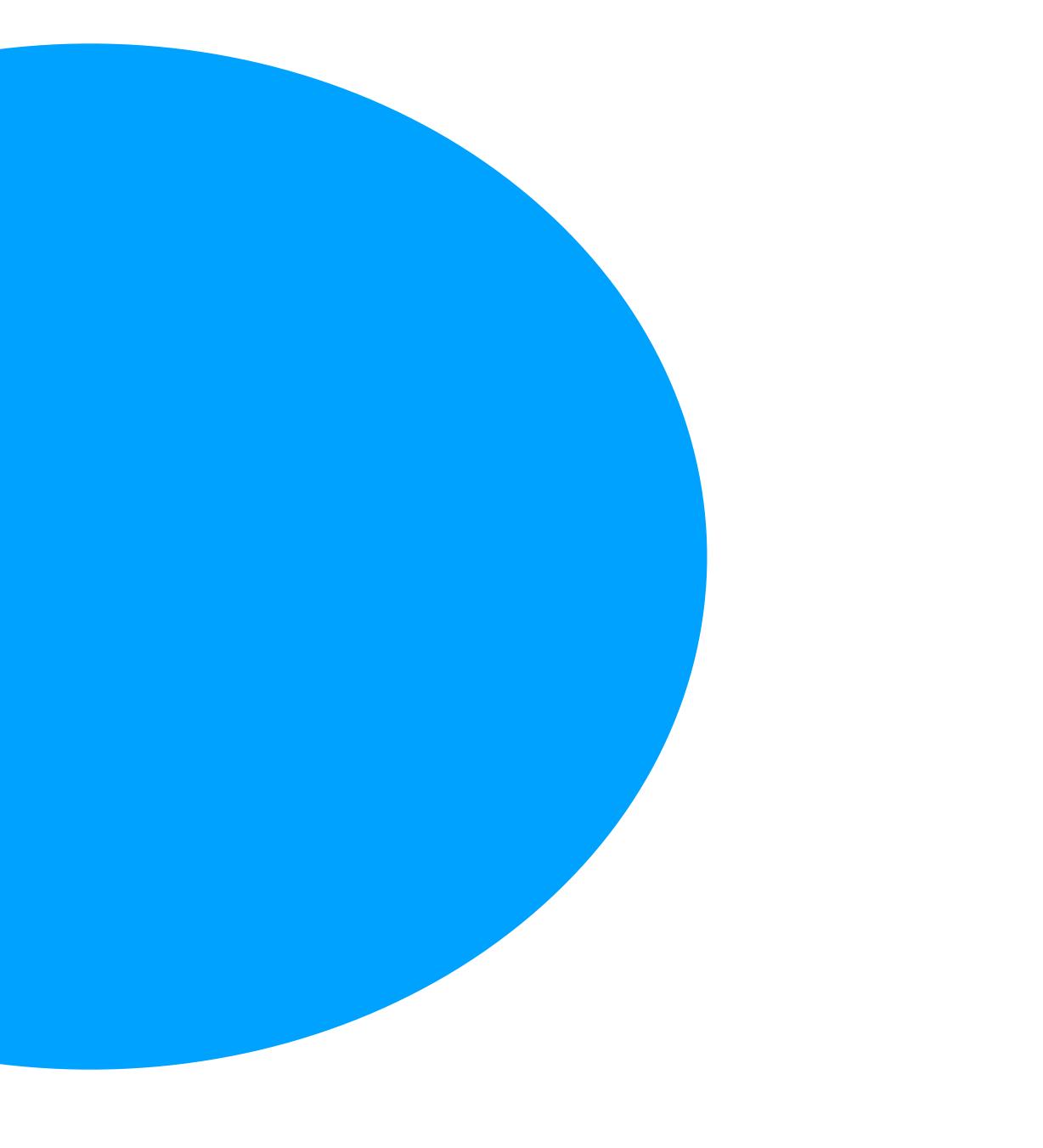


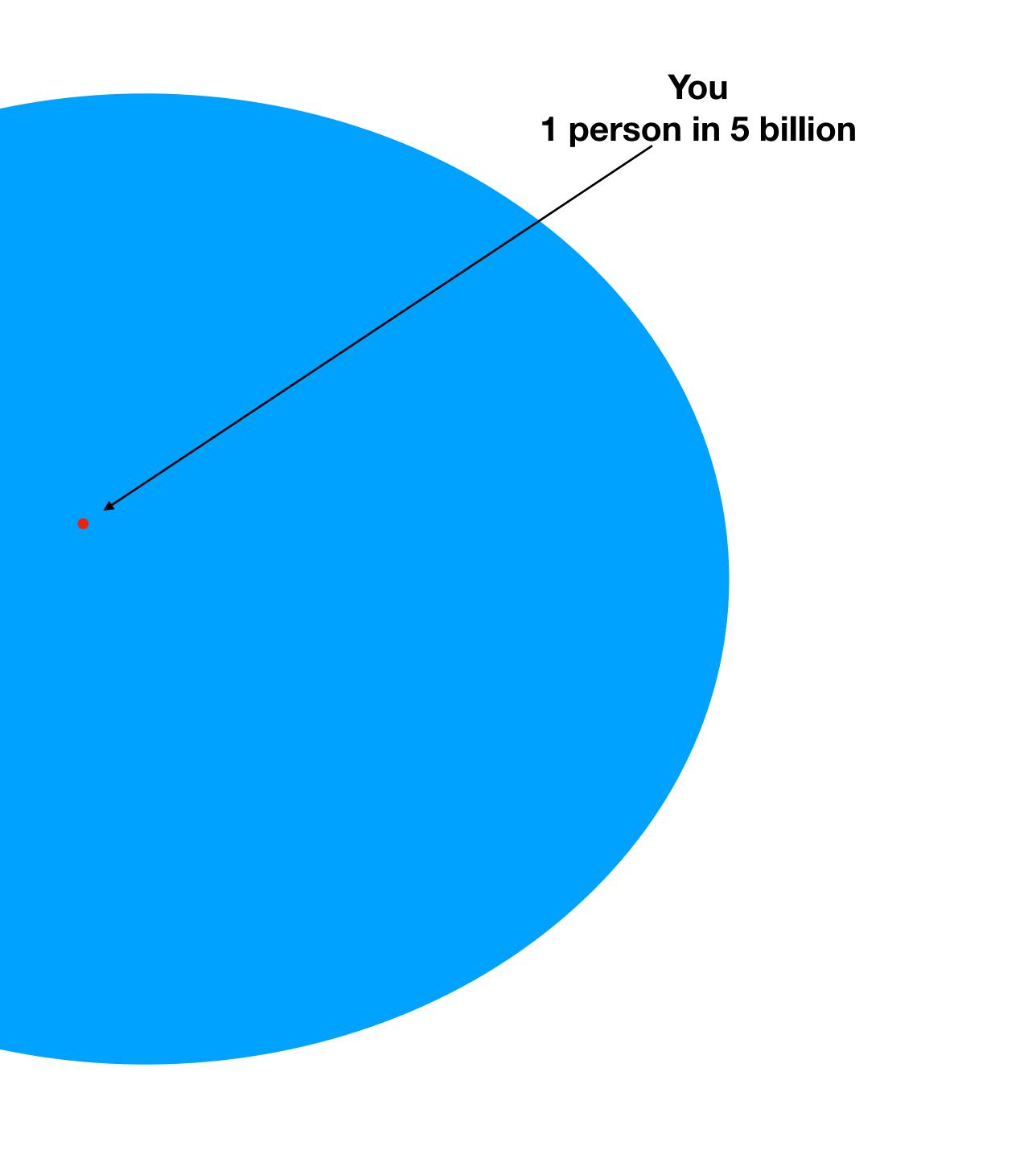
Fingerprinting, how?

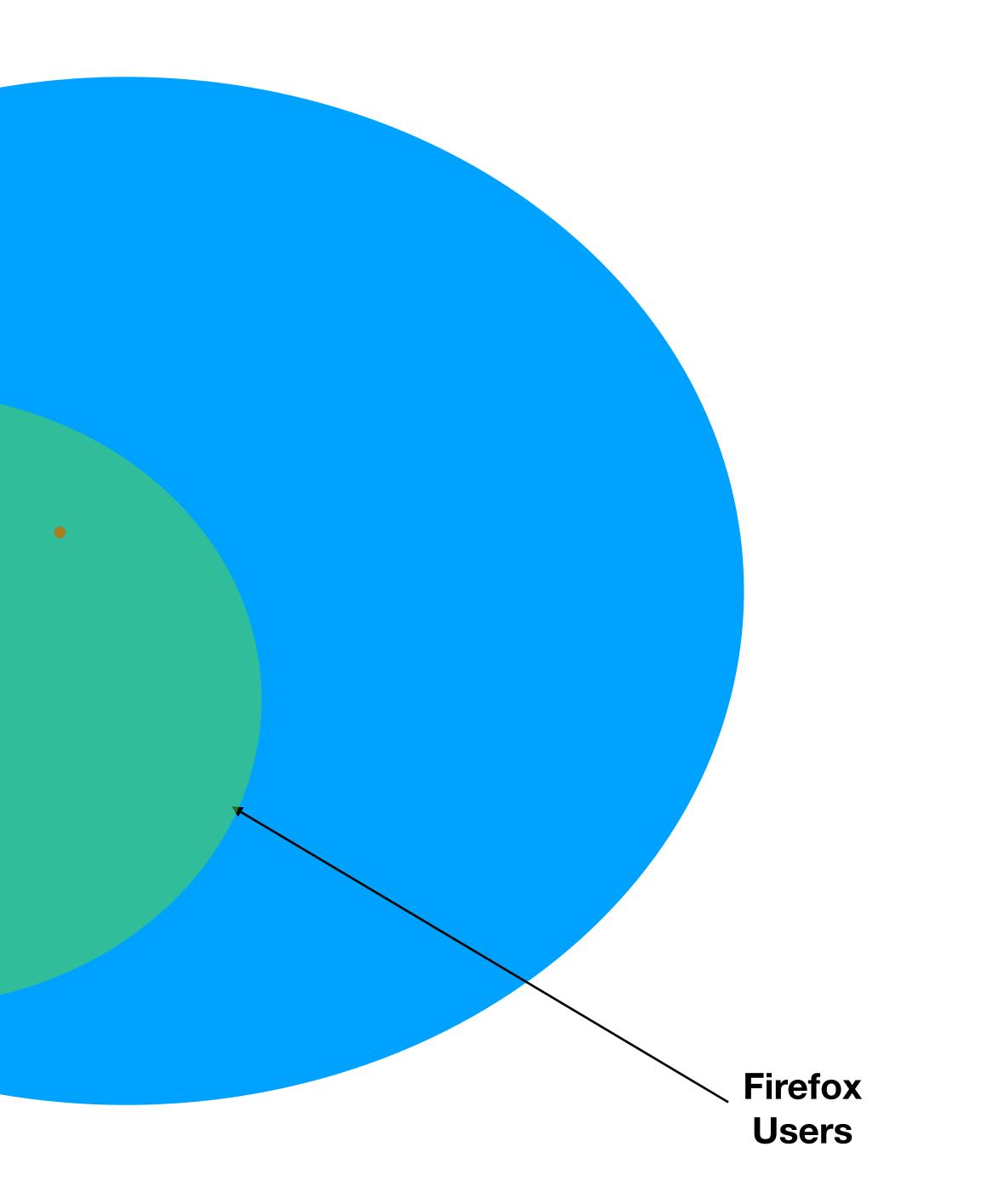
- Large number of semi-identifiers
 - Browser size
 - Extra fonts
 - Audio hardware
 - Video hardware
 - Installed plugins
 - Color depth
- Add the semi identification up...



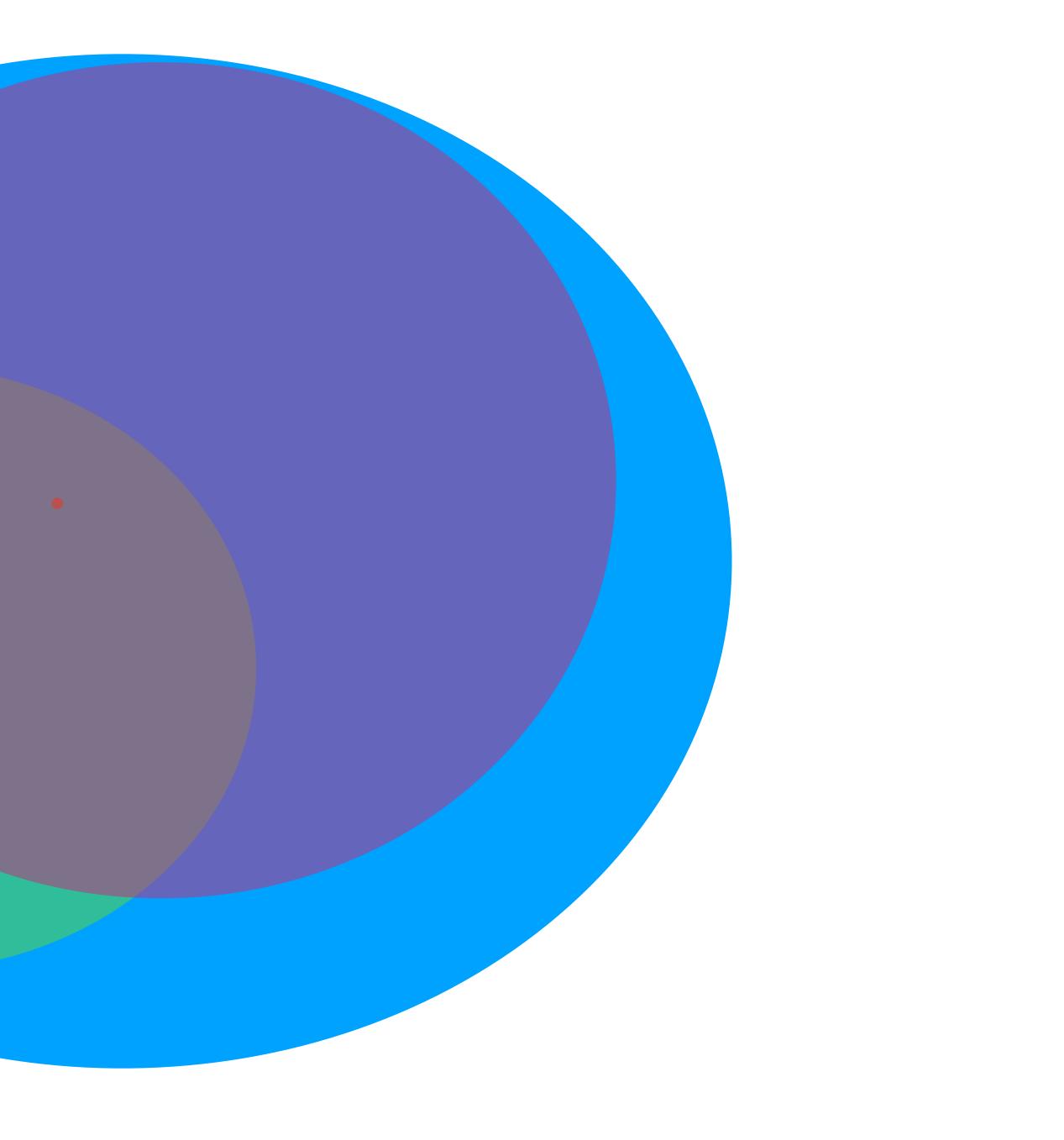
All browser users

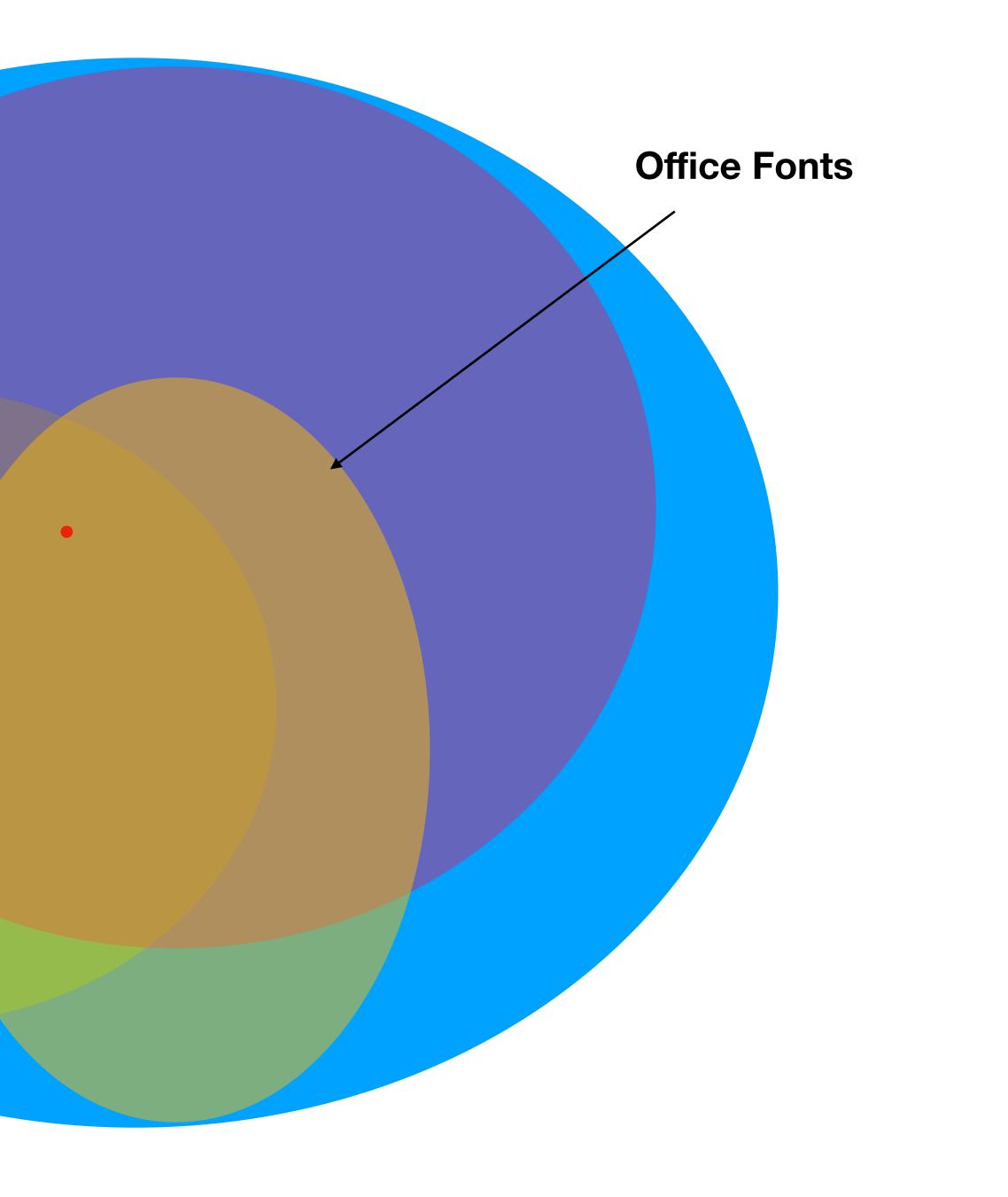


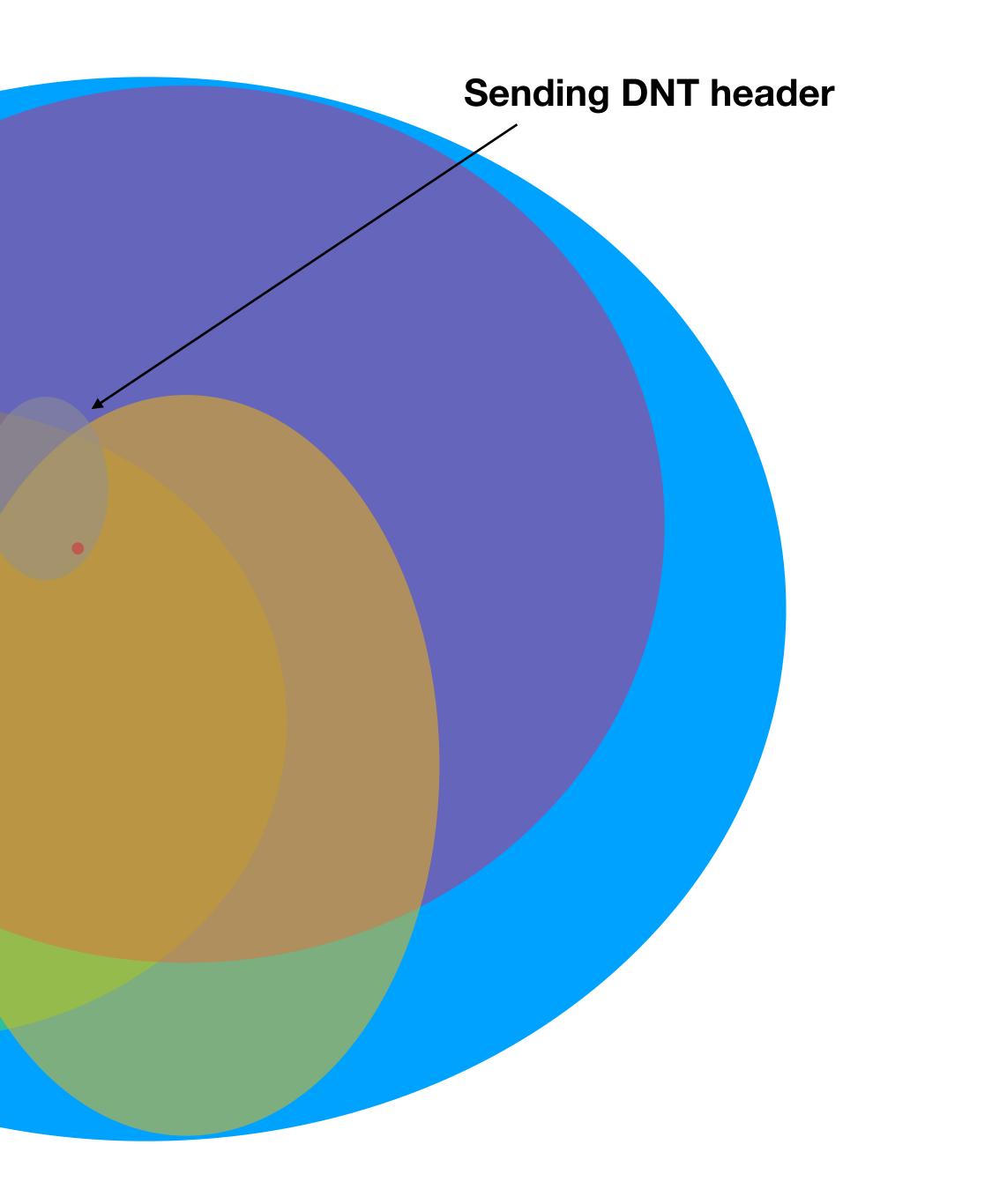




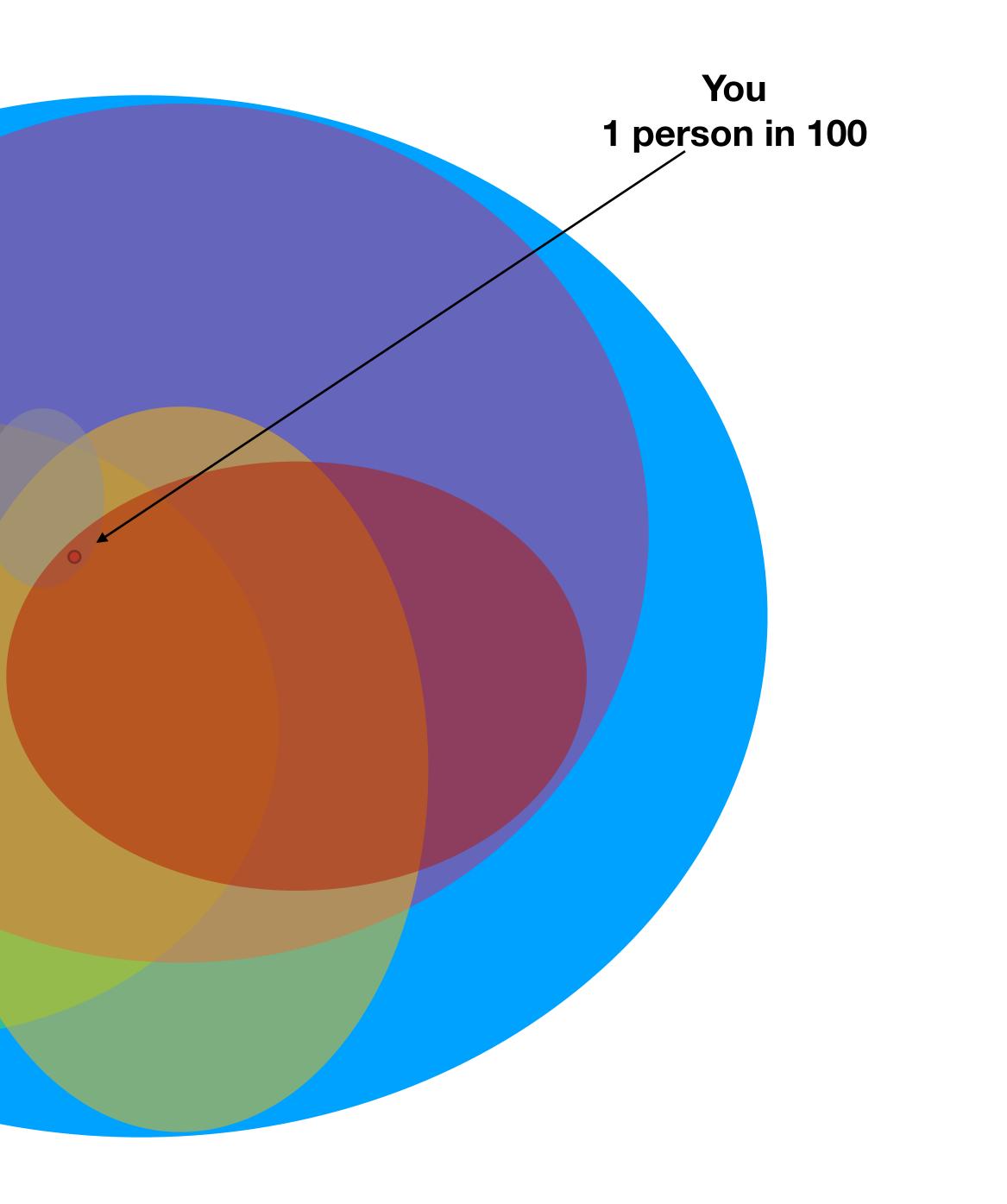
Windows users











Fingerprinting, abstracted

- Still needs a common value across boundaries Sites, sessions, time, etc
- Value needs to be unique Otherwise it mixes you up with others
- Value needs to be consistent Otherwise it doesn't (re)identify you





Possible Defenses

- Try to make browsers look similar Reduce the "bits" available to fingerprinters
- Try to block bad parties Keep the "bad folks" out
- **Privacy budgets** Only allow sites to do so much identifying, e.g., 10 bits but not more
- **Randomization** Make browser look ilntentionally different, within each boundary



Fingerprinting: Defenses

	Chrome	Safari	Edge	Firefox	Brave
Restricted hardware					
Feature selection / removal					
Block fingerprinters					
Randomization					





Tracking Techniques

- Third-party DOM storage
- Network state
- Bounce tracking
- Browser fingerprinting
- IP address
- Personal identifiers



IP Addresses are pretty unique!

- ...especially if you look for clusters The 3 ips you most commonly connect from is very unique
- **IPv6** makes it a lot worse Obviously...:-/
- Four general approaches
 - Contracts / promises
 - proxies
 - mix nets
 - block bad parties



IP Addresses Defenses

	Chrome	Safari	Edge	Firefox	Brave
Websites promise					
Proxies		Private relay		Optional VPN	
Mix networks					Optional Tor
Block bad parties					





Tracking Techniques

- Third-party DOM storage
- Network state
- Bounce tracking
- Browser fingerprinting
- IP address
- Personal identifiers





Personal Identifiers

- Names, email addresses, CCN, etc "old school"
- **Can be combined with offline sources** Credit agencies, public legal records, tax documents, etc
- **Baked into the web** Authentication, user accounts, etc



Partitioning to the Rescue (?)

- User holds the "true" value e.g., true email address
- **Browser holds a secret** e.g., secret = rand()
- **Derive per site identities** e.g. hash(email + secret + eTLD+1) + @private-email.com
- **Applicable to a range of identifiers** Email, CCN, Crypto addresses



Personal Identifiers: Defenses

	Chrome	Safari	Edge	Firefox	Brave
Partition email					
Partition Web3					
Block scripts					



Overview

- Why Privacy Matters
 A sloppy manifesto
- Defining Tracking
 Abstracting the problem
- Tracking in Practice
 Methods and defenses
- Privacy Beyond Tracking
 Other issues and concerns



Privacy is more than Absence of Tracking

- **Browsers shouldn't share information unless its helpful to user** e.g., FLoC
- **Browsers should serve users first and exclusively** e.g., Reporting API, FLEDGE
- **Browsers shouldn't introduce capabilities that remove user choice** WebBundles
- **Browsers shouldn't confuse users!** First-party sets, SXG
- First-parties are suspect too...



Other privacy protections

- **Governments increasing provide legal protections** GDPR, CCPA, etc
- **Browsers can help users assert their privacy rights** e.g., GlobalPrivacyControl
- Implemented in Brave and DDG
- Beware of conflating with "consent management" systems

Authored by activists, academics, New York Times, DuckDuckGo, Brave



A final plea...

- You are all plainly, amazingly smart people
- You'll be able to (mostly) choose your job
- **Privacy harms are particularly difficult to remediate**
- **Consider the privacy implications of a job before you take it**



A final plea...

- You are all plainly, amazingly smart people
- You'll be able to (mostly) choose your job
- **Privacy harms are particularly difficult to remediate**
- **Consider the privacy implications of a job before you take it**

Thanks!

