

CSE 484: Computer Security and Privacy

Privacy + Tracking

Spring 2023

David Kohlbrenner

dkohlbre@cs

Thanks to Franz Roesner, Dan Boneh, Dieter Gollmann, Dan Halperin, David Kohlbrenner, Yoshi Kohno, Ada Lerner, John Manferdelli, John Mitchell, Vitaly Shmatikov, Bennet Yee, and many others for sample slides and materials ...

Logistics

- Lab 2 is due next week
 - Remember we have a lot of resources/recordings on lab2 stuff!
- Lab 3 will go out shortly(?) after Lab 2 is due

Lab 2 debugging notes

- Step 1: log *every* page visit in your php script
 - I suggest logging more than the cookie param, maybe also a timestamp/etc
 - This will help you identify if the bot ever hits your page
- Step 2: verify that you can 'xss yourself'
 - If you visit the error page, does your capture script see a hit?
- Step 3: debug the error page
 - The error page is exactly what the bot is going to visit, so you can debug here
 - Look at the final source of the page (view-source)
 - Check for errors in the javascript console
 - Look at the requests being made in dev-tools/etc

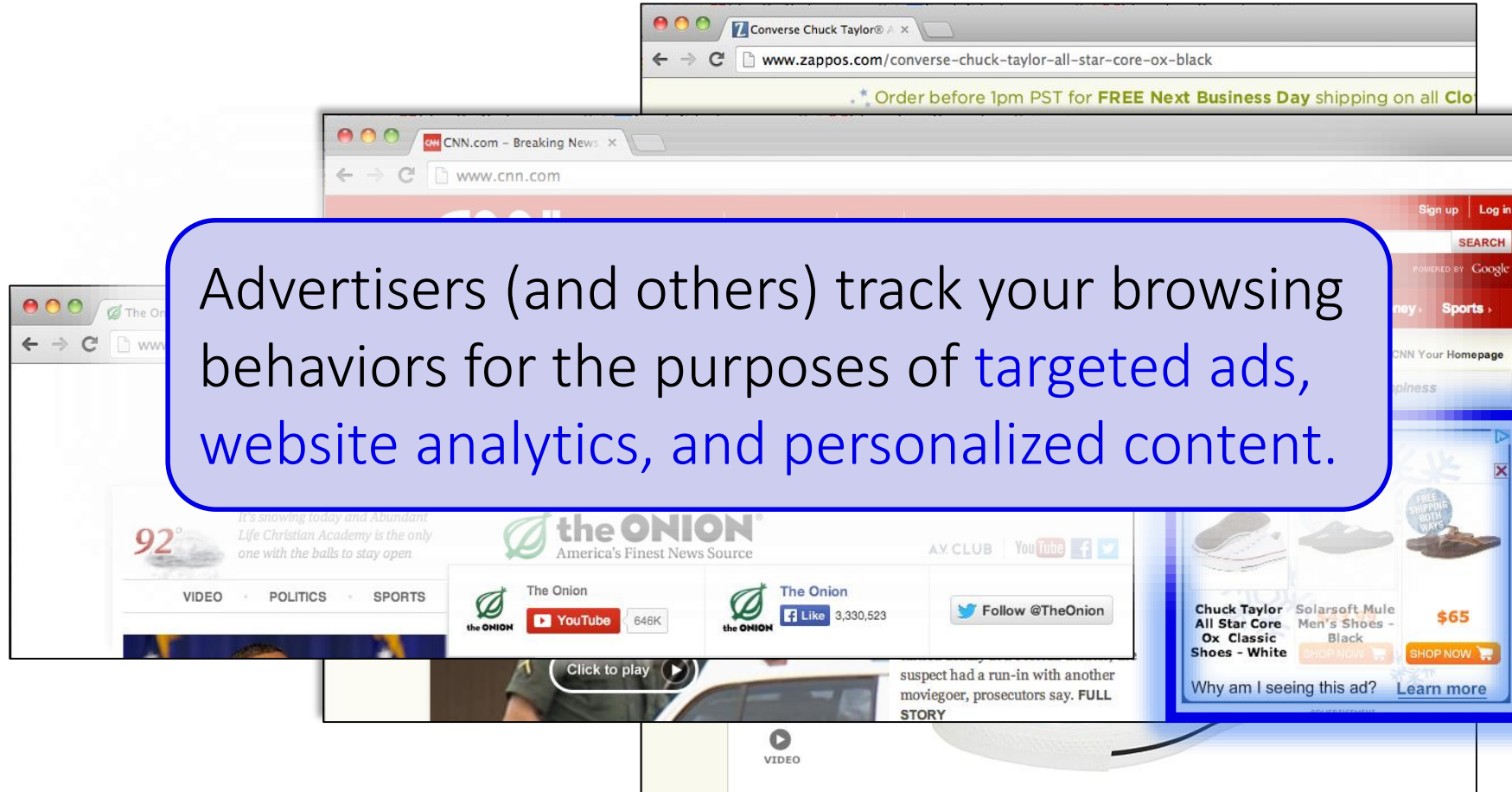
Privacy and web tracking

A topic in flux

- Tracking via cookies
- Tracking via other methods
- Fingerprinting

Ads That Follow You

Advertisers (and others) track your browsing behaviors for the purposes of **targeted ads**, website analytics, and personalized content.



Third-Party Web Tracking

The image shows a collage of browser windows. On the left, a window for 'The Onion' displays a Zappos ad. On the right, a window for 'CNN.com' also displays a Zappos ad. In the center, a blue box with a white border contains the text 'Browsing profile for user 123:' followed by a list of domains: 'cnn.com', 'theonion.com', 'adult-site.com', and 'political-site.com'. To the right of the list is a red sad face icon. The Zappos ads in the background show various shoe models and prices, with a 'SHOP NOW' button.

Browsing profile for user 123:

- cnn.com
- theonion.com
- adult-site.com
- political-site.com



These ads allow **criteo.com** to link your visits between sites, **even if you never click on the ads.**

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

2019

7,040 solutions

2018

6,829 solutions

2017

5,381 solutions

2016

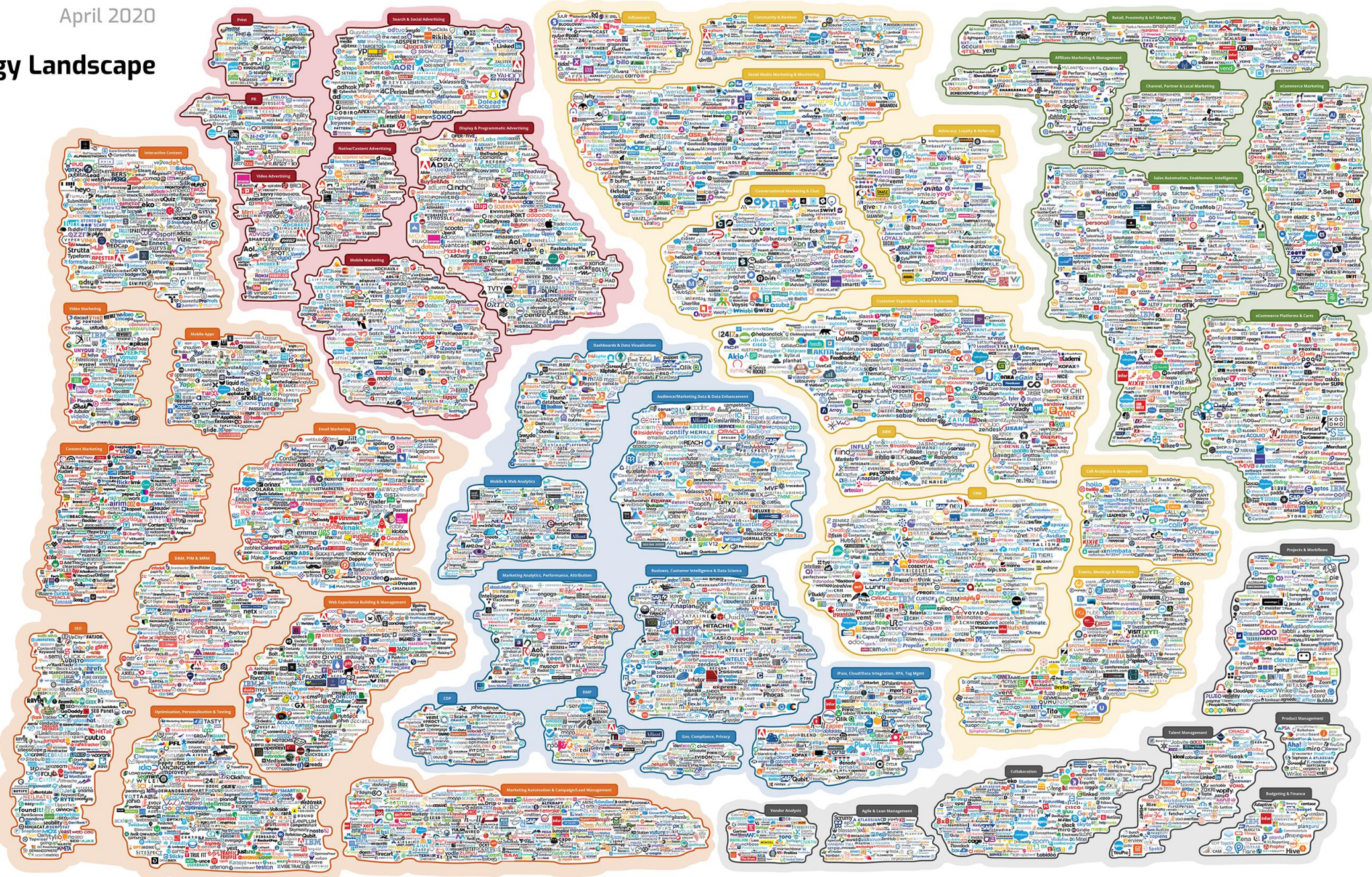
3,874 solutions

2015

1,876 solutions

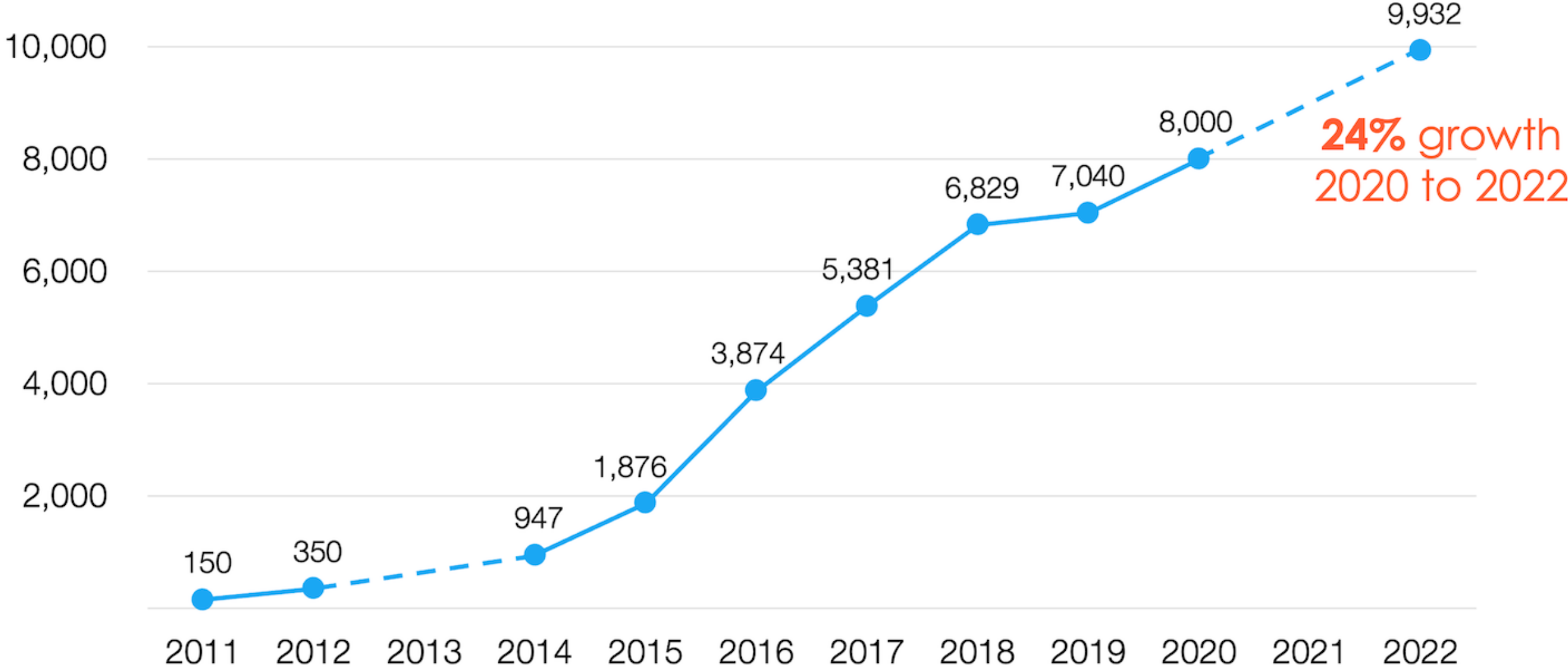
2014

947 solutions



| Advertising & Promotion | Content & Experience | Social & Relationships | Commerce & Sales | Data | Management |
|---|--|--|---|--|--|
| Display & Programmatic Advertising | Content Marketing | ABM | Retail, Proximity & IOT | Marketing Analytics Performance & Attribution | Agile & Lean Management |
| Mobile Marketing | CMS & Web Experience Management | Call Analytics & Management | Affiliate Marketing & Management | Audience/Marketing Data & Data Enhancement | Collaboration |
| Native/Content Advertising | DAM & MRM & PIM | Customer Experience Service & Success | Sales Automation Enablement & Intelligence | Ipaas Cloud/Data Integration & Tag Management | Talent Management |
| PR | Email Marketing | Influencers | eCommerce Platforms & Carts | Business/Customer Intelligence & Data Science | Budgeting & Finance |
| Print | Interactive Content | CRM | Advocacy Loyalty & Referrals | DMP | Projects & Workflow |
| Search & Social Advertising | Marketing Automation & Campaign/Lead Management | Community & Reviews | Events, Meetings & Webinars | Dashboards & Data Visualization | Governance Compliance And Privacy |
| Video Advertising | Mobile Apps | Social Media Marketing & Monitoring | eCommerce Marketing | Mobile & Web Analytics | Product Management |
| | Optimization Personalization & Testing | Live Chat & Chatbots | Channel Partner & Local Marketing | Customer Data Platform | Vendor Analysis & Management |
| | SEO | | | | |
| | Video Marketing | | | | |

6,521% growth 2011 to 2022



<https://chiefmartec.com/2022/05/marketing-technology-landscape-2022-search-9932-solutions-on-martechmap-com/>

Concerns About Privacy

THE WALL STREET JOURNAL.
WHAT THEY KNOW | JULY 30, 2010

The New York Times
May 6, 2011, 5:01 pm | 3 Comments

'Do Not Track' Privacy Bill Appears in Congress

By TANZINA VEGA

And the privacy legislation just keeps on coming.

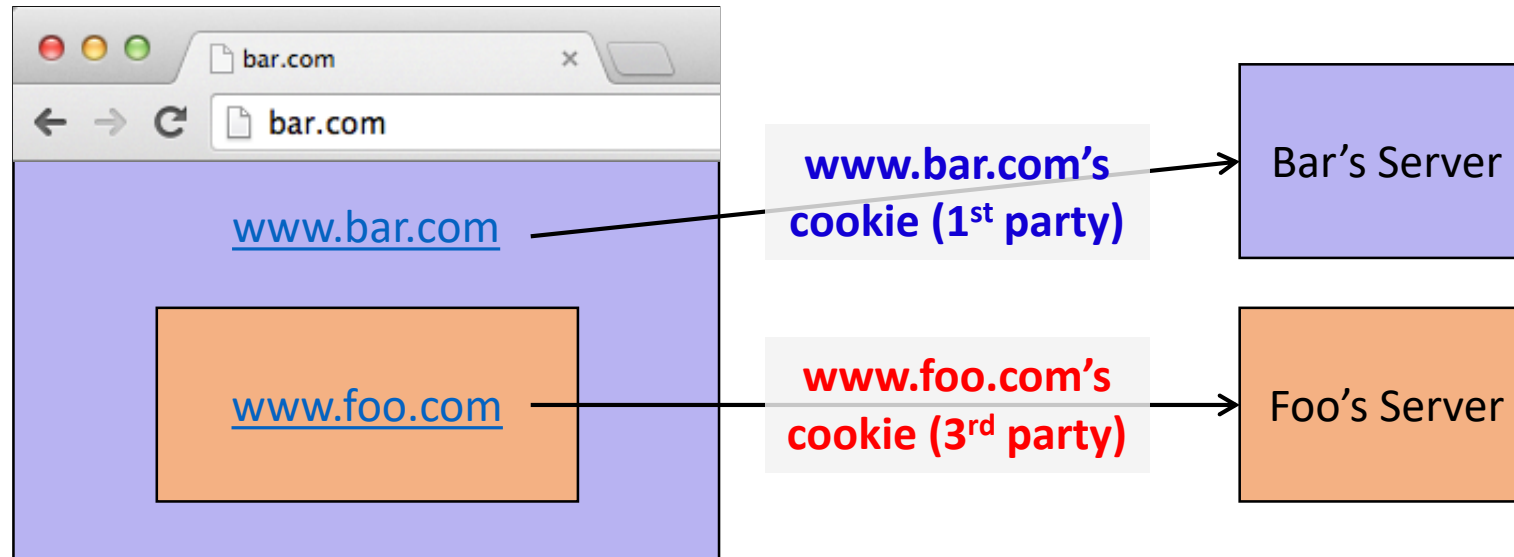
On Friday, two bills were introduced in Washington in support of a Do Not Track mechanism that would give users control over how much of their data was collected by advertisers and other online companies.

By JENNIFER VALENTINO-DEVRIES,
JEREMY SINGER-VINE and ASHKAN SOLTANI
December 24, 2012

als
ion

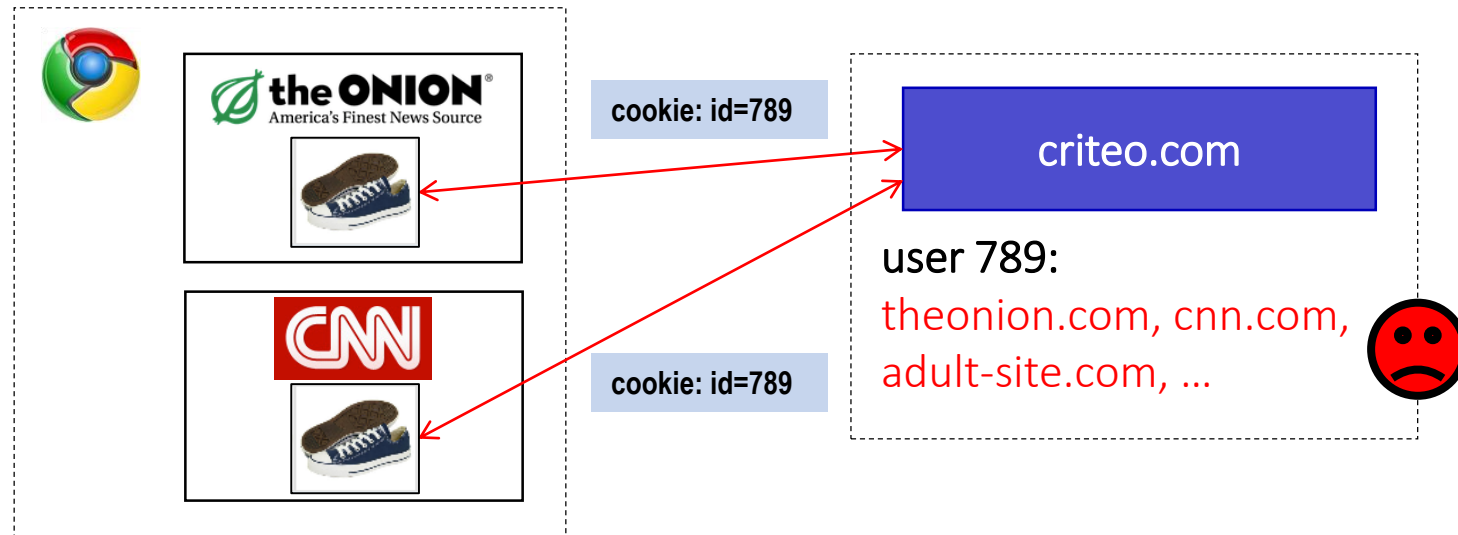
First and Third Parties

- **First-party cookie:** belongs to top-level domain.
- **Third-party cookie:** belongs to domain of embedded content (such as image, iframe).



Anonymous Tracking

Trackers **included in other sites** use **third-party cookies** containing unique **identifiers** to create browsing profiles.



Basic Tracking Mechanisms

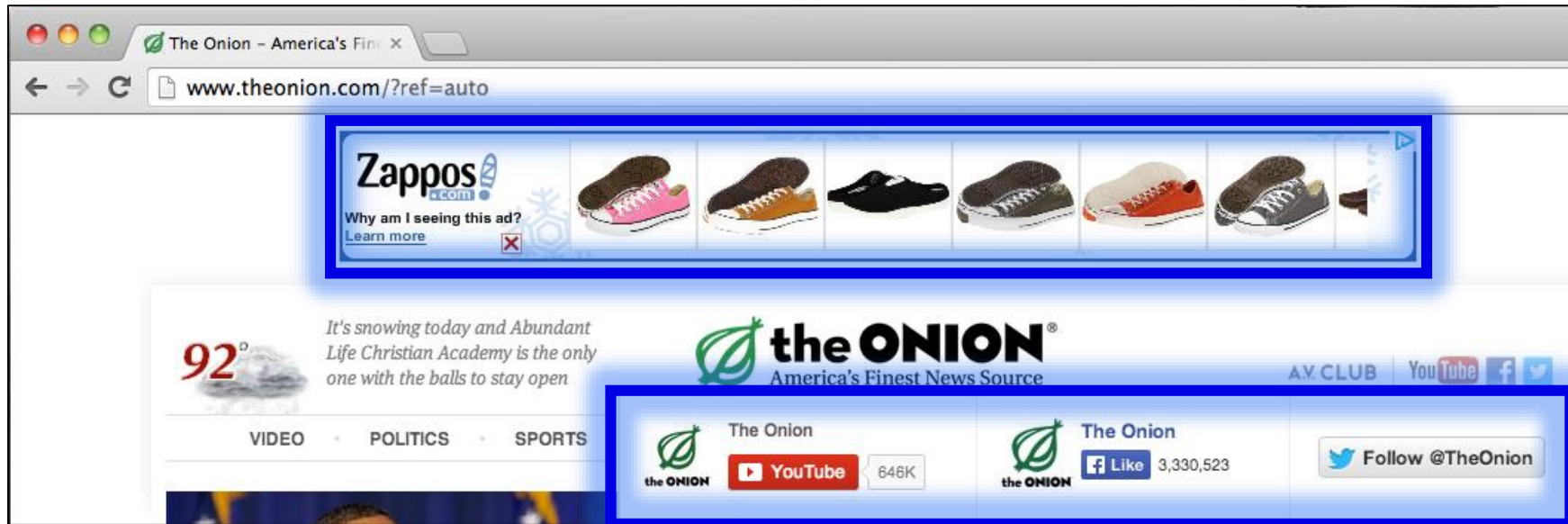
- Tracking requires:
 - (1) re-identifying a user.
 - (2) communicating id + visited site back to tracker.

```
▼ Hypertext Transfer Protocol
  ▸ GET /pixel/p-3aud4J6uA4Z6Y.gif?labels=InvisibleBox&busty=2710 HTTP/1.1\r\n
    Host: pixel.quantserve.com\r\n
    Connection: keep-alive\r\n
    Accept: image/webp,*/*;q=0.8\r\n
    User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.36\r\n
    Referer: http://www.theonion.com/\r\n
    Accept-Encoding: gzip,deflate,sdch\r\n
    Accept-Language: en-US,en;q=0.8\r\n
    Cookie: mc=52a65386-f1de1-00ade-0b26e; d=ENkBRgGHD4GYEA35MMIL74MKiyDs1A2MQI1Q
```

Tracking Technologies

- HTTP Cookies
- HTTP Auth
- HTTP Etags
- Content cache
- IE userData
- HTML5 protocol and content handlers
- HTML5 storage
- Flash cookies
- Silverlight storage
- TLS session ID & resume
- Browsing history
- window.name
- HTTP STS
- DNS cache
- “Zombie” cookies that respawn (<http://samy.pl/evercookie>)

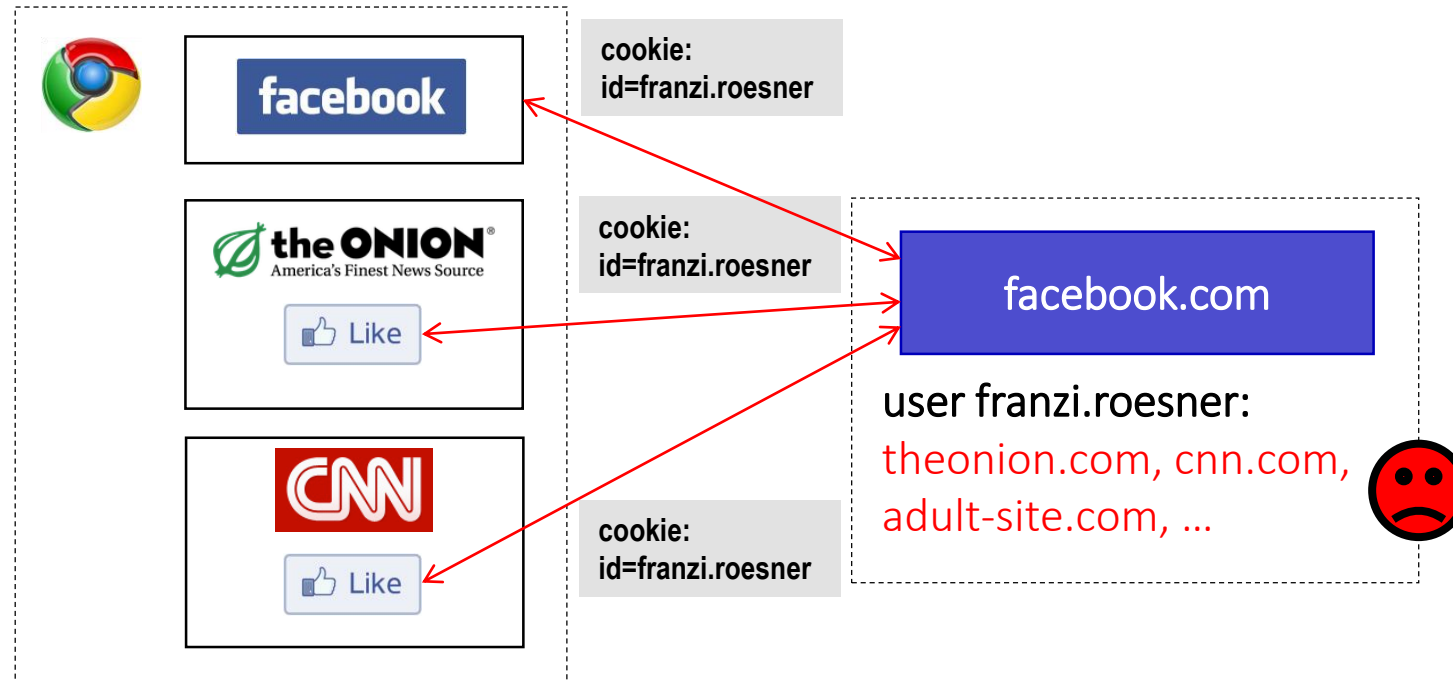
Other Trackers?



“Personal” Trackers



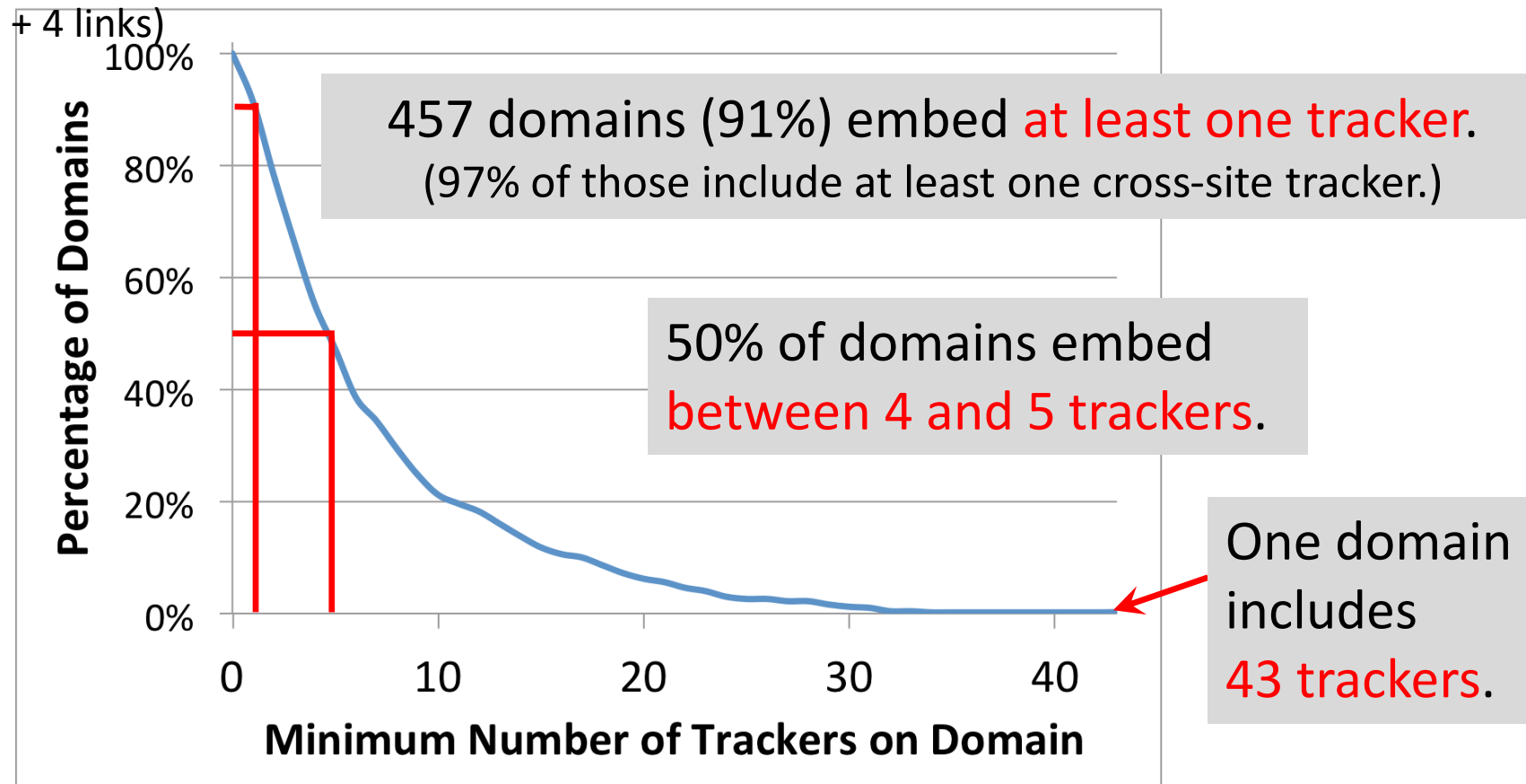
Personal Tracking



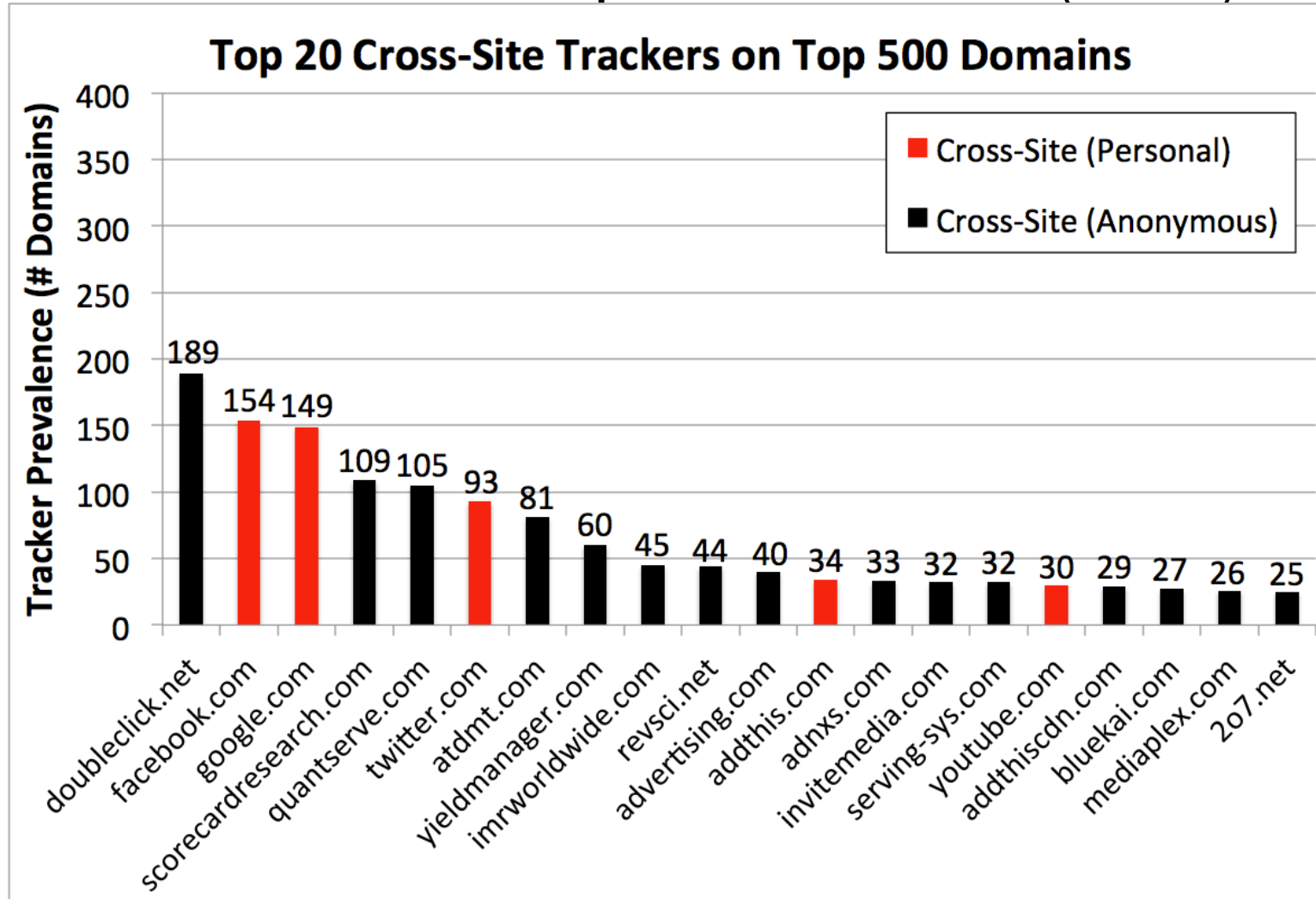
- Tracking is **not anonymous** (linked to accounts).
- Users **directly visit tracker's site** → evades some defenses.

How prevalent is tracking? (2011)

524 unique trackers on Alexa top 500 websites (homepages)



Who/what are the top trackers? (2011)

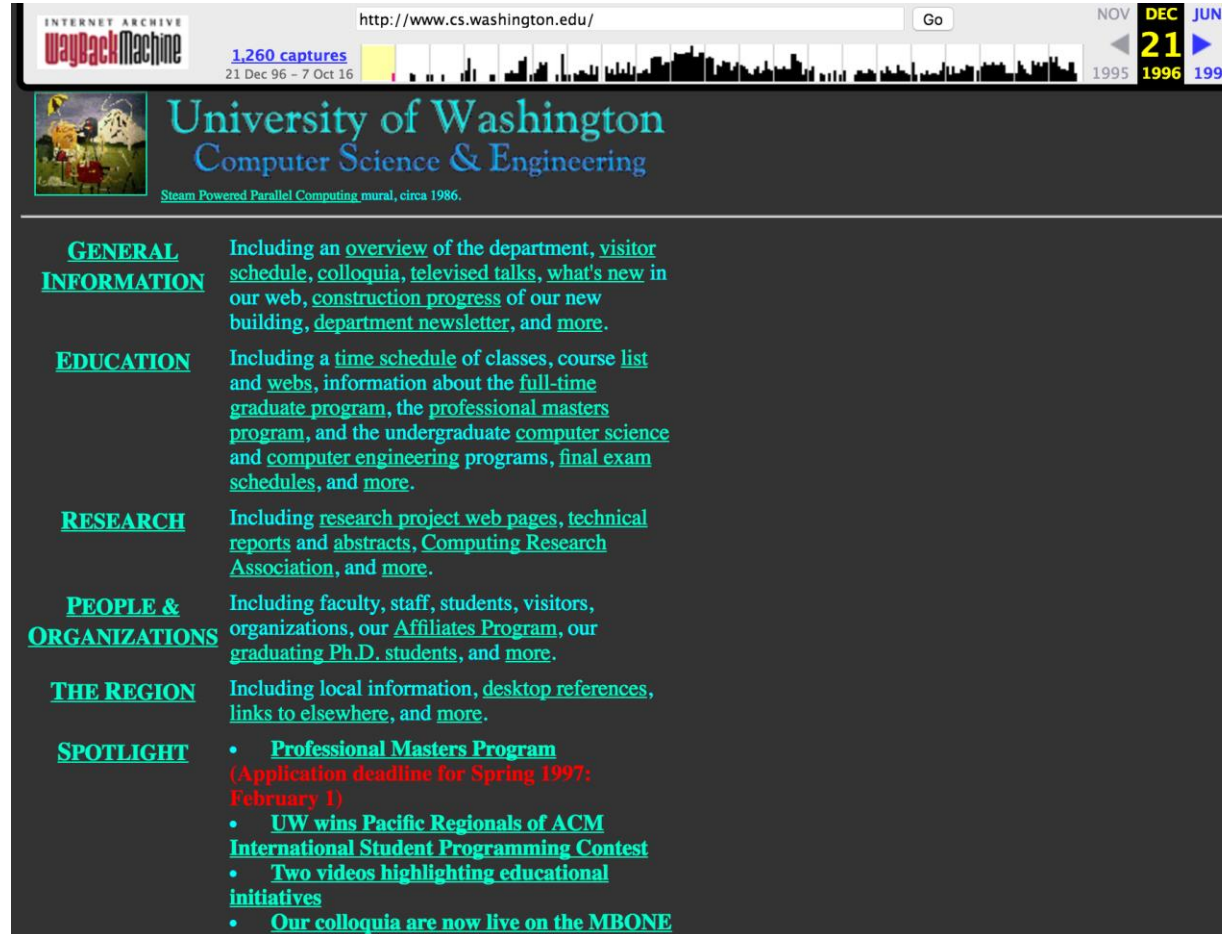


How has this changed over time?

- The web has existed for a while now...
 - What about tracking before 2011?
 - What about tracking before 2009?
- Solution: time travel!



The Wayback Machine to the Rescue



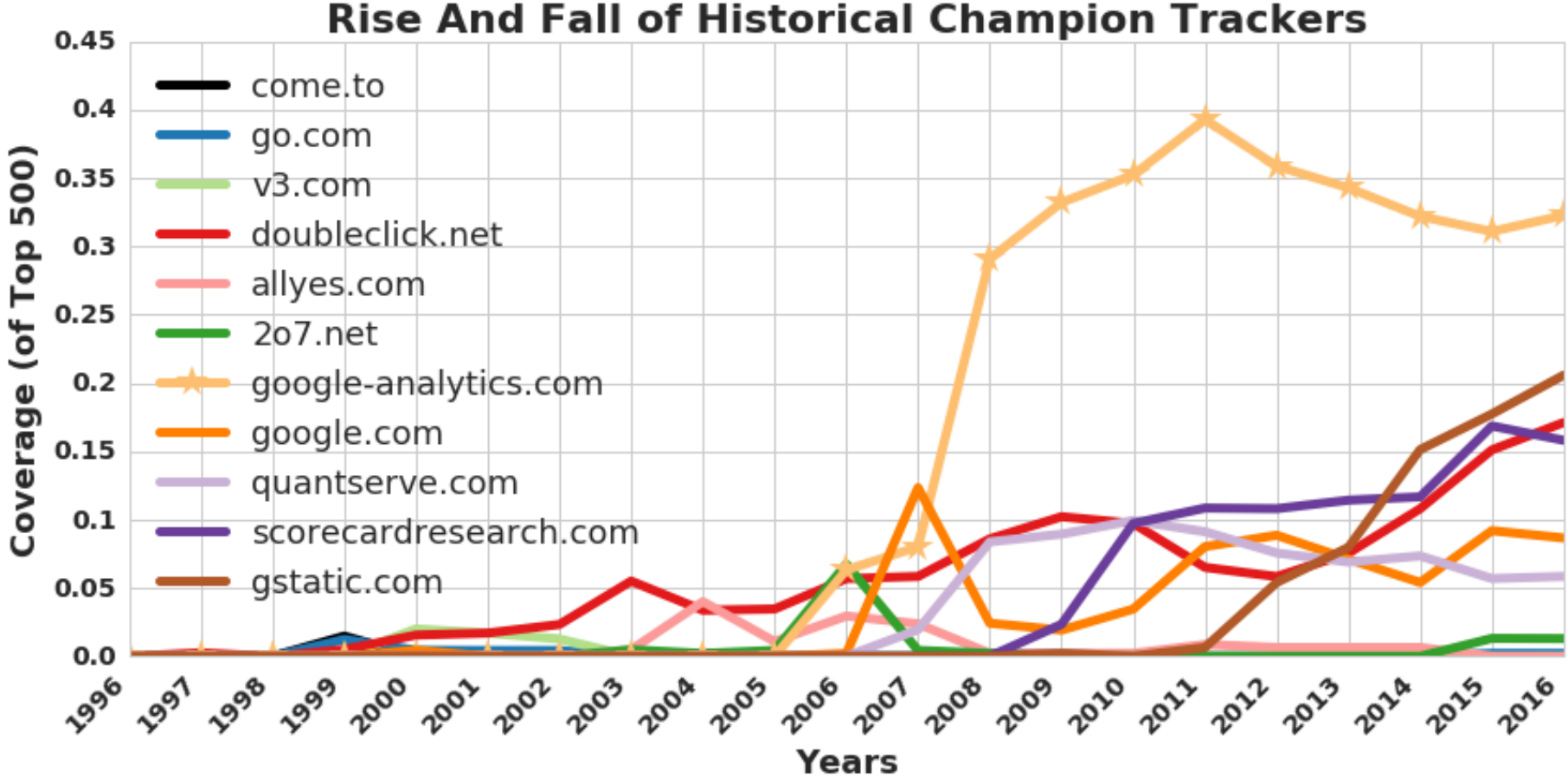
The screenshot shows a Wayback Machine interface for the URL <http://www.cs.washington.edu/>. The page is dated December 21, 1996. The header features the Wayback Machine logo, a capture count of 1,260, and a calendar navigation showing the current date as 21/12/1996. The main content area is titled "University of Washington Computer Science & Engineering" and includes a small image of a mural titled "Steam Powered Parallel Computing mural, circa 1986". The page is organized into several sections with blue headers and black text:

- GENERAL INFORMATION**: Including an [overview](#) of the department, [visitor schedule](#), [colloquia](#), [televised talks](#), [what's new](#) in our web, [construction progress](#) of our new building, [department newsletter](#), and [more](#).
- EDUCATION**: Including a [time schedule](#) of classes, [course list](#) and [webs](#), information about the [full-time graduate program](#), the [professional masters program](#), and the undergraduate [computer science](#) and [computer engineering](#) programs, [final exam schedules](#), and [more](#).
- RESEARCH**: Including [research project web pages](#), [technical reports](#) and [abstracts](#), [Computing Research Association](#), and [more](#).
- PEOPLE & ORGANIZATIONS**: Including faculty, staff, students, visitors, organizations, our [Affiliates Program](#), our [graduating Ph.D. students](#), and [more](#).
- THE REGION**: Including local information, [desktop references](#), [links to elsewhere](#), and [more](#).
- SPOTLIGHT**:
 - [Professional Masters Program](#) (Application deadline for Spring 1997: February 1)
 - [UW wins Pacific Regionals of ACM International Student Programming Contest](#)
 - [Two videos highlighting educational initiatives](#)
 - [Our colloquia are now live on the MBONE](#)

Time travel for web tracking: <http://trackingexcavator.cs.washington.edu>

1996-2016: More & More Tracking

- More trackers of more types, more per site, **more coverage**



Defenses to Reduce Tracking

- Do Not Track?

Send a 'Do Not Track' request with your browsing traffic

Do Not Track is not a technical defense:
trackers must honor the request.

Defenses to Reduce Tracking

- Do Not Track proposal?
- Private browsing mode?

Private browsing mode doesn't protect against network attackers fully.

You've gone incognito

Now you can browse privately, and other people who use this device won't see your activity. However, downloads and bookmarks will be saved. [Learn more](#)

Chrome won't save the following information:

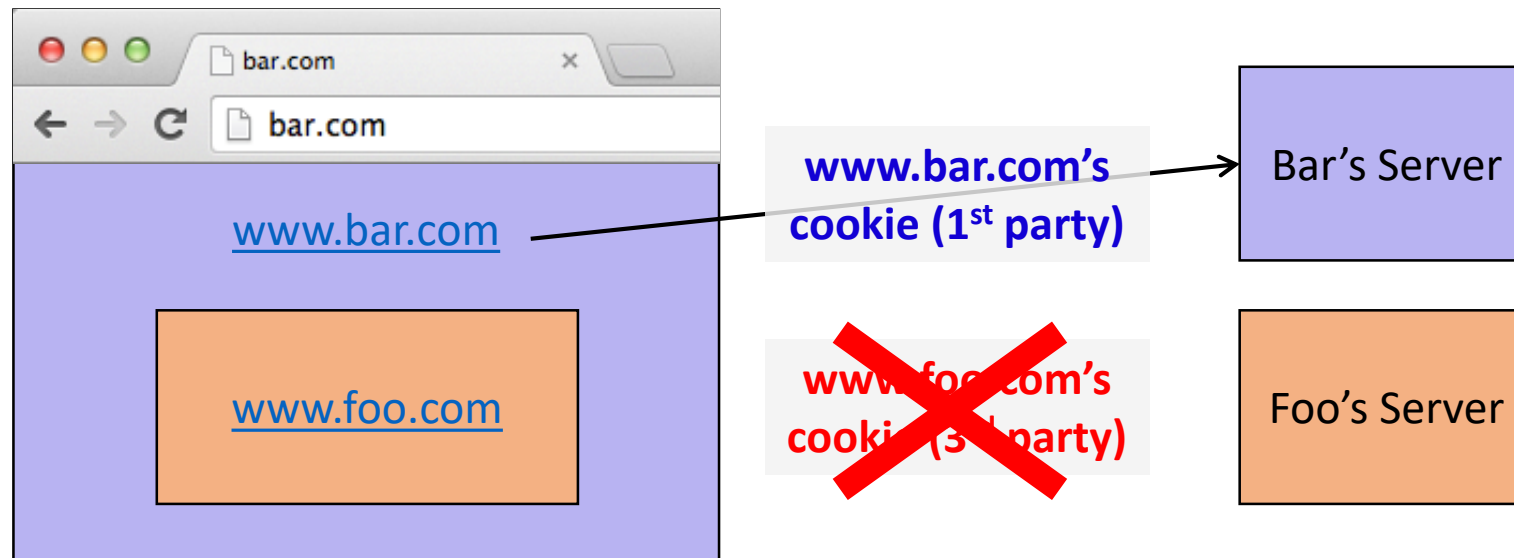
- Your browsing history
- Cookies and site data
- Information entered in forms

Your activity might still be visible to:

- Websites you visit
- Your employer or school
- Your internet service provider

Defenses to Reduce Tracking

- Do Not Track proposal?
- Private browsing mode?
- Third-party cookie blocking?



3rd party cookies

- Safari and FF (mostly) now block 3rd party cookies
 - <https://webkit.org/blog/10218/full-third-party-cookie-blocking-and-more/>
 - <https://blog.mozilla.org/blog/2019/09/03/todays-firefox-blocks-third-party-tracking-cookies-and-cryptomining-by-default/>

- Chrome...

“By undermining the business model of many ad-supported websites, blunt approaches to cookies encourage the use of opaque techniques such as fingerprinting (an invasive workaround to replace cookies), which can actually reduce user privacy and control. We believe that we as a community can, and must, do better.”

Aug 2022: Remove 3rd party cookies by 2024

How should Google respond?

- Canvas!
- Pretend someone fired all the ad/chrome execs and hired your group instead
- Safari and Firefox have removed ad's ability to track users via 3rd party cookies, and Google has committed to the same in Chrome by 2024.
- How should google respond to 3rd party cookies being removed?
- Think about the technical solutions, policy solutions, and even business model solutions available to you!

Cookie Ghostwriting

- flickr.com
 - `<script src=siftscience.com/s.js />`
- S.js runs
 - Fp = fingerprintjs2
 - Setcookie(fp)
 - Hexagon-analytics.com/cookieeciever?cookie=fp
- Every time you load flickr.com what happens?

Journey to the Center of the Cookie Ecosystem: Unraveling Actors' Roles and Relationship

Cookie Ghostwriting

- ~~flickr.com~~ patreon.com
 - `<script src=siftscience.com/s.js />`
- S.js runs
 - Fp = fingerprintjs2
 - Setcookie(fp)
 - Hexagon-analytics.com/cookieeciever?cookie=fp
- Every time you load ~~flickr.com~~ patreon.com what happens?

Journey to the Center of the Cookie Ecosystem: Unraveling Actors' Roles and Relationship

Fingerprinting is out there

- Better than a ‘voluntary’ cookie: involuntary, unchangeable id!
 - “Fingerprint”
- Idea: Measure ‘behavior’ of browser
 - Smash into unique ID

Fingerprinting Web Browsers

- User agent
- HTTP ACCEPT headers
- Browser plug-ins
- MIME support
- Clock skew
- Installed fonts
- Cookies enabled?
- Browser add-ons
- Screen resolution
- HTML5 canvas
(differences in graphics SW/HW!)

HTML5 Canvas Fingerprinting - Text

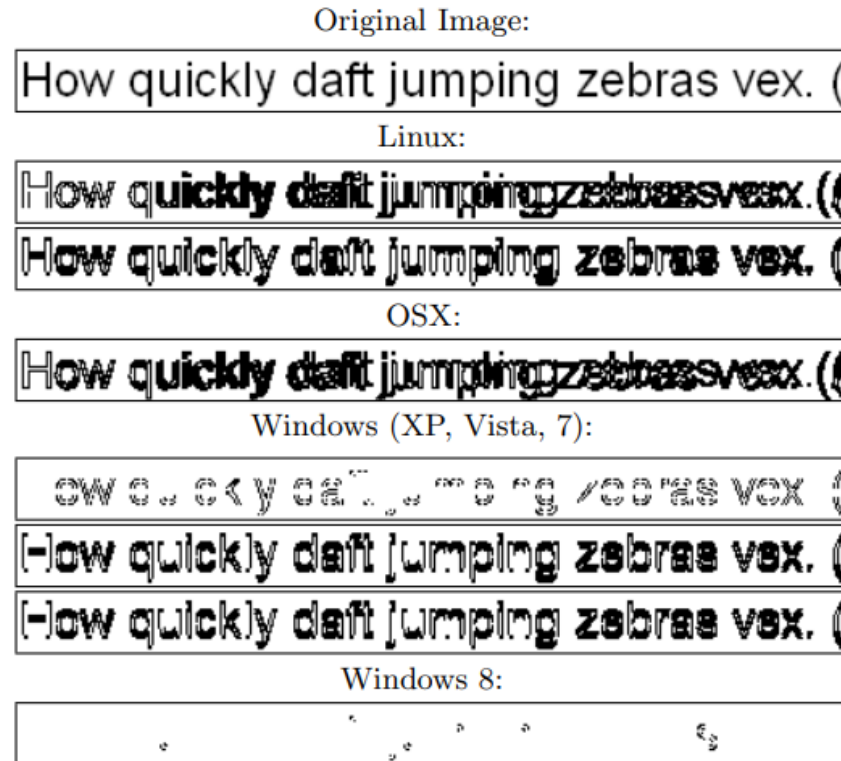


Figure 7: Difference maps for a group on text_arial

Mowery and Shacham, 2012

HTML5 Canvas Fingerprinting - Image

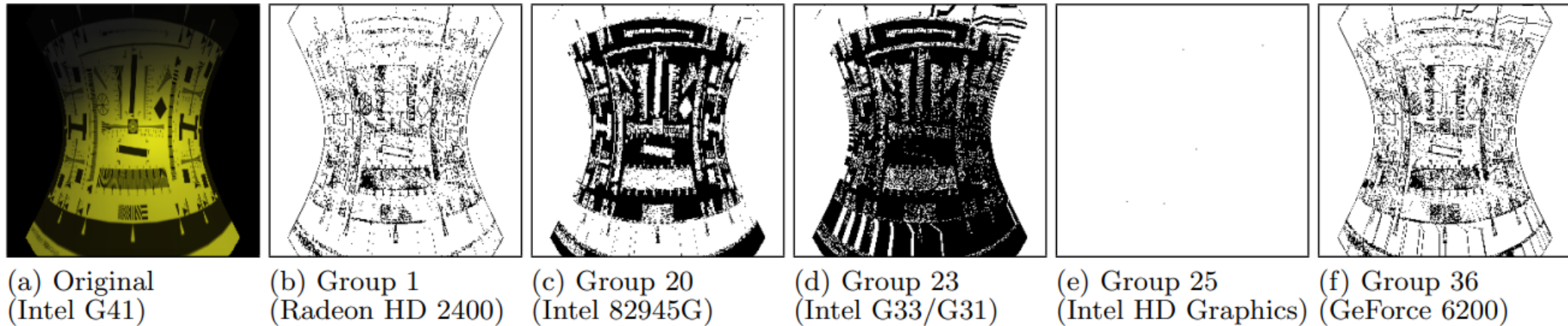


Figure 10: Original render and difference maps for Group 24

Mowery and Shacham, 2012

And its out there!



Figure 4: Different images printed to canvas by fingerprinting scripts. Note that the phrase “*Cwm fjordbank glyphs vext quiz*” in the top image is a *perfect pangram*, that is, it contains all the letters of the English alphabet only once to maximize diversity of the outcomes with the shortest possible string.

COVER YOUR TRACKS

See how trackers view your browser

[Learn](#)

[About](#)

HOW TO READ YOUR REPORT

You will see a summary of your overall tracking protection. The first section gives you a general idea of what your browser configuration is blocking (or not blocking). Below that is a list of specific browser characteristics in the format that a tracker would view them. We also provide descriptions of how they are incorporated into your fingerprint.

HOW CAN TRACKERS TRACK YOU?

Trackers use a variety of methods to identify and track users. Most often, this includes tracking cookies, but it can also include browser fingerprinting. Fingerprinting is a sneakier way to track users and makes it harder for users to regain control of their browsers. This report measures how easily trackers might be able to fingerprint your browser.

HOW CAN I USE MY RESULTS TO BE MORE ANONYMOUS?

Knowing how identifiable you are, or whether you are blocking trackers, can help you take steps to better protect your privacy. Browser add-ons or protection mechanisms built into the browser can help. Even so, the sneakiest trackers have ways around even the strongest security.

Here are your Cover Your Tracks results. They include an overview of how visible you are to trackers, with an index (and glossary) of all the metrics we measure below.

Our tests indicate that you have strong protection against Web tracking.

IS YOUR BROWSER:

| | |
|---|---|
| Blocking tracking ads? | <u>Yes</u> |
| Blocking invisible trackers? | <u>Yes</u> |
| Protecting you from <u>fingerprinting</u> ? | <u>Your browser has a nearly-unique fingerprint</u> |

Still wondering how fingerprinting works?

[LEARN MORE](#)

Note: because tracking and p

Your R

One in 145,235 browsers have the same fingerprint

measure all forms of

Within our dataset of several hundred thousand visitors tested in the past 45 days, only **one in 145,235 browsers have the same fingerprint as yours.**

Fingerprinting as a security measure

- Blocking bots (e.g. reCAPTCHA)
- Validating users over-time

How should we view tracking and fingerprinting efforts?

“Privacy preserving” personalized ads aka FLoC

- <https://github.com/WICG/turtledove>
 - The browser, not the advertiser, holds the information about what the advertiser thinks a person is interested in.
 - Advertisers can serve ads based on an interest, but cannot combine that interest with other information about the person — in particular, with who they are or what page they are visiting.
 - Web sites the person visits, and the ad networks those sites use, cannot learn about their visitors' ad interests.

“Privacy preserving” personalized ads aka Topics

- <https://github.com/patcg-individual-drafts/topics>
 - The browser, not the advertiser, holds the information about what the advertiser thinks a person is interested in.
 - Advertisers can serve ads based on an interest, but cannot combine that interest with other information about the person — in particular, with who they are or what page they are visiting.
 - Web sites the person visits, and the ad networks those sites use, cannot learn about their visitors' ad interests.

<https://support.google.com/google-ads/answer/11899856?hl=en>

“Privacy preserving” personalized ads aka ~~FLEDGE~~ Protected Audience API

- <https://github.com/WICG/turtledove/blob/main/FLEDGE.md>
- FLEDGE has been renamed to Protected Audience API.
- First Experiment (FLEDGE)
 - This document describes an early prototype for ads serving in the TURTLEDOVE family, appropriate for experimentation before a fully-featured system is ready. It would be the First "Locally-Executed Decision over Groups" Experiment.
 - This first experiment is currently a Chrome Origin Trial. The goal is for us to gain implementer experience, and for the ads ecosystem to evaluate its usability, as soon as it is feasible to do so. We need a robust API to take flight before the removal of third-party cookies shown on Chrome's Privacy Sandbox timeline.

Privacy is far more than web tracking

- We've only started talking about it, in only 1 context.