

# 17. Tracking and Trust on the Web



Blase Ur and David Cash  
February 19<sup>th</sup>, 2021  
CMSC 23200 / 33250



THE UNIVERSITY OF  
CHICAGO

# 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 & Platform Games

Android OS

Banking

Beaches & Islands

Bollywood & South Asian Film

Business & Productivity Software

Action & Adventure Films

Adventure Games

Autos & Vehicles

Bars, Clubs & Nightlife

Blues

Books & Literature

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

# Online Tracking

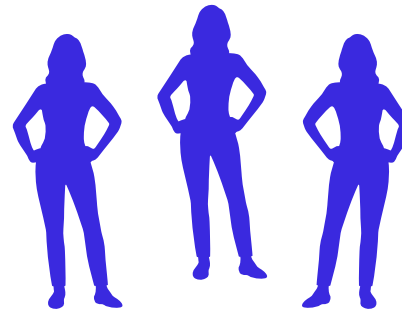
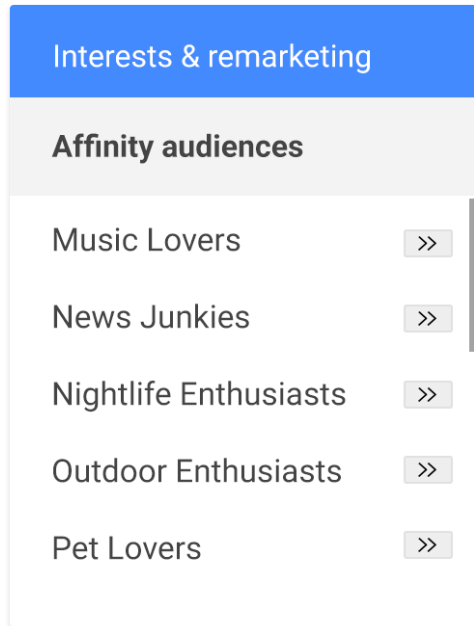
- First party = the site you are visiting (whose address is in the URL bar)
- Third party = other sites contacted as a result of your visit to that site
- First-party tracking (e.g., for search)
  - Consider DuckDuckGo and alternatives

# Data-Driven Inferences

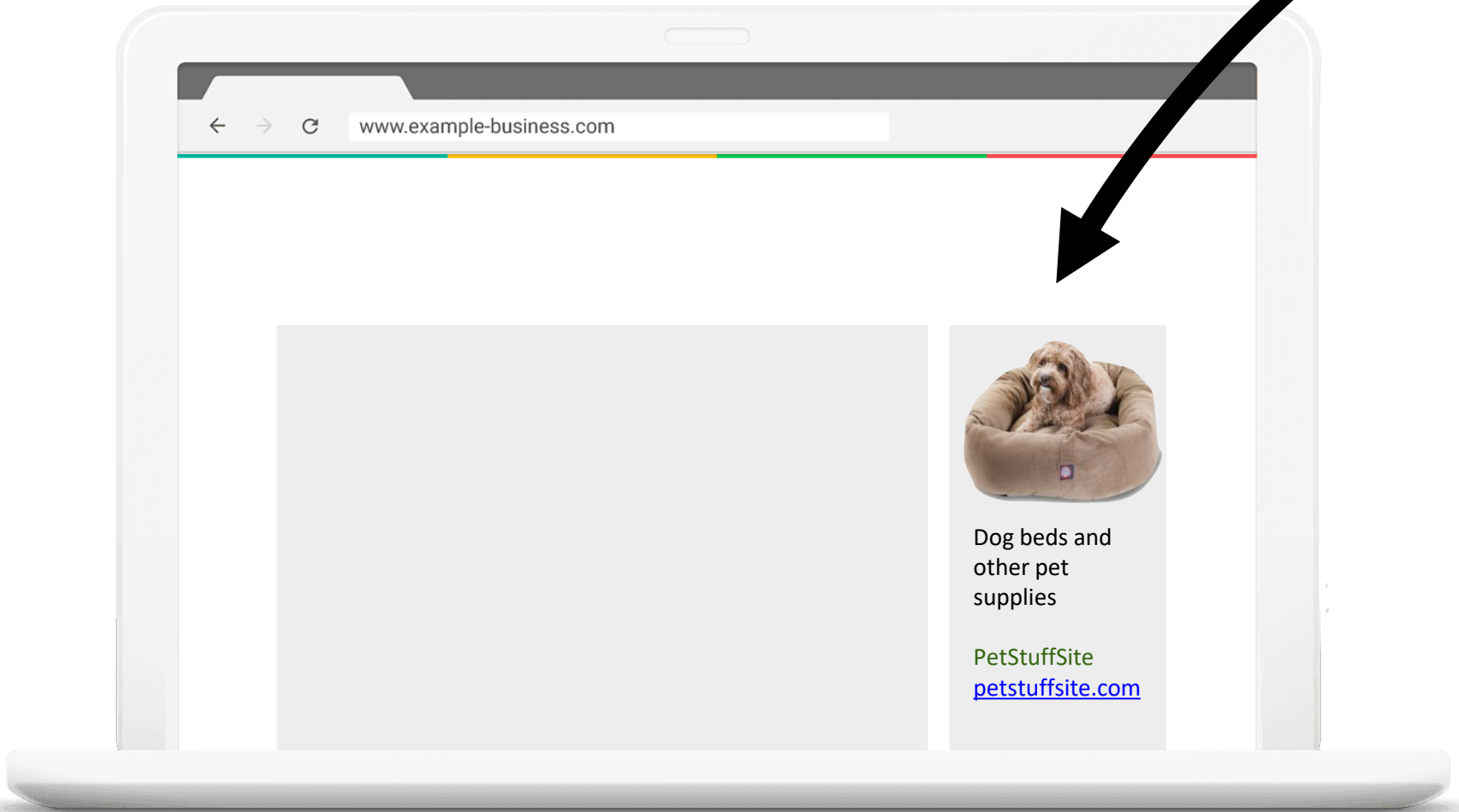
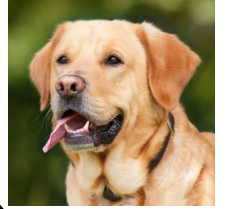


You might like dogs!

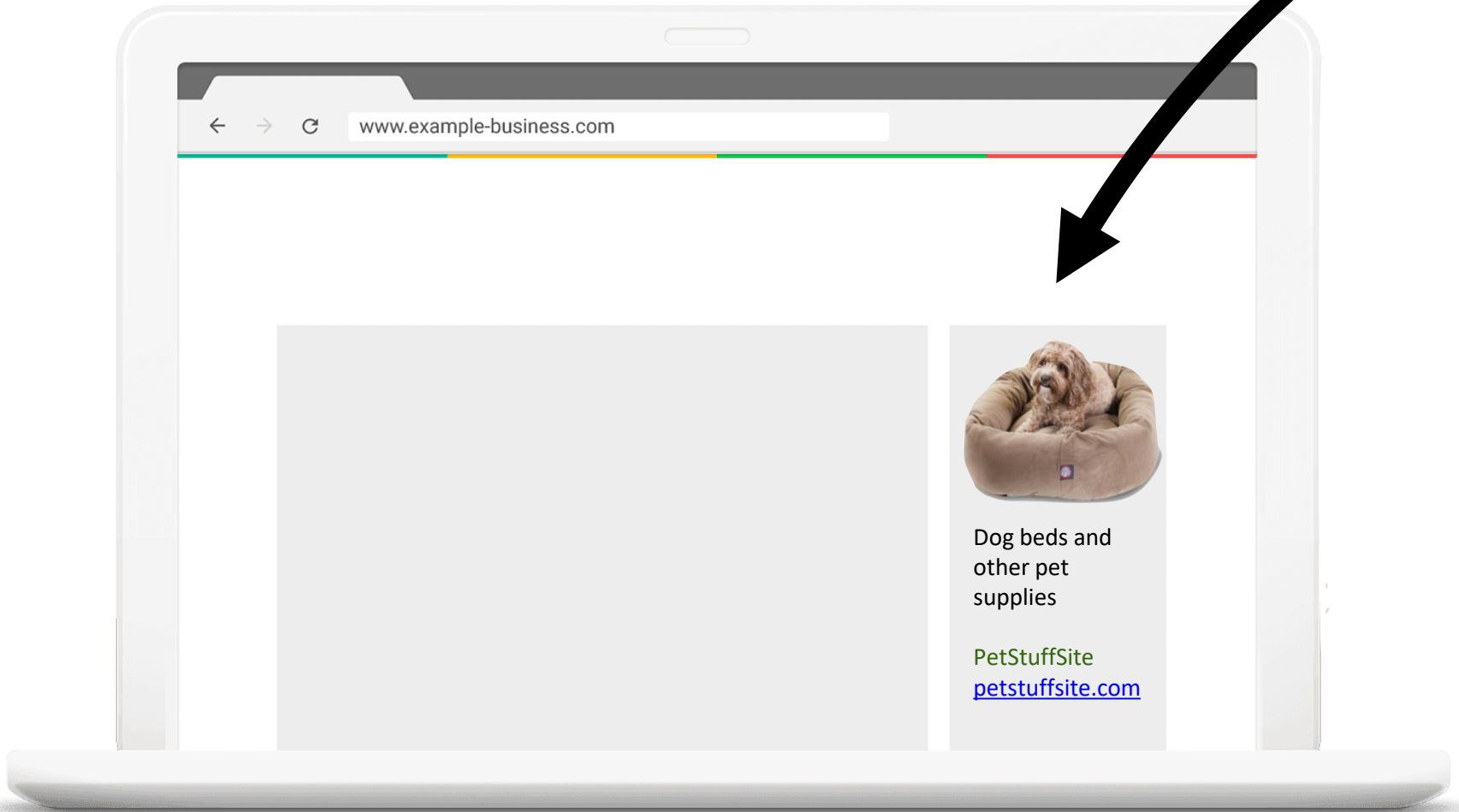
# Targeted Advertising



# Targeted Advertising



# Targeted Advertising



# Mechanics of Tracking

- Canonically, tracking is accomplished via HTTP cookies
  - Third-party cookies



# Online Tracking

- JavaScript / images from advertising networks loaded as part of your page
  - In iframes
  - Or sometimes not
  - Why does this matter?
  - Does this also apply to email? (Yes)

# Ubiquity of Online Tracking

The screenshot displays the Ghostery browser extension interface. On the left, a summary panel shows '22 Trackers found on www.mynews.com' with a donut chart indicating '4 Blocked'. Below this are buttons for 'Trust Site', 'Restrict Site', and 'Pause Ghostery', along with a link to 'Map these trackers'. The right panel, titled 'TRACKERS', lists individual trackers with checkboxes to block them. A 'Block All' checkbox is at the top right of this panel.

TRACKERS		Block All <input type="checkbox"/>
<b>Advertising</b> 10 TRACKERS 3 Blocked		
Advertising.com	<input checked="" type="checkbox"/>	
DoubleClick	<input type="checkbox"/>	
Google Adsense	<input type="checkbox"/>	
Korrelate	<input type="checkbox"/>	
Moot	<input checked="" type="checkbox"/>	
NetRatings Site Center	<input checked="" type="checkbox"/>	
Polar Mobile	<input checked="" type="checkbox"/>	
ScoreCard Research Beacon	<input type="checkbox"/>	
Tacoda	<input type="checkbox"/>	

# (My Group's) Tracking Transparency

The screenshot shows a Firefox browser window with three tabs: 'Debugging with Firefox Developer', 'Tracking Transparency', and 'The New York Times - Breaking'. The address bar shows 'https://www.nytimes.com'. The page displays the New York Times homepage with the date 'Friday, March 16, 2018'. A red notification box from the 'Tracking Transparency' extension is overlaid on the right side of the page. The notification text reads: 'On The New York Times - Breaking ..., there are 5 trackers. One of these trackers is Google, which knows about your activity on this page and 3829 others. In total, 169 trackers have seen you visit 12313 pages. The Tracking Transparency extension has determined that these companies could have inferred your interest in 162 topics.' Below the notification is a button that says 'Show me more about what the trackers know'. The background page shows the New York Times masthead, navigation links (World, U.S., Politics, N.Y., Business, Opinion, Tech, Science, Health, Sports), and article teasers including 'Trump's Steel Tariffs Open Lobbying Floodgate' and 'Controlled' by Daniel McCarthy.

Tracking Transparency

On The New York Times - Breaking ..., there are 5 trackers.

One of these trackers is **Google**, which knows about your activity on this page and **3829** others.

In total, *169 trackers* have seen you visit *12313 pages*. The Tracking Transparency extension has determined that these companies could have inferred your interest in *162 topics*.

Show me more about what the trackers know

Trump's Steel Tariffs Open Lobbying Floodgate

By ANA SWANSON and KENNETH P.

Controlled

By DANIEL MCCARTHY


The duty of the president's team is not to second guess him or the outcome of the last

Elite

By SUSAN JACOBY


The energy that elites spend being ashamed of their advantages would be

# (My Group's) Tracking Transparency

[Tracking Transparency](#) [Interests](#) [Trackers](#) [Sites](#) [Activity](#) [About](#) 


Home

## What are *trackers* and *interests*?






When you browse online, your online activity can be tracked by ad networks and analytics companies.

We call these *trackers*.

→ 

These companies track your browsing to make guesses about what topics you might be interested in.

We call these topics *interests*.

→   


Companies can personalize your online experience based on these interests. *Click on the circles above to learn more.*


### Your Top Trackers


- Google
- Chartbeat
- Optimizely
- Microsoft
- Amazon.com

### Your Top Interests

- Law & Government
- Online Communities
- People & Society
- News
- Shopping

39  
 Trackers encountered

7  
 Pages visited

6  
 Potential interests

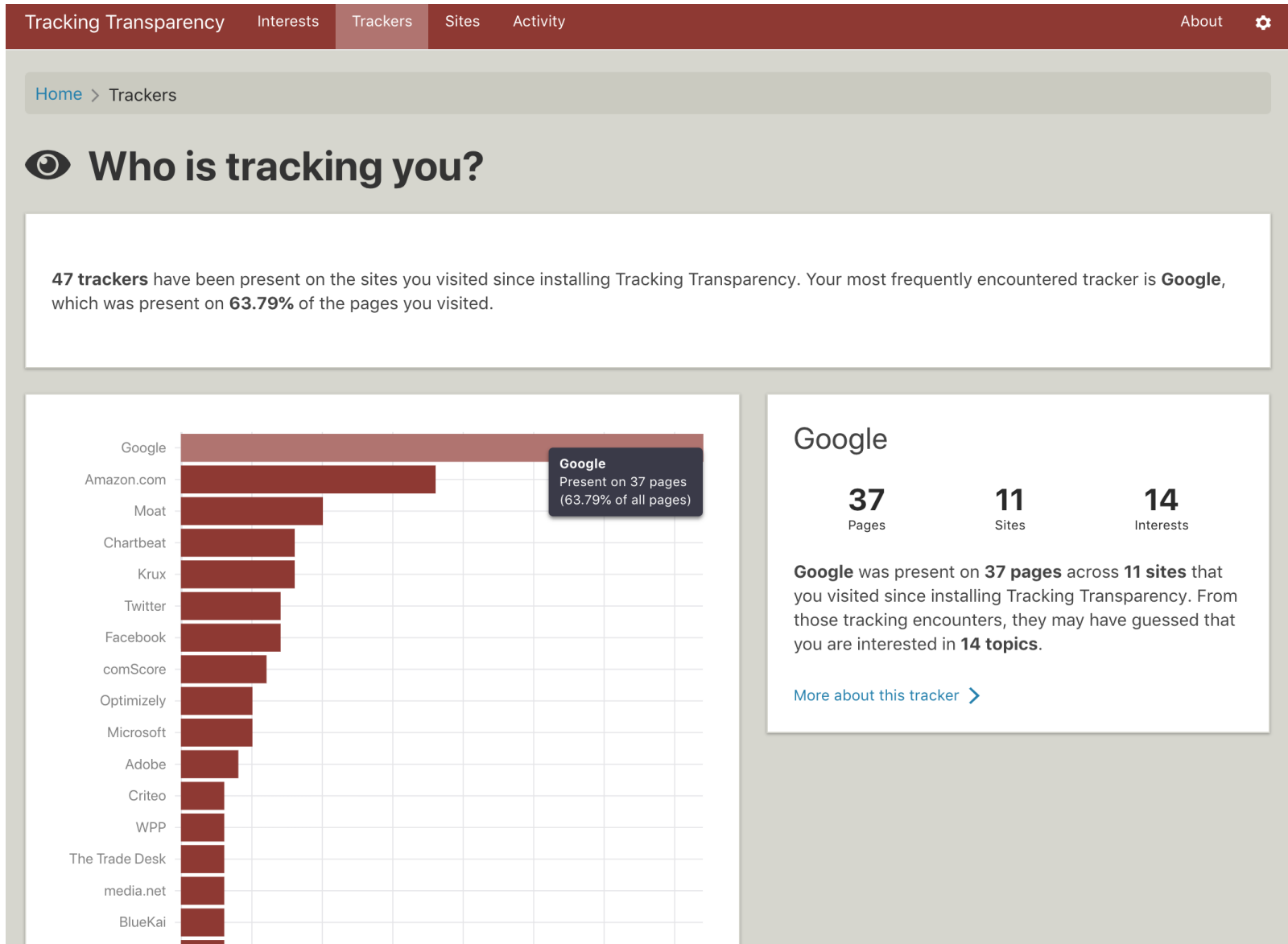
#### Recent Interests

- Law & Government
- Computers & Electronics
- Shopping
- News
- People & Society

#### Recent Sites

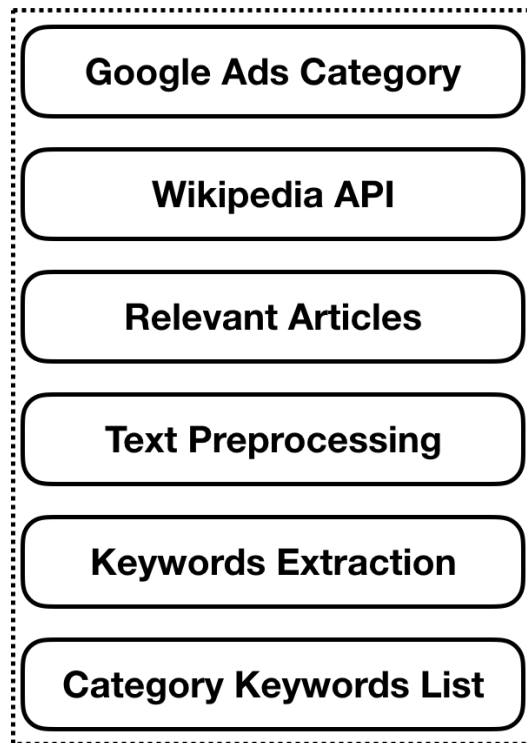
- twitter.com
- eff.org
- amazon.com
- cnn.com
- nytimes.com

# (My Group's) Tracking Transparency

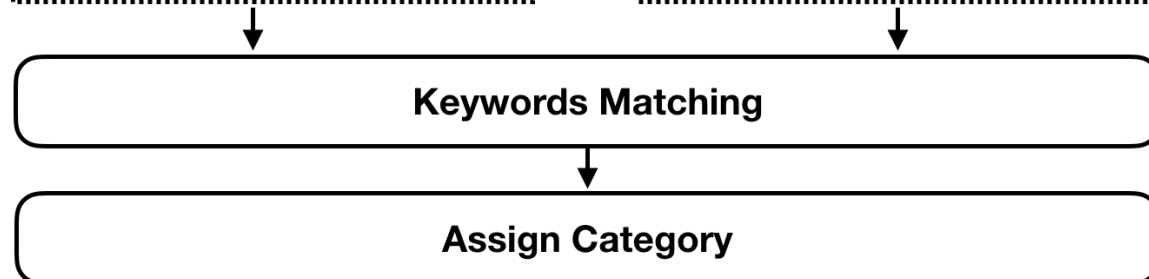
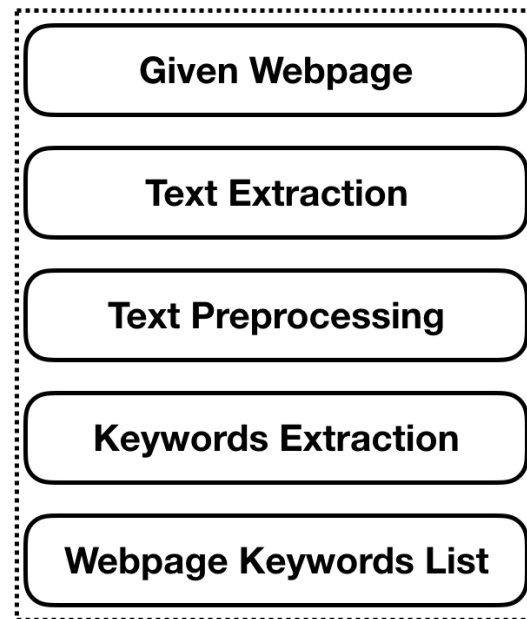


# (My Group's) Tracking Transparency

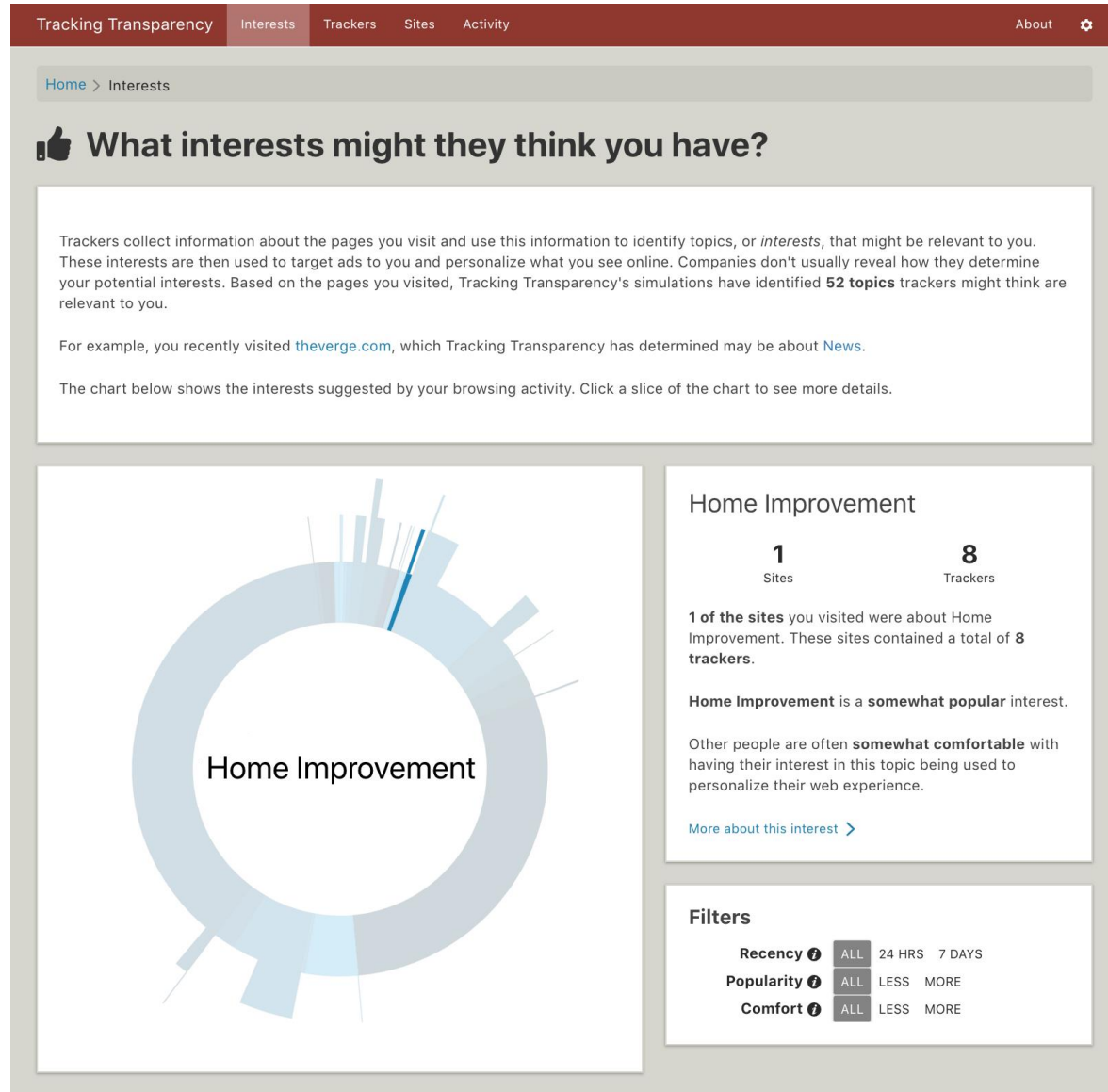
## 1) Categories keywords extraction



## 2) Webpage keywords extraction



# (My Group's) Tracking Transparency






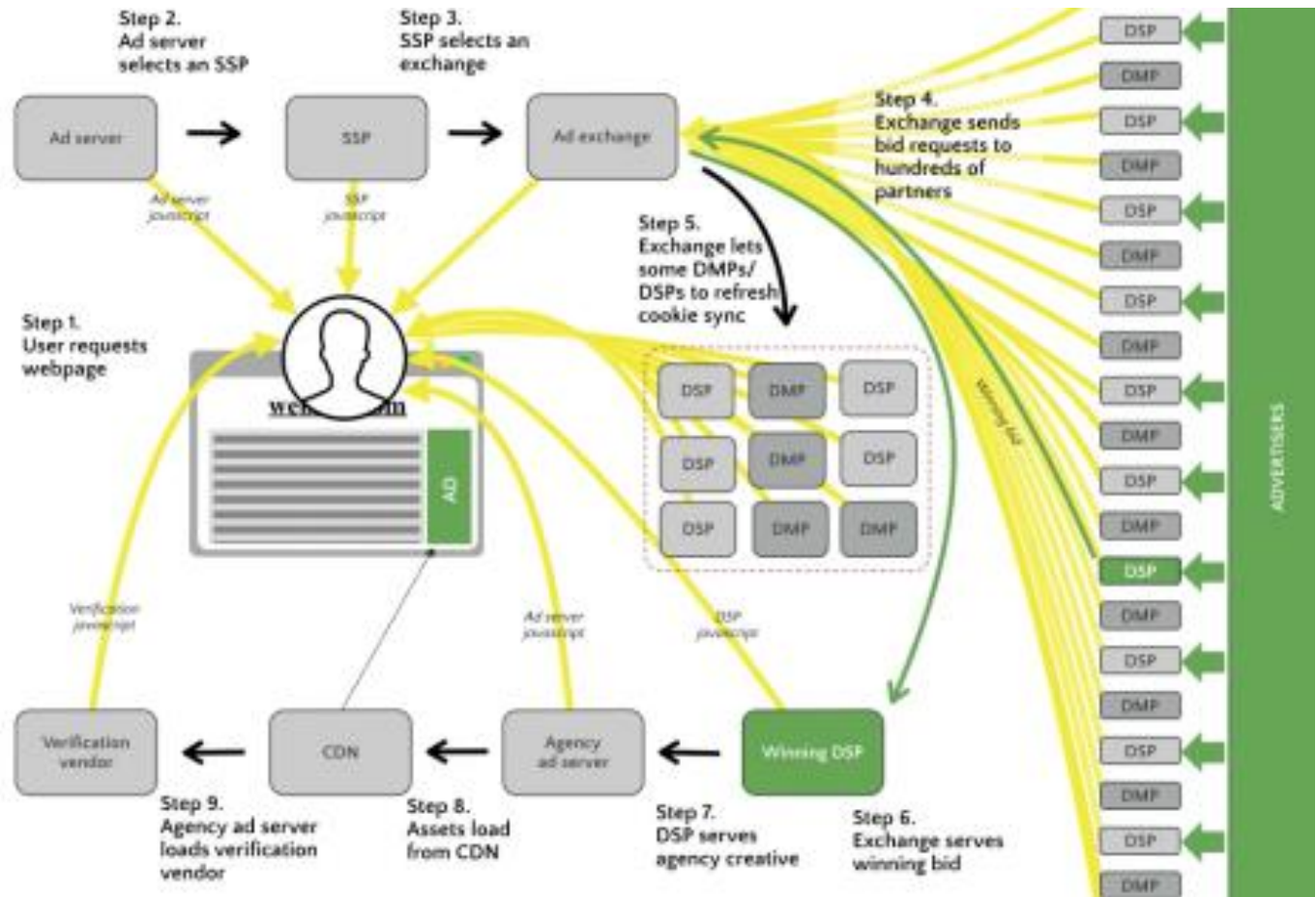
# Ad Bidding Marketplaces

## DATA LEAKAGE IN ONLINE ADVERTISING

This is the current process of real-time bidding that is used in online behavioural advertising.

### Legend

-  Channel of data leakage
-  Money
-  Personally identifiable information





# Existing Privacy Tools

The Disconnect browser extension interface is shown. At the top, it says "DISCONNECT" with "Help" and "Share" links. Below this are social media sharing buttons for Facebook (0), Google+ (1), and Twitter (1). The main section lists categories of trackers blocked:

- Advertising**: 2 requests. Includes Adobe (1 request) and Nielsen (1 request).
- Analytics**: 7 requests.
- Social**: 0 requests.
- Content**: 0 requests.




At the bottom, there are options to "Whitelist site" and "Visualize page", checkboxes for "Show counter" and "Cap counter", and a bar chart showing "Time saved" and "Bandwidth saved". A green button at the bottom says "Get Mobile Protection".

The Blur browser extension interface is shown on the ESPN website. At the top, it says "espn.com" with a close button and "8 trackers blocked". Below this, it says "Tracker blocking is on for this website". A list of blocked trackers is shown:




- Google AdSense blocked
- Demdex blocked
- Twitter Badge blocked
- Omniure blocked

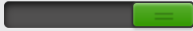


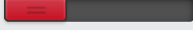
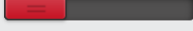

Below the list is a link: "see your tracker blocking stats and learn more about these companies". At the bottom, a blue bar says "21 trackers blocked since Feb '17". Below this is a link: "Correct how Blur works in the form below". At the very bottom, the "oBLUR" logo is shown, along with links for "Settings", "Help", and "Go Premium".

# Existing Privacy Tools

 **Privacy Badger**  

Privacy Badger detected 45 potential **trackers** on this page. These sliders let you control how Privacy Badger handles each one. You shouldn't need to adjust them unless something is broken.



  

weather.api.cnn.io	
rtax.criteo.com	
ad.doubleclick.net	
googleads.g.doubleclick.net	
securepubads.g.doubleclick.net	
connect.facebook.net	


[Disable Privacy Badger for This Site](#)

[Did Privacy Badger break this site? Let us know!](#)

[Donate to EFF](#)

 **GHOSTERY** 

15 Trackers found on [www.cnn.com](http://www.cnn.com)
















14 Blocked

[Trust Site](#)

[Restrict Site](#)

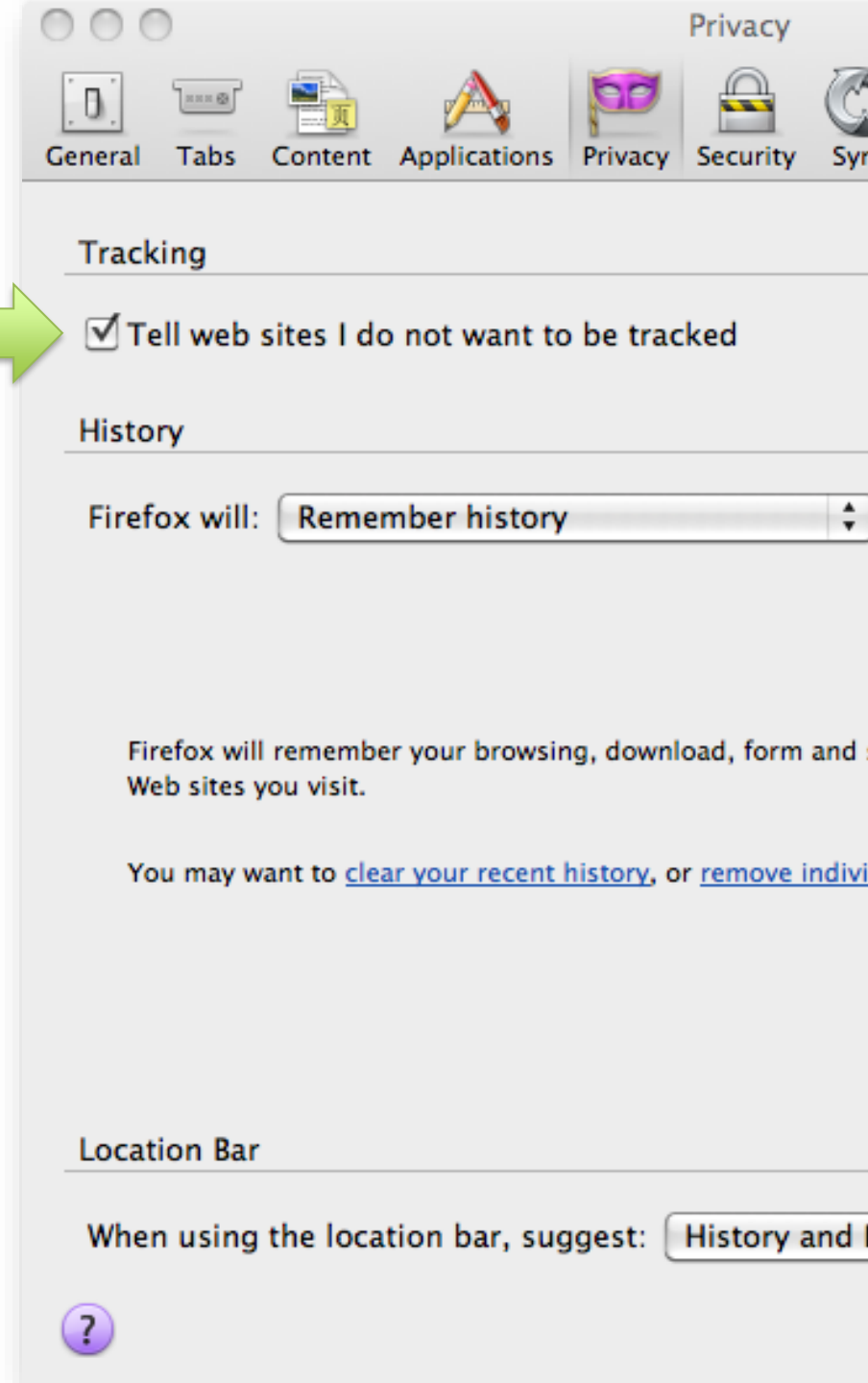
[Pause Ghostery](#)

[Map These Trackers](#)

Trackers	Block All 
 <b>Advertising</b> 10 Trackers 10 Blocked	
<a href="#">Amazon Associates</a>	
<a href="#">ChartBeat</a>	
<a href="#">Criteo</a>	
<a href="#">DoubleClick</a>	
<a href="#">Google Publisher Tags</a>	
<a href="#">KruX Digital</a>	
<a href="#">NetRatings SiteCensus</a>	
<a href="#">Outbrain</a>	
<a href="#">Rubicon</a>	
<a href="#">ShareThrough</a>	
 <b>Site Analytics</b> 2 Trackers 2 Blocked	

# Do not track

- Proposed W3C standard
- User checks a box
- Browser sends “do not track” header to website
- Website stops “tracking”
- W3C working group trying to define what that means



# Tools to stop tracking, effective?

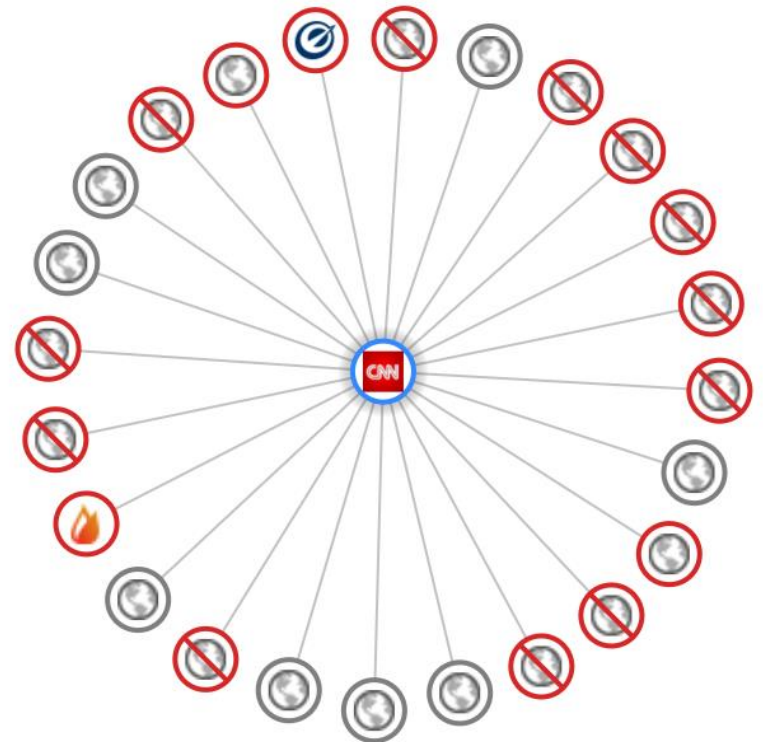
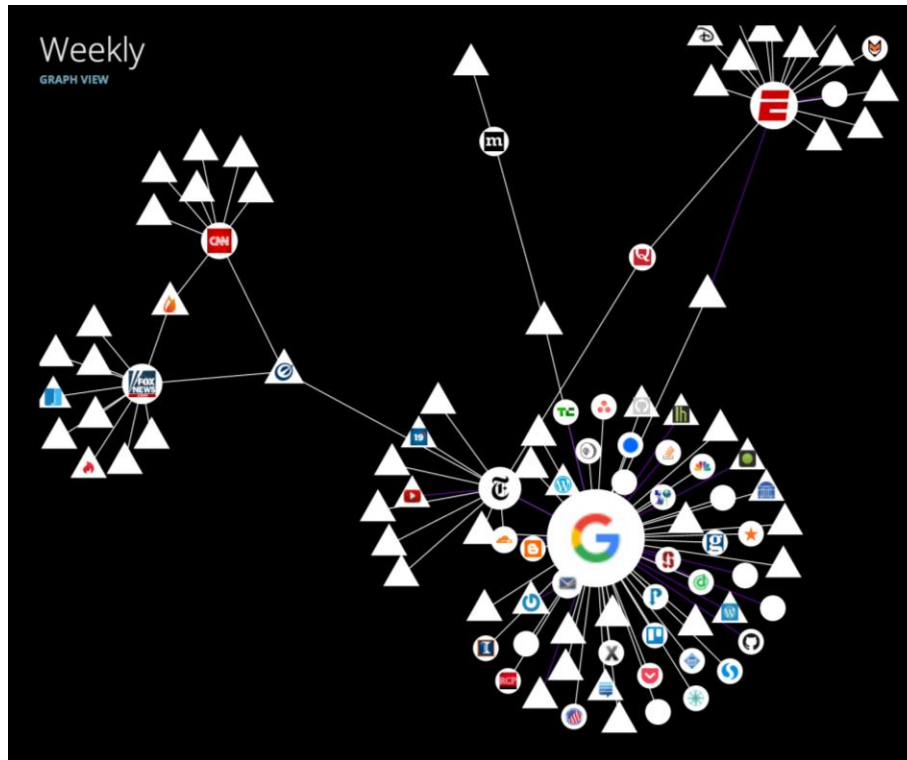
- Browser privacy settings
  - Cookie blocking
  - P3P
  - Tracking Protection Lists
  - Do Not Track
- Browser add-ons
- Opt-out cookies
- Digital Advertising Alliance (DAA) AdChoices icon and associated opt-out pages



DoNotTrackMe



# Visualization: Connection Graphs

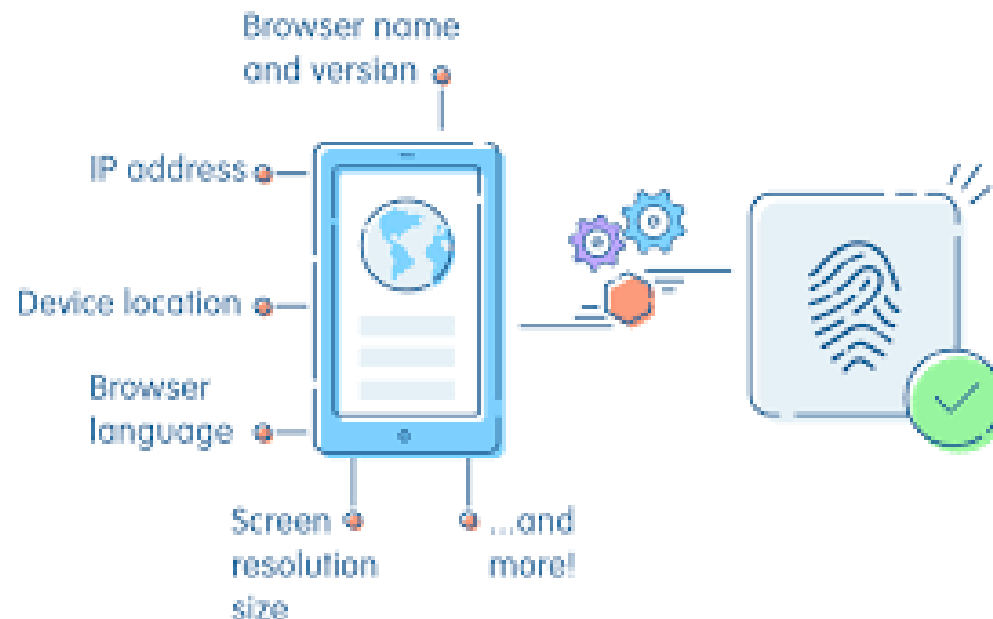


# Browser fingerprinting

- 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?)*

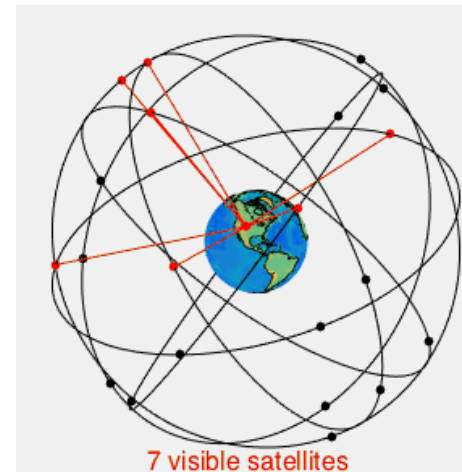
# Device Fingerprinting

- Use unique(-ish) combination of device features as an identifier
- <https://panopticklick.eff.org/>



# Location Tracking

- IP Geolocation
  - Hierarchy of IP addresses
- GPS (Global Positioning System)
  - ~31 satellites in semi-synchronous orbit in OUTER SPACE with atomic clocks
  - Always ~6 satellites in line of sight
  - Multilateration





# What Does HTTPS Hide? (Ghost)

- Body of the HTTP request / response is hidden
- ...So what's left to be seen / inferred?

# Side Channels

- Using metadata or outside observations to make inferences about the data



# Web Side Channels Include:

- Size of packets
  - How can this reveal what pages you are visiting?
- Timing

## Remote Timing Attacks are Practical

David Brumley  
*Stanford University*  
dbrumley@cs.stanford.edu

Dan Boneh  
*Stanford University*  
dabo@cs.stanford.edu

### Abstract

Timing attacks are usually used to attack weak computing devices such as smartcards. We show that timing attacks apply to general software systems. Specifically, we devise a timing attack against OpenSSL. Our experiments show that we can extract private keys from an OpenSSL-based web server running on a machine in the local network. Our results demonstrate that timing attacks against network servers are practical and therefore security engineers should defend against them.

The attacking machine and the server were in different buildings with three routers and multiple switches between them. With this setup we were able to extract the SSL private key from common SSL applications such as a web server (Apache+mod\_SSL) and a SSL-tunnel.

**Interprocess.** We successfully mounted the attack between two processes running on the same machine. A hosting center that hosts two domains on the same machine might give management access to the admins of each domain. Since both domains are hosted on the same machine, one admin could use

# Web Side Channels Include:

- Color

- [link one](#)
- [second link](#)
- [link three \(visited\)](#)
- [fourth link](#)