17. Tracking and Trust on the Web

Blase Ur and David Cash
February 19th, 2021
CMSC 23200 / 33250
Online Tracking

- Advertisers want to show you advertisements targeted to your interests and demographics
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

Here Are the Most Dog-Friendly Restaurants in the United States

You might like dogs!
Targeted Advertising

Interests & remarketing

Affinity audiences

Music Lovers
News Junkies
Nightlife Enthusiasts
Outdoor Enthusiasts
Pet Lovers

Dog
Guitar
Three women silhouettes
Targeted Advertising

Dog beds and other pet supplies

PetStuffSite
petstuffsite.com
Targeted Advertising

Dog beds and other pet supplies

PetStuffSite
petstuffsite.com
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
(My Group’s) 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
(My Group’s) Tracking Transparency

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
1. Google
2. Chartbeat
3. Optimizely
4. Microsoft
5. Amazon.com

Your Top Interests
1. Law & Government
2. Online Communities
3. People & Society
4. News
5. 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

Who is tracking you?

47 trackers have been present on the sites you visited since installing Tracking Transparency. Your most frequently encountered tracker is Google, which was present on 63.79% of the pages you visited.

Google

37 Pages
11 Sites
14 Interests

Google was present on 37 pages across 11 sites that you visited since installing Tracking Transparency. From those tracking encounters, they may have guessed that you are interested in 14 topics.
(My Group’s) Tracking Transparency

1) Categories keywords extraction
   - Google Ads Category
   - Wikipedia API
   - Relevant Articles
   - Text Preprocessing
   - Keywords Extraction
   - Category Keywords List

2) Webpage keywords extraction
   - Given Webpage
   - Text Extraction
   - Text Preprocessing
   - Keywords Extraction
   - Webpage Keywords List

Keywords Matching
Assign Category
What interests might they think you have?

Trackers collect information about the pages you visit and use this information to identify topics, or interests, that might be relevant to you. These interests are then used to target ads to you and personalize what you see online. Companies don’t usually reveal how they determine your potential interests. Based on the pages you visited, Tracking Transparency's simulations have identified 52 topics trackers might think are relevant to you.

For example, you recently visited theverge.com, which Tracking Transparency has determined may be about News.

The chart below shows the interests suggested by your browsing activity. Click a slice of the chart to see more details.

---

Home Improvement

1 Sites
8 Trackers

1 of the sites you visited were about Home Improvement. These sites contained a total of 8 trackers.

Home Improvement is a somewhat popular interest.

Other people are often somewhat comfortable with having their interest in this topic being used to personalize their web experience.

More about this interest

---

Filters

- Recency
- Popularity
- Comfort

- ALL
- 24 HRS
- 7 DAYS
- LESS
- MORE
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

PageFair
Existing Privacy Tools
Existing Privacy Tools

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.

Ghostery found 15 trackers on www.cnn.com. 14 of them were blocked.

Options include:
- Trust Site
- Restrict Site
- Pause Ghostery

Site Analytics:
- 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
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://panopticlick.eff.org/](https://panopticlick.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 motivate the need for faster attacks.

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 domain are hosted on the same machine, an attacker could...
Web Side Channels Include:

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