CSE 127 Computer Security

Stefan Savage, Spring 2020, Lecture 18

Privacy, Law and Ethics

Quick topics

- Final on Monday
  - We're going to do similar to the midterm
  - 14 hour window, online quiz on canvas
  - It will mainly focus on material since the midterm, but may also include a few questions from the whole quarter
  - Subject matter from today's lecture will not be on the final

- Things are crazy right now

Today

- Arming you as technologists to both protect yourselves and to "do the right thing"

- Privacy
  - Commercial and Government monitoring
  - What can one do?

- Ethical and (US) legal issues that abound in security

Privacy

- We talked a bunch in class about *confidentiality* (how to keep information secret in a system/protocol)
- Privacy is about how information that isn't secret is controlled
  - It is a much thornier problem…
- Today there is a huge amount of information gathered about you
  - Far far far more than you realize
  - Lets start with the commercial side of this…

Elements of the technical issue

- What data is being captured?
- What identifiers are tracked?
- What information can be inferred?

Data captured about you

- What Web sites you visit?
- What magazines/newspapers you subscribe to?
- Organizations you're in?
- Credit report, income, bank balance, property owned?
- Where you've lived?
- Political donations?
- Marital status, criminal/civil actions, adoption status?
- Where you are/have been?
- What you buy, where you've flown, where you drive?
- Did you vote?
- What you're watching on TV? What music you listen to?
- Who your friends are?
- What you post online? (pretty much every service gets scraped)
- Etc…

Identifiers that tie you to this data

- IP address
- Cookies
- App accounts
- Account logins (esp FB, Google, etc used for federated authentication)
- Phone number/ESN
- VIN/License plate numbers
- Built-in tracking in your television/cable box
- Visa/MC/Check number
- Name
- Address
- E-mail addresses
- Characteristics of your computer/phone
- Facial characteristics (via recognition)
- SSN, Passport number, etc…

Information Inference

- Correlation of data by identifier
  - Flight to Newark, Car Rental, Hotel in Manhattan, tickets purchased for Broadway show -> **vacation to NYC**
- Classification via aggregation
  - Unknown characteristics may be *probabilistically revealed* based on other features (what sites you visit, what you buy, who friends are)
    - E.g., gender, race, ethnic background, sexual orientation, income, weight, health status, religion, beliefs on social issues, level of education, etc
    - This works incredibly well if you don't care about perfect accuracy

Single biggest driver?

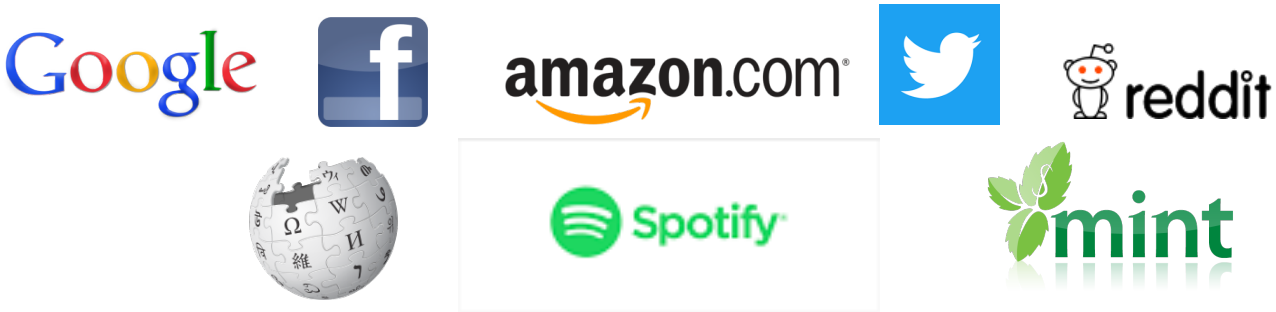
- Advertising

- Goal: target message to convince you to take some action (e.g., buy a car, vote for a candidate)

- Lets briefly look at one common domain:
  Web advertising and cookies

Context

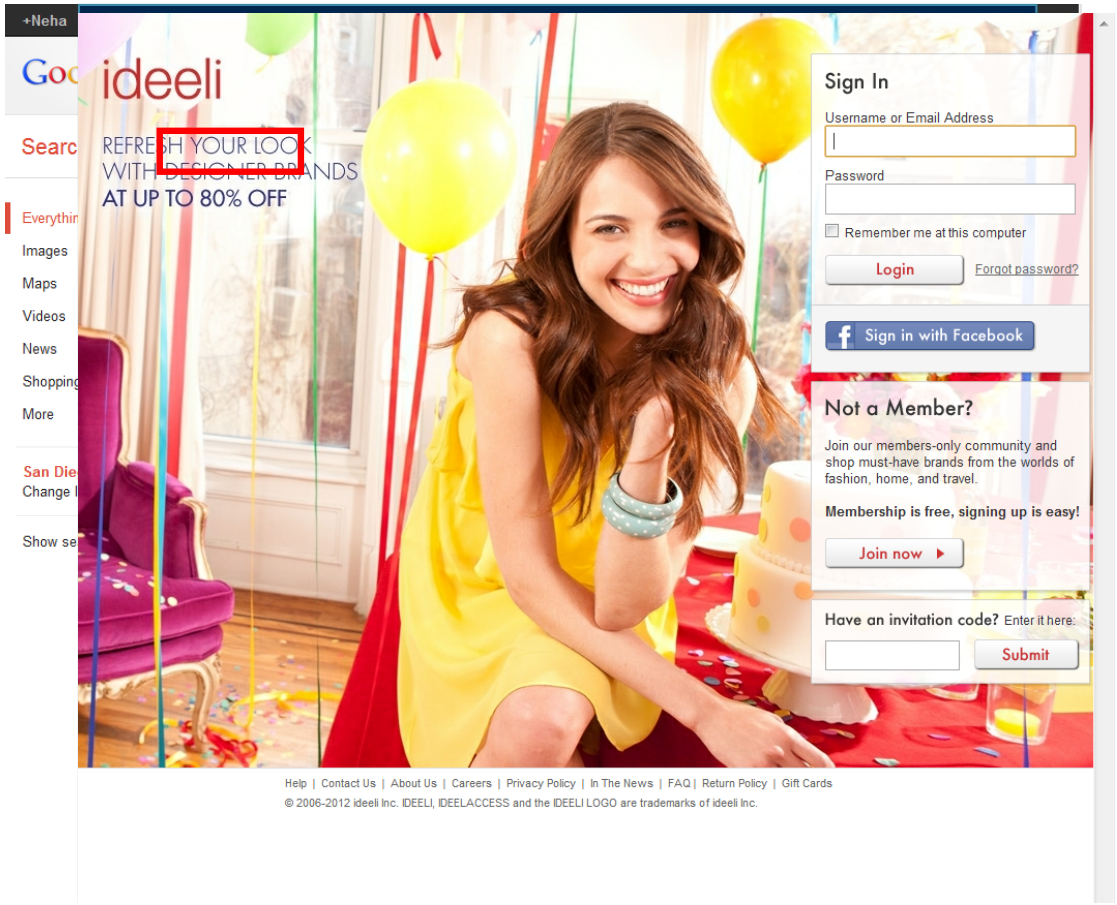
- Web has been immensely successful



- Majority of the content is free!
- Companies make money through ads
  - Revenue estimated at $125 billion in 2019 [iab.net 2020]

How to increase profit?

Increasing revenue



Increasing revenue

- For ad provider:
  - Show most relevant ads
  - Limited viewing time
  - Limited space
- Track users for interests, demographics, etc.

- For content providers:
  - Improve or personalize content for more traffic
  - More viewers per ad
- Track what users like to see on a site

- This form of tracking is called *behavioral tracking*

Behavioral Tracking

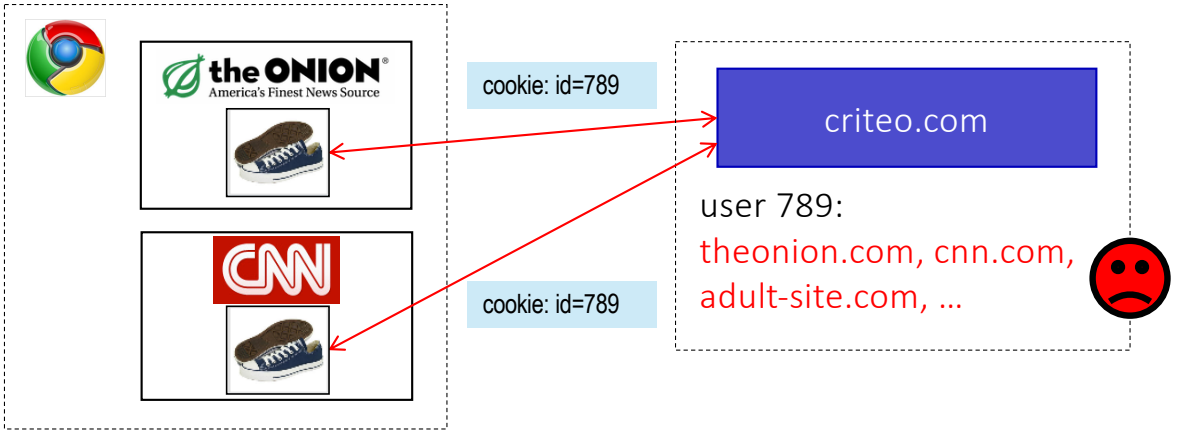
- Trend is to aggregate vast amounts of user information
  - Location information, different interests, sexual orientation, gender, favorite bands, colors…
  - Via references to outside tracking sites (sometimes called trackers)
    - Either only for purpose of tracking,
    - Or broker ads and also track user

- Consolidation of tracking sites
  - Provide ads at scale
  - Top tracking 10 sites could track users across 70% sites in 2008 [Krishnamurthy 2009]. Today it is even higher.
  - A few organizations know most of where you go on the Web

Ad Ecosystem

“Anonymous” Tracking

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

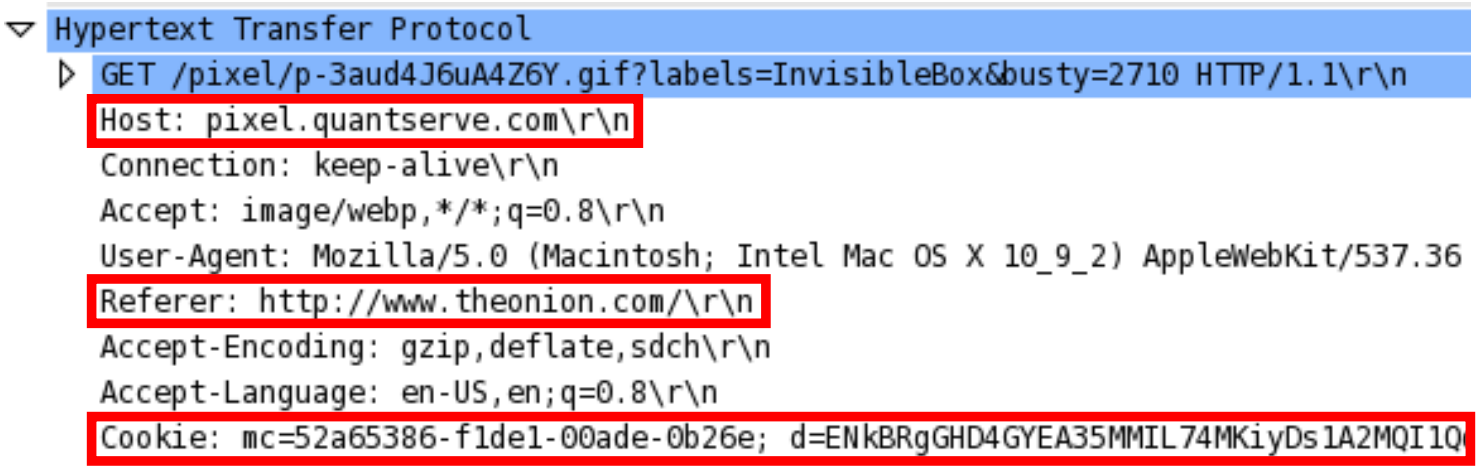


- Called anonymous because cookies don’t identify who you are, but any external data (e.g., login, contact e-mail) can make that association; also cookie syncing allows trackers to share data

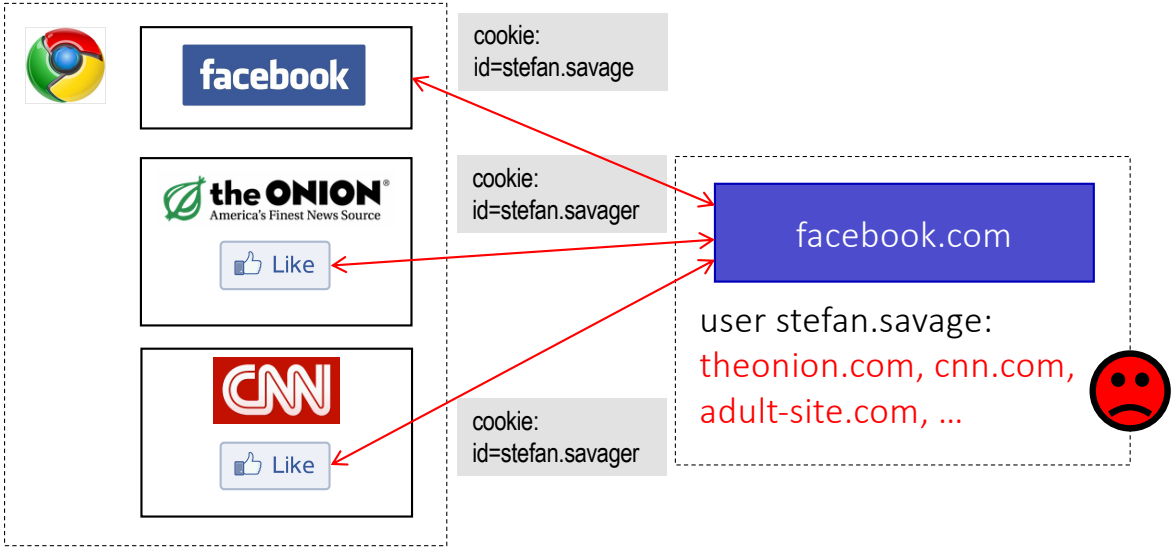
Slides courtesy Franzi Roesner

Basic Tracking Mechanisms

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



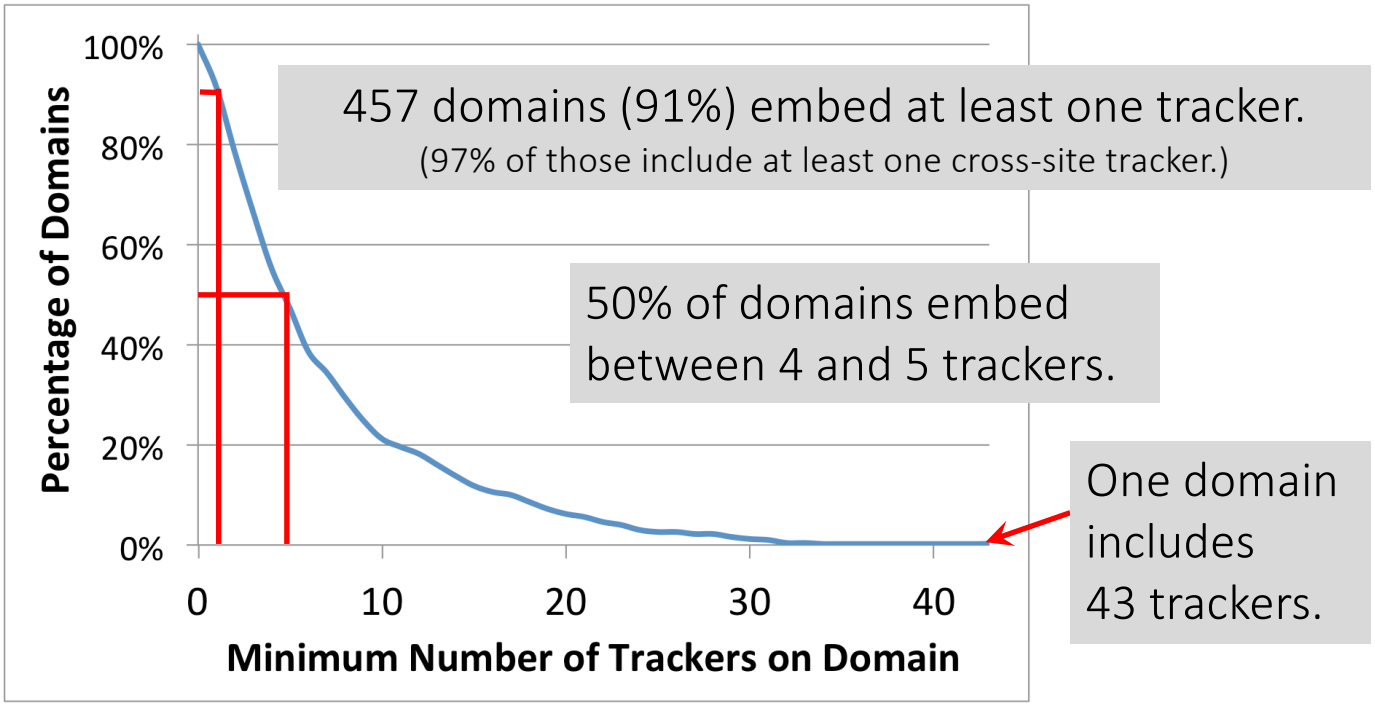
Personal Tracking



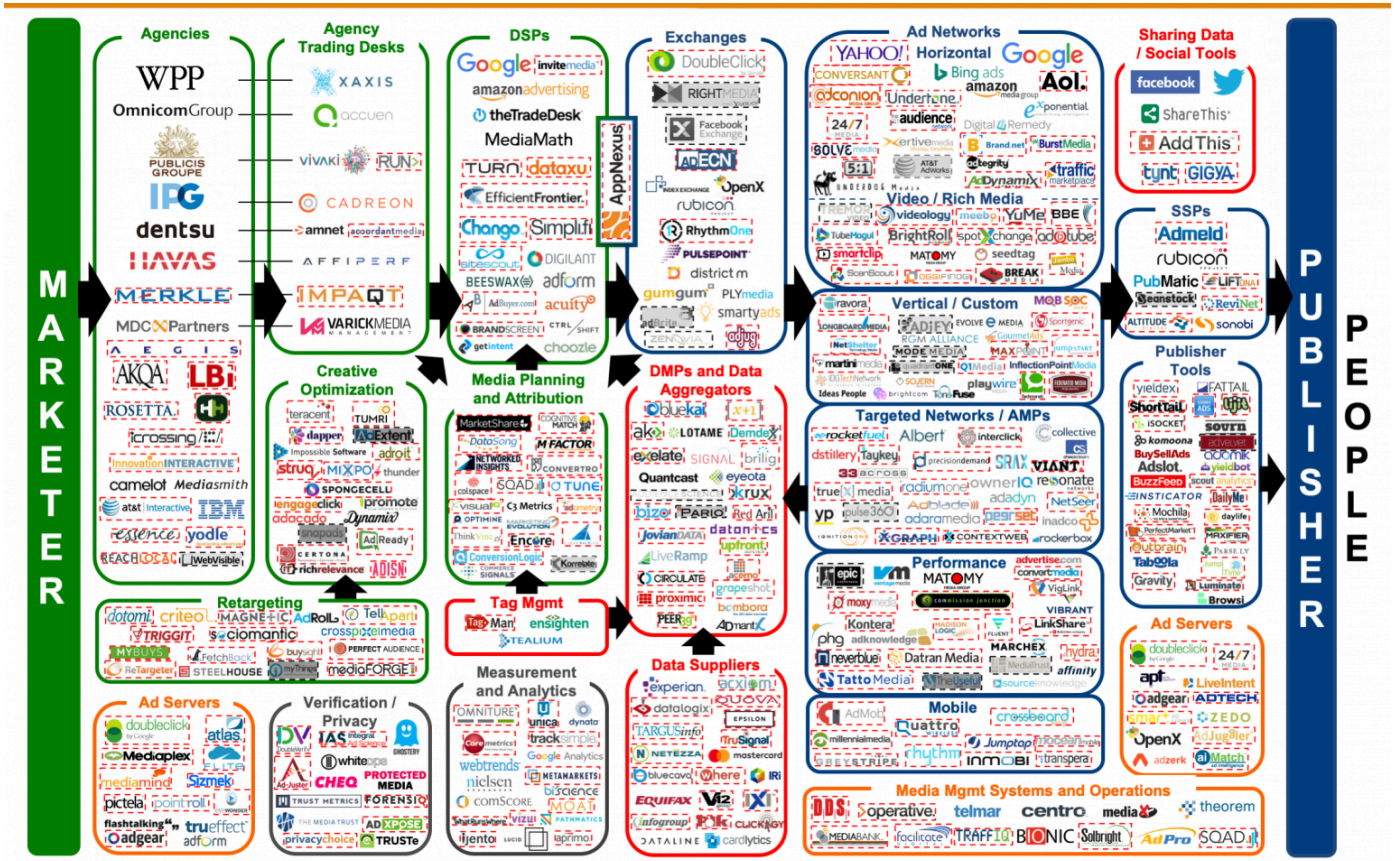
- Tracking is not anonymous (linked to accounts).
- Users directly visit tracker's site → evades some defenses.

How prevalent is tracking?

- 524 unique trackers on Alexa top 500 websites
  (Rosener et al, NSDI '14)



It's a big ecosystem: Display Advertising Landscape



Can block cookies but…

- Lots of other ways to track users
  - IP address
  - Various persistent objects in HTML5
  - Device fingerprints (Canvas, WebRTC, AudioContext, Battery)
  - Tracking codes in URLs
  - Beacons
  - Etc

- Its pretty crazy out there (remember that picture of the ad market)

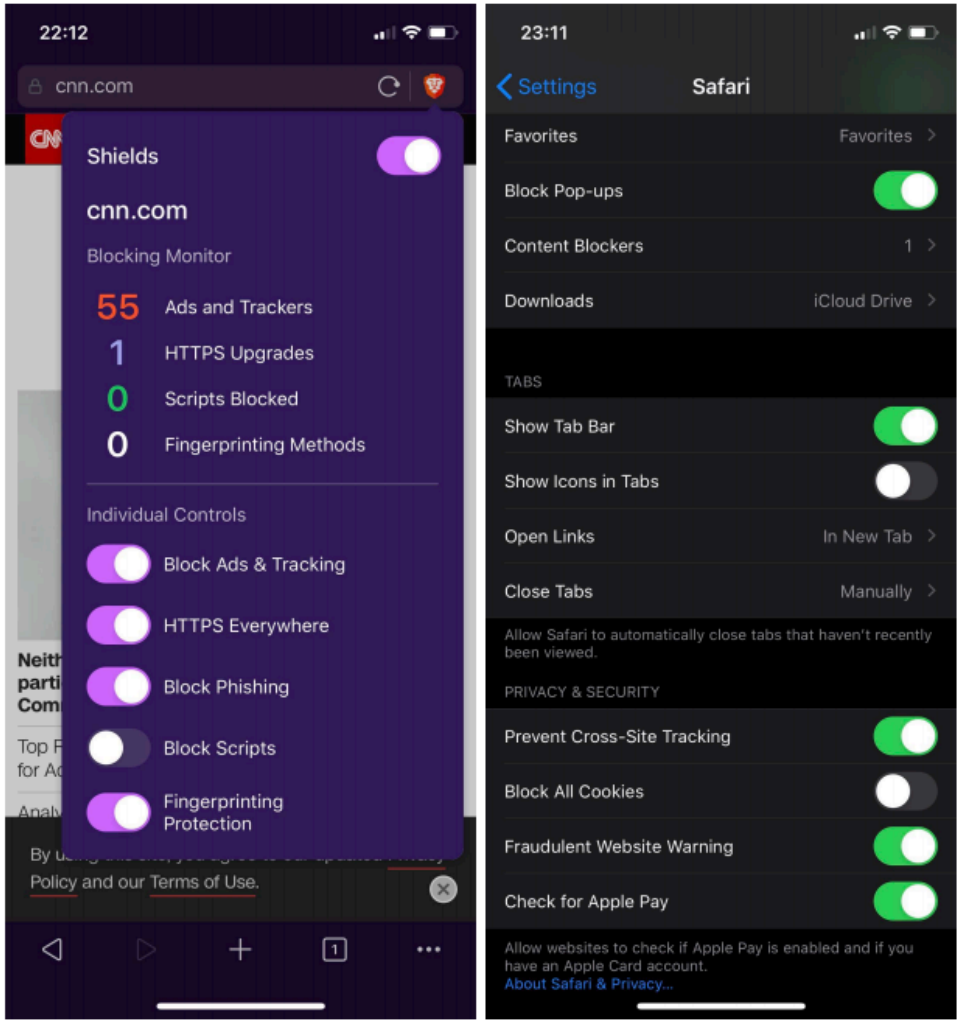
Example: History Sniffing

- Unintended consequence of combination of three independently desirable features
  - Visited-link indication to the user
  - JavaScript monitoring of page rendering
- Malicious site renders URLs of interest of screen and checks their color using Javascript

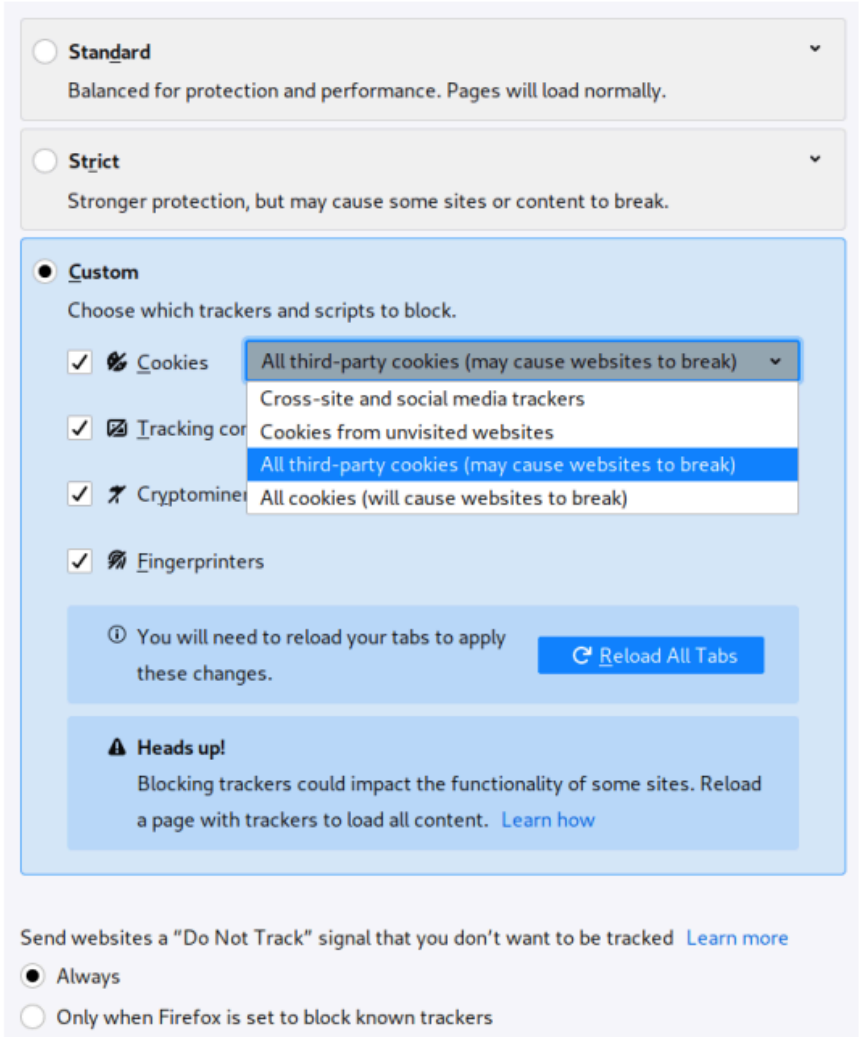
Things you can do

- Don't use Google/Facebook for third-party site login
  (always have a per-site login; ideally with unique password)
- Don't use third-party gadgets (e.g., facebook "like" buttons)
- Under California law (CCPA) you can request what information each company has about you, who they have sold it to, you can ask them not to sell it going forwards and to delete it (time consuming)
- Some browsers better about privacy than others
  (e.g., Brave, Safari, Tor Browser)
- Privacy-oriented browser extensions (tracker blockers)

Privacy-enhanced browsing (Brave & Safari)

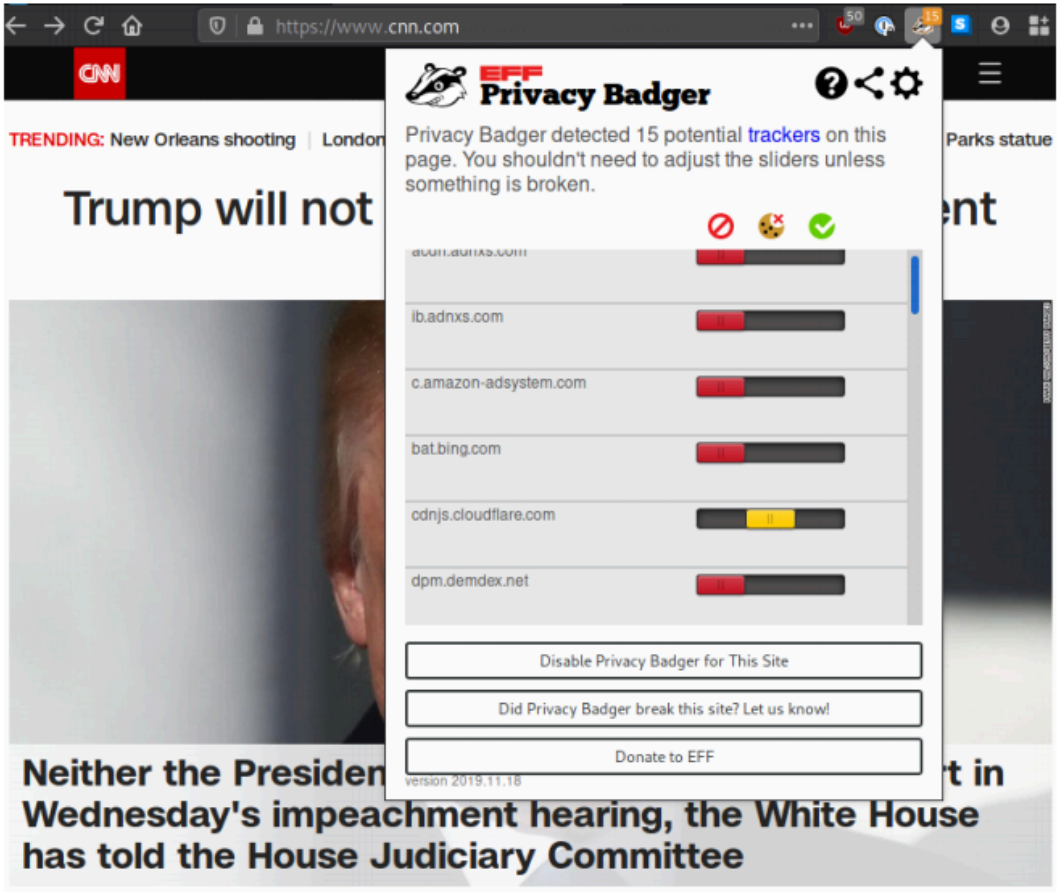


Privacy-enhanced browsing (Firefox)



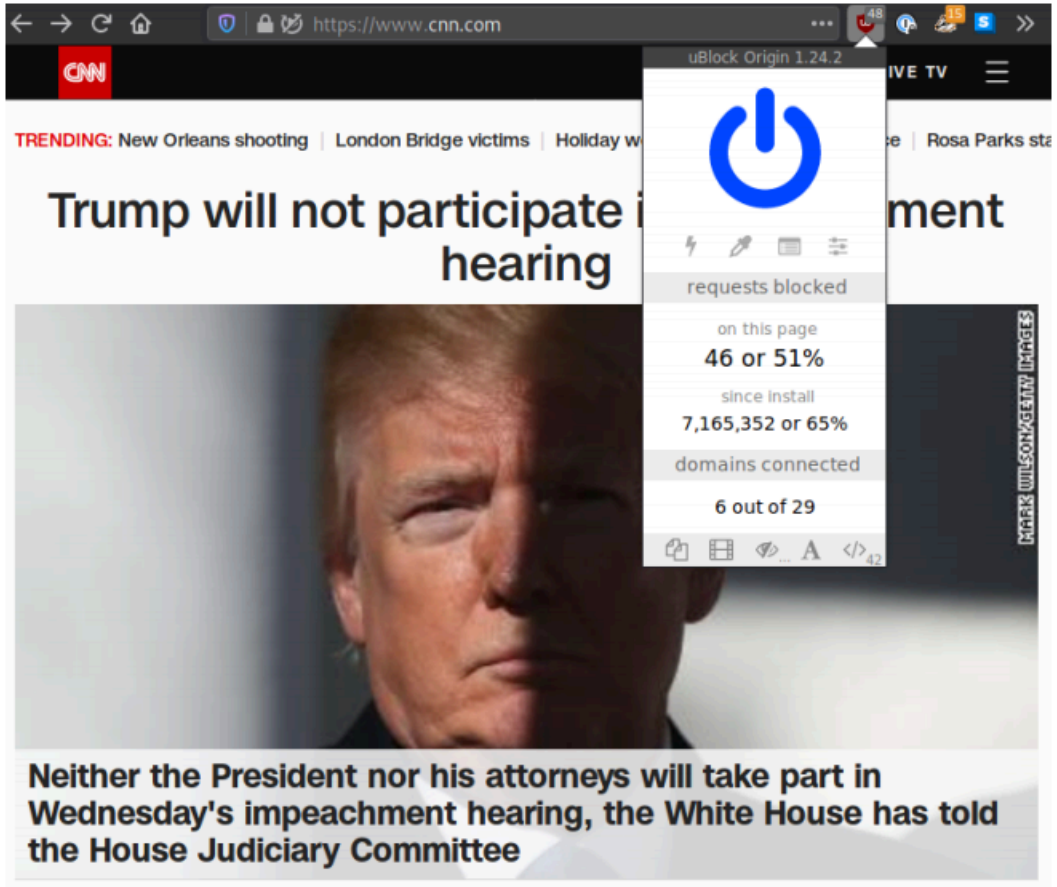
Privacy-enchaning extensions

- Privacy Badger blocks trackers; uBlock Origin blocks ads; many others



Privacy-enchaning extensions

- Privacy Badger blocks trackers; uBlock Origin blocks ads; many others



Quick mention about ToR

- The Tor project is dedicated to providing *individual anonymity*
  - For lots of reasons, but particularly motivated by concerns about state surveillance
  - Tor Browser is one piece (special modified version of FireFox), but the more important service provided is address anonymity
- The Onion Router (TOR) – anonymous Internet access
  - Large numbers of volunteer systems participate in TOR, agreeing to route, accept and deliver traffic
  - Client picks three nodes at random and routes your TCP traffic through them
  - Onion encryption (you encrypt data with key of last node, then add header identifying last router, encrypt with key of second node, etc…)
  - Only the last node sees the data and none of them know the whole path
  - Effective, but slow

Government Interests and Privacy

- HTTPS is pretty good, but what about the endpoints?
  (i.e., what about your mail at Google, post on Instagram, etc)

- What if government takes an interest?
  (e.g., as part of a criminal investigation)
  - With appropriate legal process all US companies will turn over your data
  - In a range of circumstances you might not be told (at least for a while)
  - You have limited standing to object (3rd party doctrine)

Kinds of criminal process for obtaining data

- Grand jury subpoena
  - Business records, some limited meta-data/subscriber information, etc – basic subscriber info (3rd party doctrine)
- Pen register (part of Electronic Communications Privacy Act – '86)
  - Prospective "dialing, routing, addressing, or signaling information" (i.e., metadata)
  - Standard: "information likely to be obtained … is relevant to an ongoing criminal investigation
- 2703(d) (also part of ECPA)
  - Historic (non-content) [requires "specific and articulable facts"]
- Warrants
  - Search/Seizure & Tracking (Rule 41)
  - ECPA (2703) warrant
    - **RCS/ECS only (unique privacy rights)**
    - Content, probable cause
- Wiretap (Title III of Crime Control and Safe Streets Act – '68)

Rights to privacy?

- 4th amendment
  - The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.
- Griswold v Connecticut (1965)
  - Douglas – "penumbras" and "emanations" of other constitutional protections
- Katz v United States (1967)
  - What about a phone booth?
  - "Reasonable expectation of privacy" std

Technological change & the 4th amendment

- What is a "**reasonable expectation** of privacy"?

- Kylo v US (2001): FBI use of Thermal Imaging
  - 'when . . . the Government uses a device that *is not in general public use*, to explore details of the home that would previously have been unknowable without physical intrusion, the surveillance is a "search" and is presumptively unreasonable without a warrant.'
- GPS trackers (Jones) (2012)
- Cell phones SITA (Riley) (2014)
- Cell site simulators (e.g. Stingrays) (2017) [split]
- Cell site location info (Carpenter) (2018)

Quick aside: What's in a search warrant?

- Sworn affidavit establishing "probable cause"

- Particularity (Attachment B)
  - Where will the search take place?
  - What will be searched?

- Reasonableness

- In general, warrant must issue from district in which search takes place (some exceptions)

Private sector responsibilities

- Criminal process obligates producing data if you have it… but what if you don't have it?
  - Specific laws:
    - Communications Assistance for Law Enforcement Assistance Act (1994). CALEA
      - Compels telecoms and equip manufacturer to provide support for lawful intercept; highly controversial
      - But doesn't cover other service providers (e.g., messaging) or consumer products
    - Assistance provision to Wiretap Act – must help, but can charge also "unobtrusively and with a minimum of interference with the service"
  - All Writs Act – having exhausted other strategies court can compel private sector to render aid (NY Tel)
    - Point of contention about scope allowed

Ongoing tensions between tech companies and government

- Access to data at motion and at rest
  - CALEA does not mandate access to non-telephony systems for interception (e.g., WhatsApp, FB Messenger, iMessage, etc)
  - If company has data they are obligated to turn it over under warrant, but if they don't have the data they aren't (e.g., San Bernadino iPhone Case)
- Law enforcement use of exploits (e.g., Playpen case)
- Use of new identification technology
  - DNA (e.g., Golden State Killer)
  - Face recognition (e.g., ClearView AI)
- Issues of Jurisdiction and International Clouds (e.g., US v Microsoft)
  - Cloud Act

Switching gears: legal/ethical issues you should think about

- Things you should know about… anyone doing real security research has a good lawyer in their contact list
- Reverse Engineering
- Vulnerability discovery/disclosure
- Thinking about Ethics in your work…

Reverse engineering

- Baseline
  - Bonito Boats v Thunder Boats "an essential part of innovation"
  - DVD copy Control Association v Andrew Bunner; presumptively legal
- Copyright issues (protection vs fair use)
  - If you both reverse engineer a product and design a competitor then there may be a claim against illegal copying
  - Best practice: separate team for reverse engineering of spec vs designing product (e.g., Sega v Accolade, but doesn't protect you from patents)
- Contract issues
  - Anti-reverse engineering clauses in software/service licenses (enforceability unclear, but lots of case law support it)
- DMCA

Digital Millennium Copyright Act (DMCA) (17 USC 1201)

- Anti-circumvention clause
  - No person shall ***circumvent a technological measure*** that ***effectively controls access to a work*** protected under this title
  - Has been used expansively…
    - Felten v RIAA – HackSDMI contest aftermath (dropped)
    - US v Elcom (& Skylarov) – copy protections in Adobe eBook (not guilty)
    - US v Crispin -- xbox modding (dropped)

- But… Library of Congress rulemaking provides exceptions
  - 2010 Jailbreaking
  - 2016/2018 Security research

(i) Computer programs, where the circumvention is undertaken on a lawfully acquired device or machine on which the computer program operates, or is undertaken on a computer, computer system, or computer network on which the computer program operates with the authorization of the owner or operator of such computer, computer system, or computer network, solely for the purpose of good-faith security research and does not violate any applicable law, including without limitation the Computer Fraud and Abuse Act of 1986.

(ii) For purposes of this paragraph (b)(11), "good-faith security research" means accessing a computer program solely for purposes of good-faith testing, investigation, and/or correction of a security flaw or vulnerability, where such activity is carried out in an environment designed to avoid any harm to individuals or the public, and where the information derived from the activity is used primarily to promote the security or safety of the class of devices or machines on which the computer program operates, or those who use such devices or machines, and is not used or maintained in a manner that facilitates copyright infringement.

Vulnerability discovery legality

- It is **legal** to find vulnerabilities if it doesn't involve accessing systems without permission (CFAA)
  - If you buy it, you're probably good (modulo anti-reversing contract… but ebay)
  - Be careful with online services!
  - Also, some awkwardness with packet sniffing and Wiretap act

- It is **legal** to sell exploits of vulnerabilities, so long as you do not have knowledge that the buyer intends to use them for criminal purposes (otherwise could be conspiracy)

Computer Fraud and Abuse Act (18 USC 1030)

- Primary criminal anti-hacking statute in US
  - Complex law with many branches, but typically some form of: "Whoever intentionally accesses a computer ***without authorization*** or ***exceeds authorized access***, and …"
  - Penalties typically 1-5 years per instance, but can go as high as 20 for certain cases (plus fines)
- What counts?
  - Jailbreaking your phone?
  - Jailbreaking your friend's phone?
  - Finding a vulnerability in a piece of third-party software? Exploiting it?
  - Finding a vulnerability in an online service (e.g., Facebook)? Exploiting it?
  - Violating the Terms of Service of an online service? (US v Nosal, Oracle v Rimini)
  - Logging into your professor's computer with a password you sniffed
  - Logging into your professor's computer using your own password? (which works for some reason)

Wiretap Act

- Makes it illegal to tap someone's phone (unless police w/T3 court order)
  - What about packet sniffing? Nope, illegal… you cant sniff my Ethernet without permission
  - But lots of companies do this on their networks. Yes, but its *their* network (and there are some tricky details even then)
  - What about WiFi? The law has generally indicated that public mediums (e.g., radio) are exempt… but Google v Joffee is an unfortunate decision in the opposite direction
- State have equivalent laws, but differ around question of disclosures
  - NY: one party disclosure (as long as one person knows, then its ok)
  - California: all party disclosure (only legal to record a call if everyone knows)
    - This is why you get those "recorded for quality assurance purposes" messages
    - What does this mean wrt a network protocol (e.g., recording your computer talking to X)
    - We have no idea… its crazy

The ethics of vulnerability discovery

- Should we be doing this at all?
  - Pro/con?

- When do we do once we find a vulnerability?
  - Uncoordinated disclosure (i.e., full disclosure): Tell the world and the vendor together
    - Pressures vendors into developing/releasing fix quickly
    - Exposes customers to attack during the intervening period
  - Coordinated disclosure (i.e., responsible disclosure): Tell vendor first, let them develop patch, then tell the world
    - Lets vendors develop patch before vuln exposed to world
    - Vendor delay may leave customers vulnerable longer
  - No Disclosure – sell to highest bidder and make a buck

Some general thoughts about ethics and computer security

- When you do things, try to anticipate harm it might cause
  - Think about tradeoffs: good achieved vs harms incurred
    - Sometimes harm is inevitable, but you can strive to minimize it
  - E.g., what if you have a vulnerability that can't be fixed, but can hurt people?
  - E.g., location services that are push (here I am) vs pull (tell me about things here)
- Outcomes >> Process
  - What counts is what will happen, not only that you did it fairly and by the rules
  - E.g., fetishization of "consent" as a mechanism to deal with privacy problems
    - Yes, we monitor the content of all their mail, but they agreed in our 19 page EULA…
  - E.g., engagement-maximization algorithms in recommender systems (e.g., FB/YouTube)
    - Super strong evidence that it drives people towards extremist content (but not by design)

Finally…

- My list of personal security advice (in order):
  - Turn on two-factor authentication (at least for your e-mail and bank)
  - Don't reuse passwords if you can (esp for e-mail and bank)
  - Use a password manager
  - Invest in regular data backups (best defense against ransomware)
  - If you are lucky enough to acquire lots of assets then ask your financial institution to require in-person signature for any wire transfer from that account

- Finally, finally…. Thanks!
  - Good luck, be strong, stay safe!