Extra Lecture: Privacy on the Web
(History stealing, Fingerprinting, DNT, etc.)

Webonomics

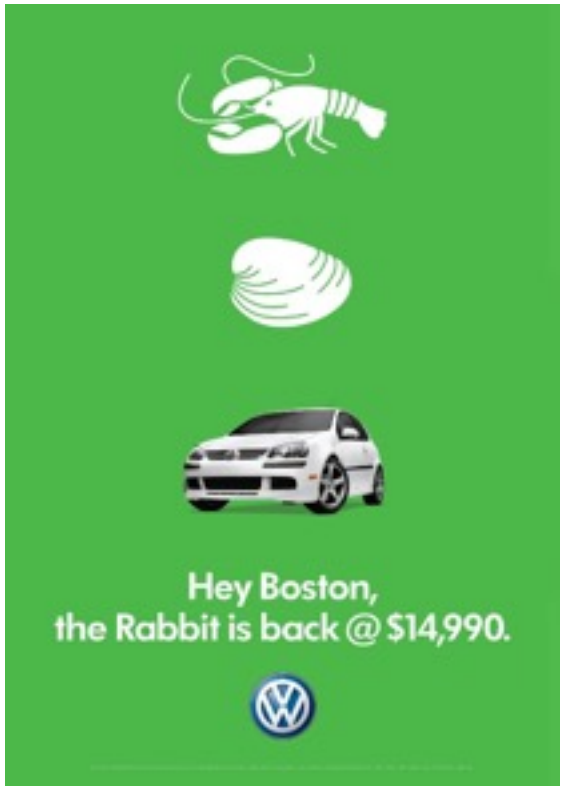
- The Web has allowed **free**, convenient services to proliferate
  - Google, Android
  - Facebook, Instagram
  - Millions of smartphone apps
- Who pays for the costs of all these services?
  - You do.
  - Not in cash, but in personal information

“If you are not paying for it, you're not the customer; you're the product being sold” – Andrew Lewis, 2002

Advertising on the Web

- By and large, advertising provides the money for web services and apps
  - 90% of Google's $6 Billion in revenue came from ads in 2014

Pre-Web advertising → Branding



Web advertising → Targeting

Your Personal Information is Valuable

How is this information collected, shared, and used for targeted advertising?

Tracking

Cookies, Flash Cookies, E-tags, Evercookies, Supercookies!

DNT

Fingerprinting

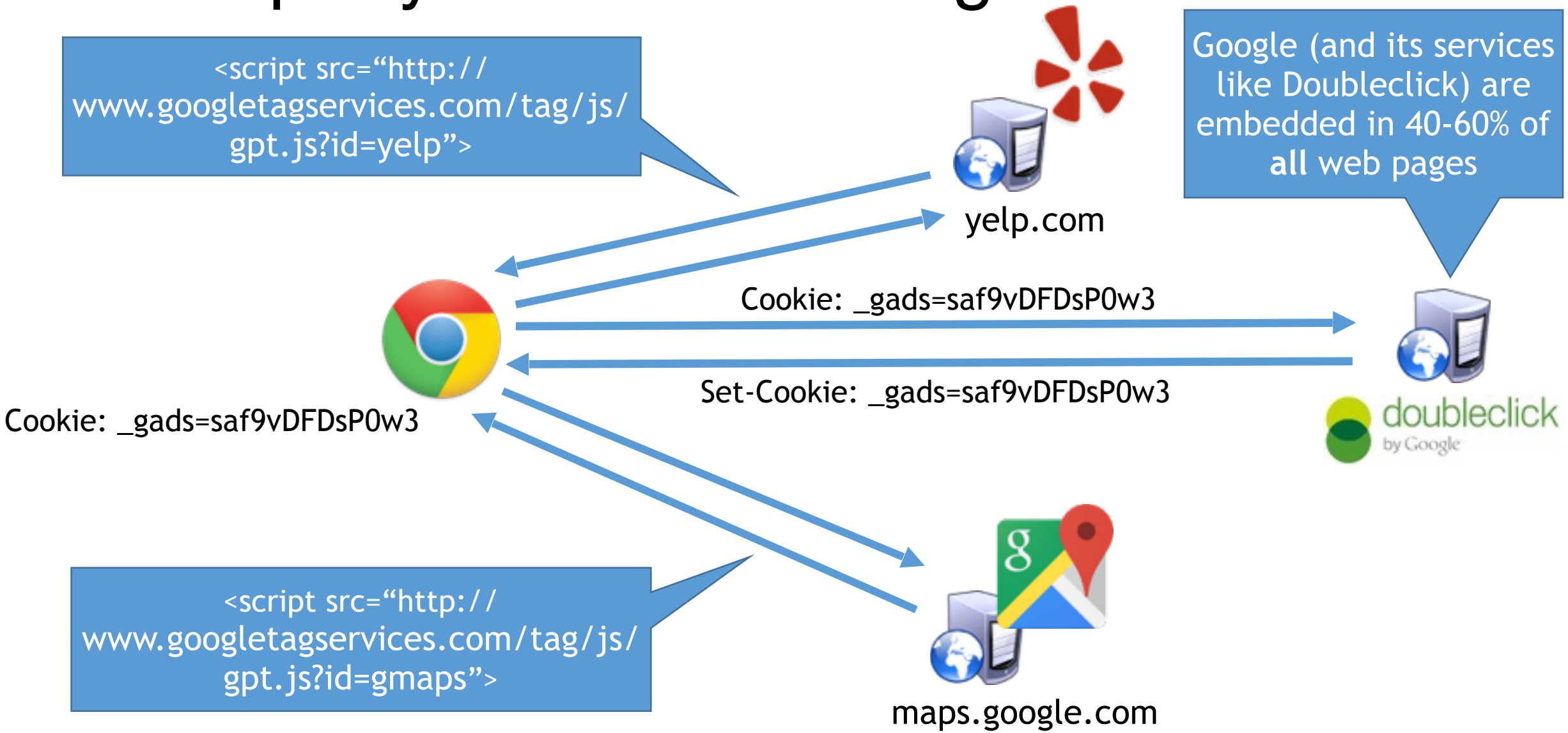
IP Address Tracking

- IP address is the most basic mechanism for tracking on the Internet
  - Everybody must have an IP address
  - Every packet you send contains your IP address
  - Your IP address remains relatively constant over time
- However, IP address is not as useful as it once was. Why?
  - NATs are ubiquitous; multiple people behind a single IP
  - Cell networks employ many layers of NATs and proxies
  - Users split their time across multiple devices with separate IPs

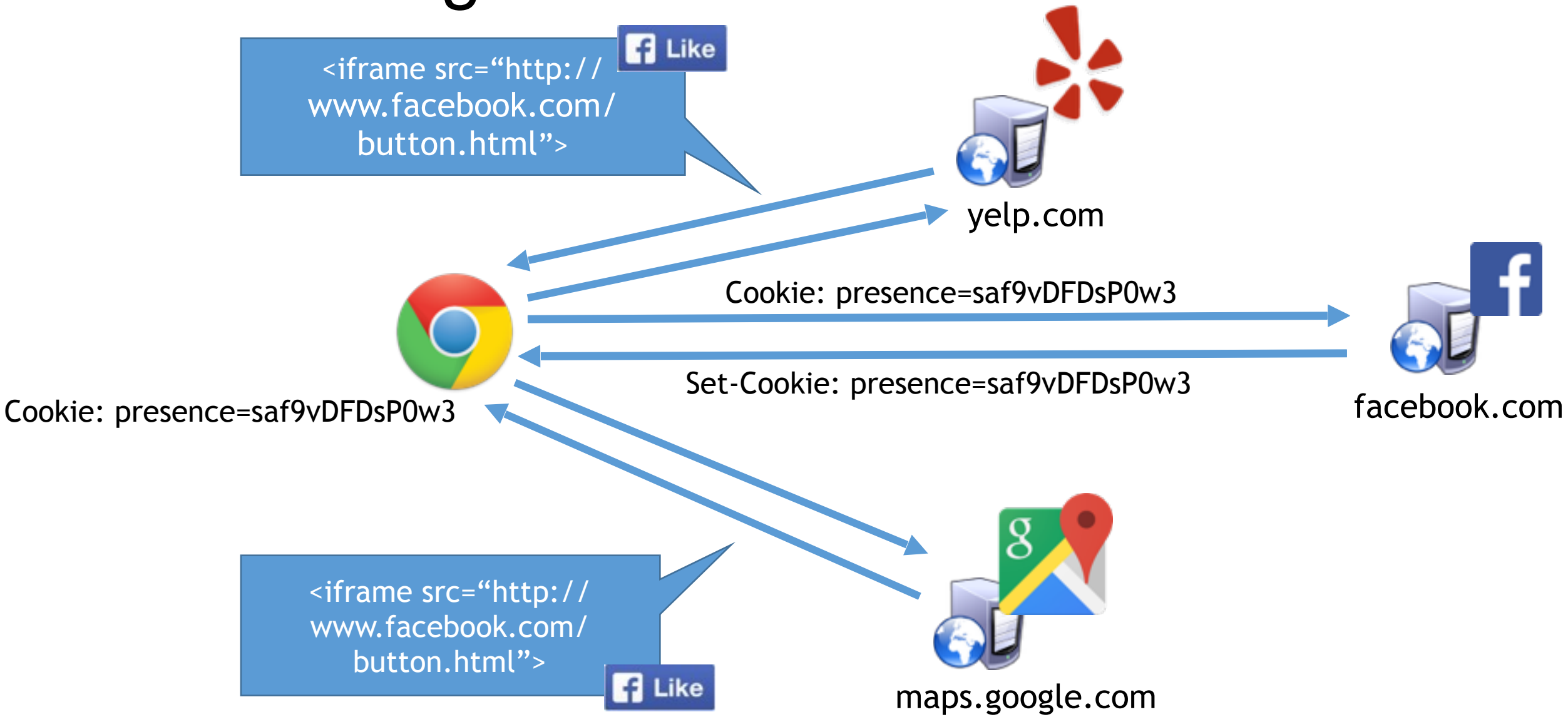
Cookies

- Allows servers to store state on client web browsers
  - Originally, invented for storing authentication information (session cookies)
  - Today, routinely used to implement tracking cookies
- Tracking cookies are so pervasive that they are now legislated
  - EU e-Privacy Directive (Cookie Law)
    - Requires that sites disclose if they use cookies and what they are used for
    - Users must opt-in before cookies can be set
  - Google was fined $22.5 Million by the FTC for circumventing cookie restrictions in Safari
    - Safari did not accept third-party cookies by default…
    - … unless they were received after a POST
    - Google used Ajax to send a POST to circumvent Safari's restriction

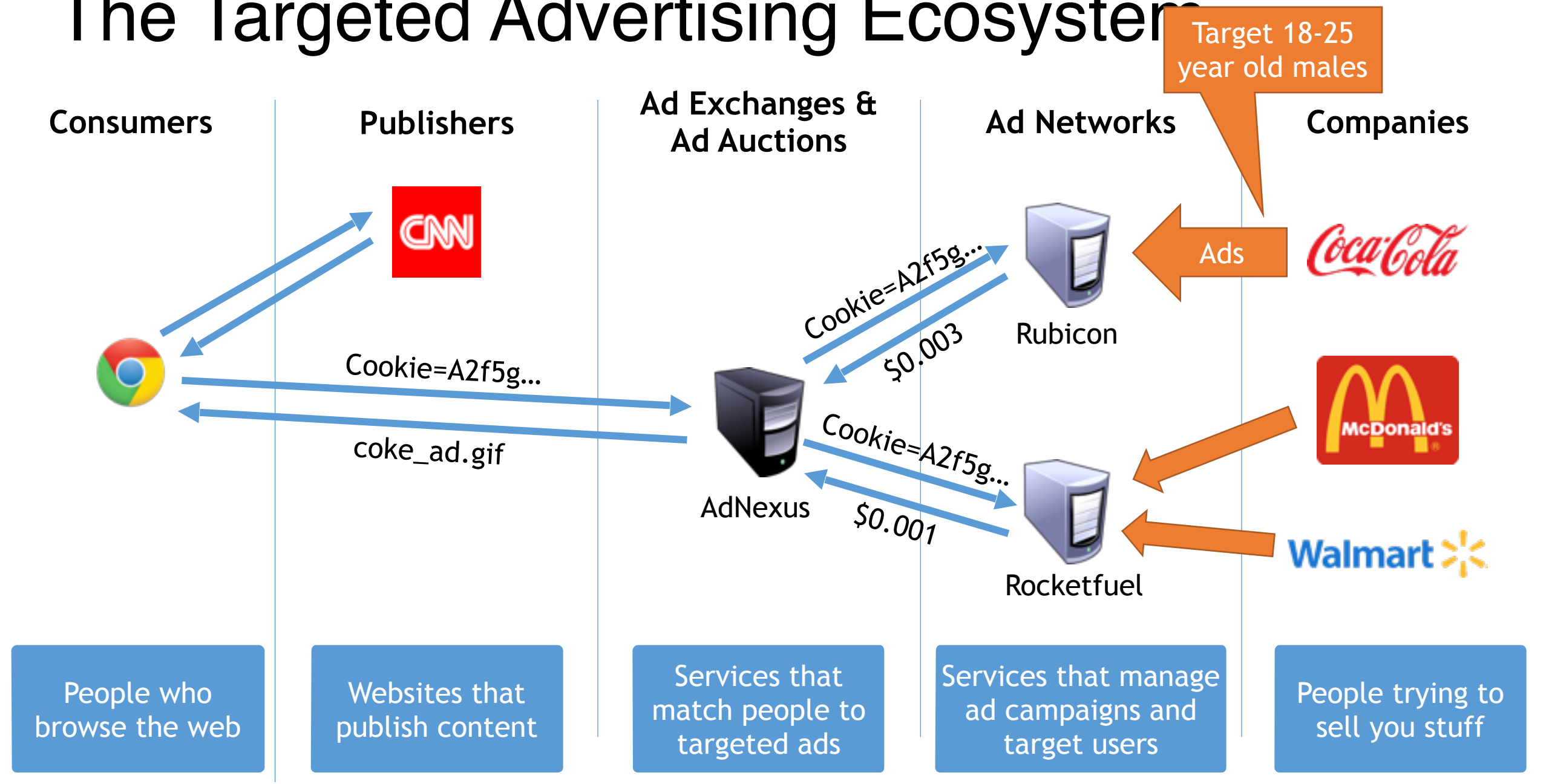
Third-party Cookie Tracking



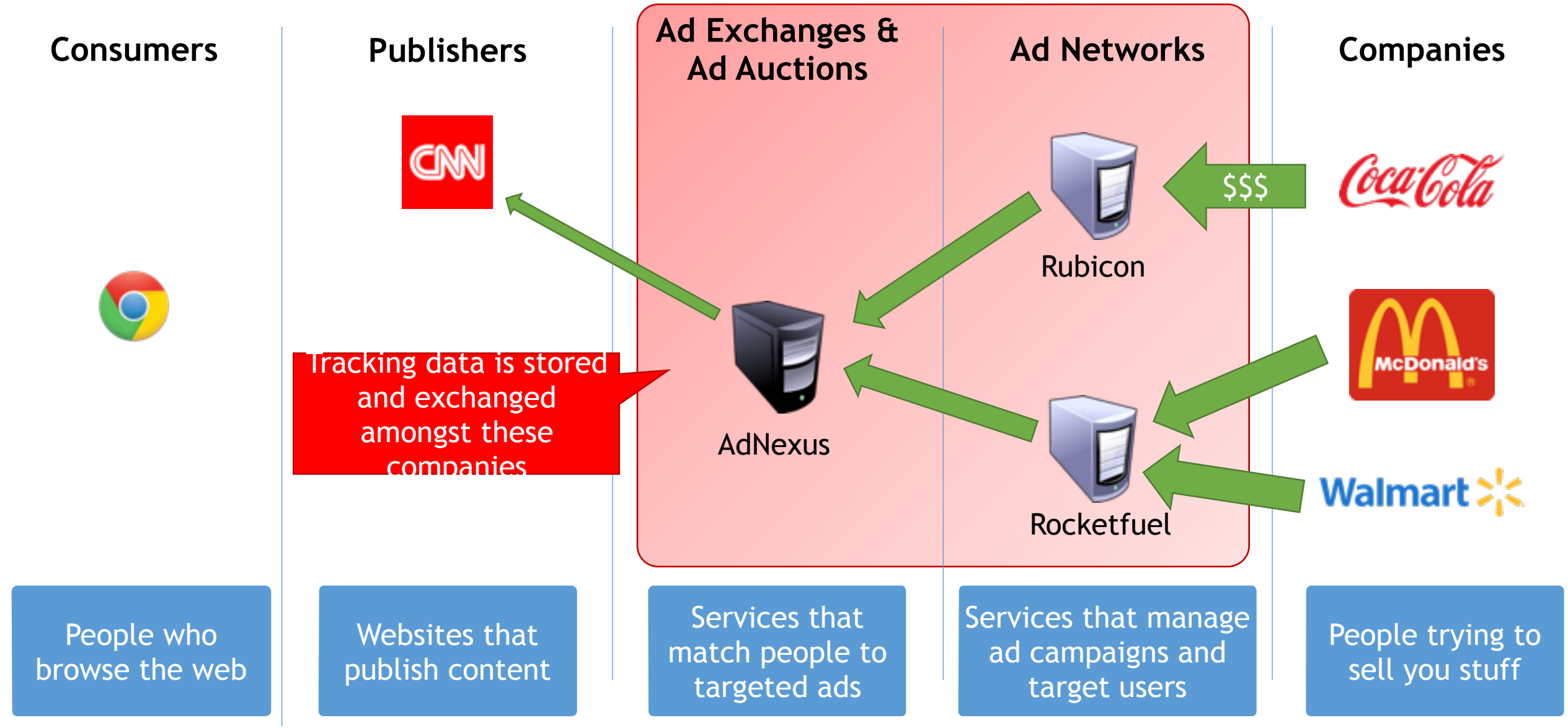
Social Widgets



The Targeted Advertising Ecosystem



The Targeted Advertising Ecosystem



Users Against Tracking Cookies

- Users did not respond well when they found out about tracking
- Many started clearing their cookies to avoid tracking
- Ad networks fought back using Evercookies
  - HTTP, HTML, and plugins provide many ways to store state on clients
  - Evercookies are placed in all available storage locations
  - If the cookie is deleted, it can be regenerated from the ‘backups’ in other locations

Evercookies

HTTP features

- Cookies
- E-tags – values set by the server that are supposed to be used for page caching
- Cached HTTP authentication credentials

HTML features

- window.name
- HTML5 localstorage
- HTML5 indexeddb
- HTML5 web database
- Web history (more on this later)

Plugins

- Flash Local Shared Objects (LSOs)
- Silverlight Isolated Storage
- Java PersistenceService

Mitigations Against Tracking Cookies

- Opting-out
  - In an effort to stave off regulation, many online ad networks have voluntarily joined the AdChoices program
  - AdChoices allows you to opt-out of some targeted advertising
  - Ironically, the opt-out is stored as a cookie in your browser
- Incognito/Private browsing mode
  - Starts a fresh browser instance with no cookies
  - All cookies are erased when the instance closes
  - Warning: plugins may still persist evercookie information
- Extensions
  - Adblock, Ghostery, Disconnect, PrivacyBadger, NoScript, uMatrix

Do Not Track

- Proposed in 2009 by Christopher Soghoian, Sid Stamm, and Dan Kaminsky
  - HTTP header that informs third-parties you do not wish to be tracked
  - Supported by most modern browsers (but typically off by default)
- The original aim was get buy in from regulators and advertisers
  - Instead, the whole effort became controversial and politicized
  - Today, no laws or regulations mandate compliance with DNT
  - Digital Advertising Alliance does not require its members to honor DNT
- Issues
  - Microsoft attempted to set DNT: 1 by default in IE 10
    - Advertisers revolted and refused to support the initiative
  - What is the expected behavior of Do Not Track?
    - Can a third-party retain data for other purposes like analytics, debugging, or security audits?
    - Can an advertiser store data but simply not use it to target ads?

Beyond Tracking Cookies

- Times are getting tough for cookie-based tracking
  - Tracker-blockers are proliferating
  - Anti-cookie legislation/regulation are increasing
- Many advertisers are experimenting with cookie-less tracking
  - Otherwise known as browser fingerprinting

Your Browser is Unique

GET / HTTP/1.1

Host: www.google.com

Connection: keep-alive

Cache-Control: max-age=0

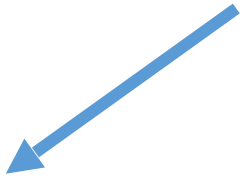
Accept: text/hmtl

User-Agent: Mozilla/5.0 (Windows NT 6.3; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/42.0.2311.68 Safari/537.36

Accept-Encoding: gzip,deflate,sdch

Accept-Language: en-US,en;q=0.8

Cookie: _session=aAB4m3rf8weG224

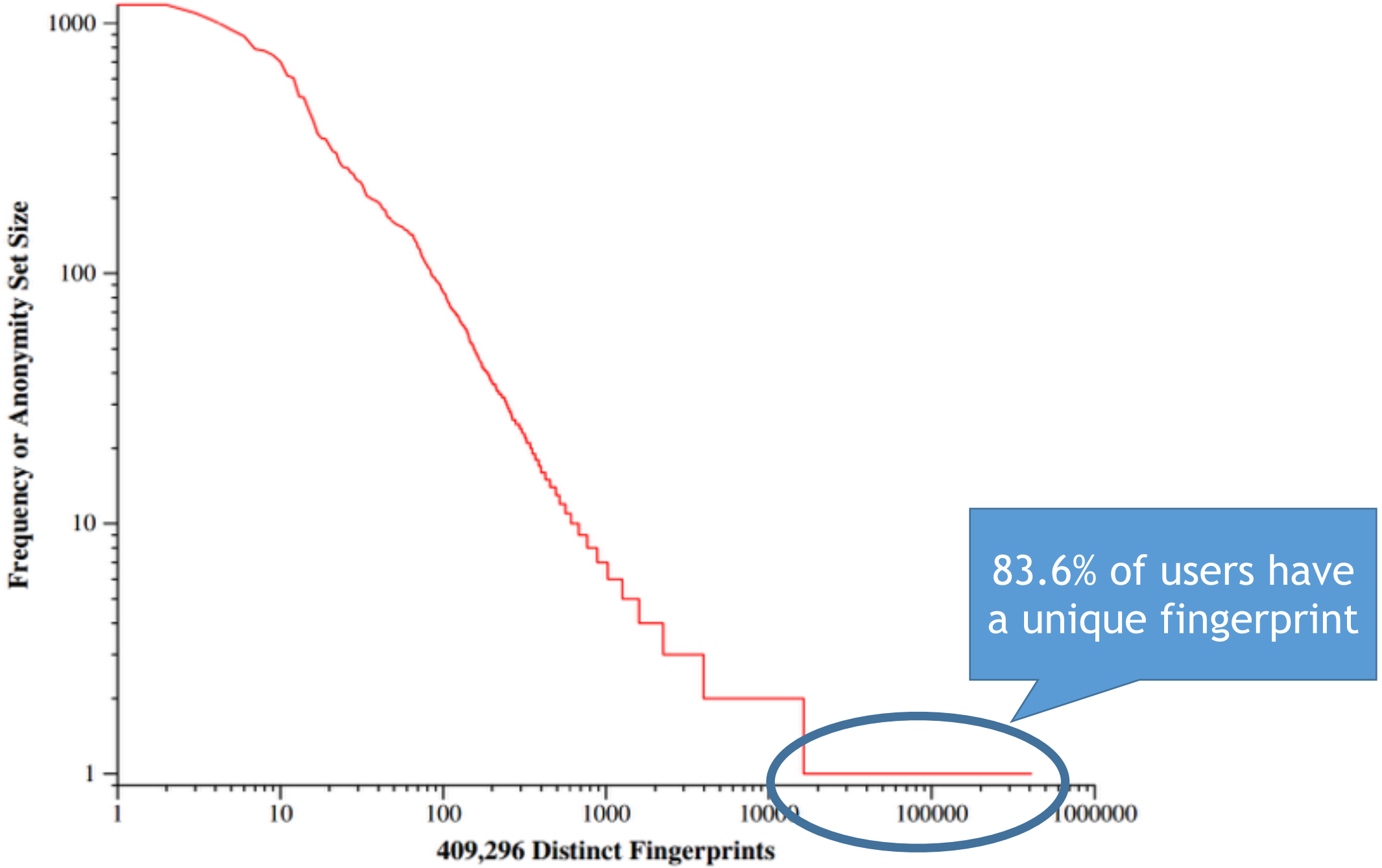


More Sources of Uniqueness

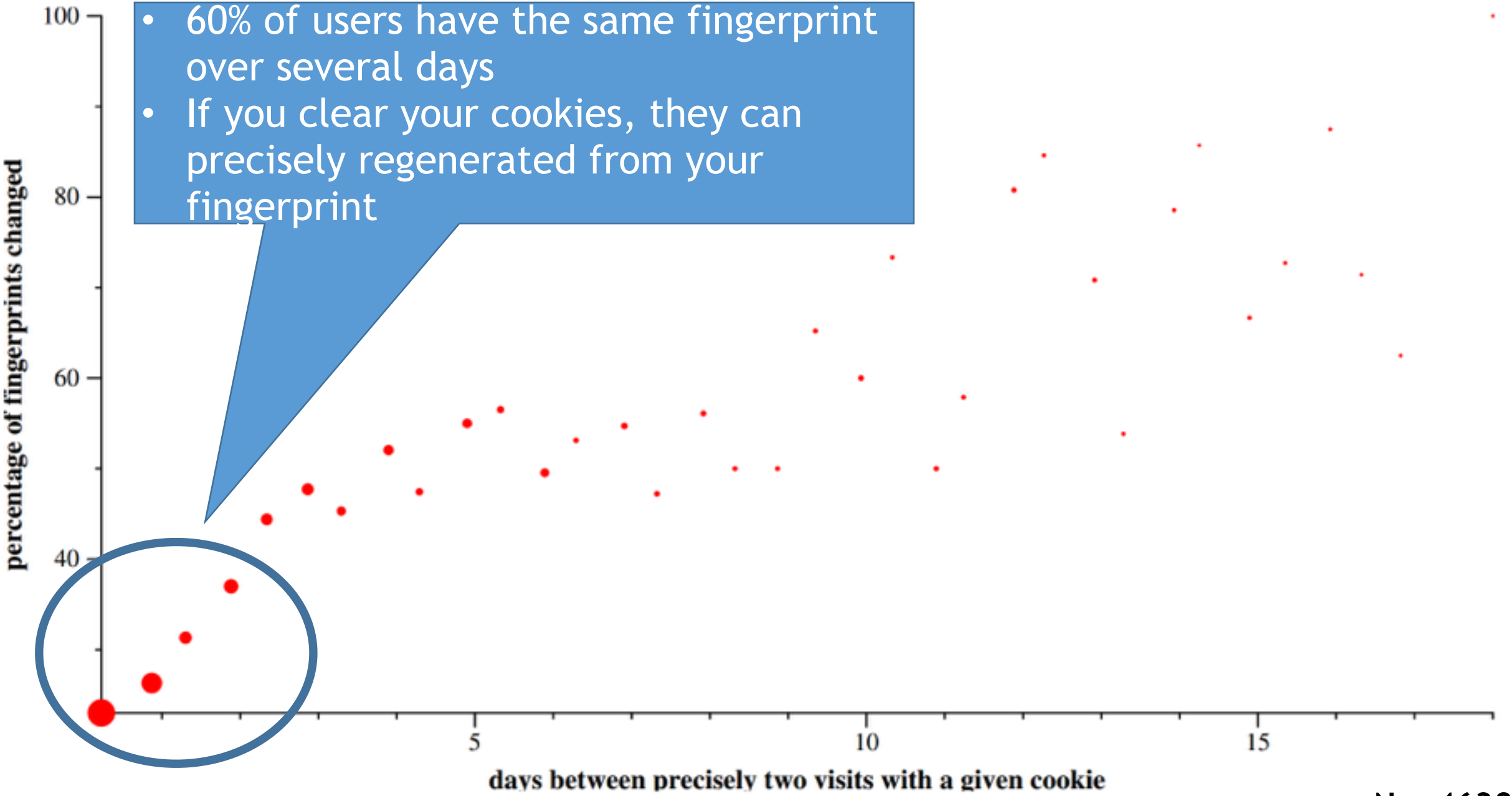
- Many more high-entropy characteristics are observable via Javascript/ plugins
  - What time zone are you in?
  - What fonts are installed on your machine?
  - What plugins are installed, and what are their versions?
  - What is your screen resolution and color depth?
  - Availability of specific JS APIs (i.e. browser version or platform dependent features)
  - Existence of specific browser extensions (e.g. AdBlock)
  - Order in which HTTP headers are sent
  - Hardware-level characteristics like CPU ID and frequency (MHz)

Browser Fingerprinting

- Fingerprinting is a method of identifying individual users based on the specific characteristics of their browser/system
  - Each characteristic is encoded as having bits of entropy
  - 15-20 total bits of entropy is enough to uniquely identify most people
- Examples:
  - Is Javascript enabled? – Roughly .4 bits of entropy (on or off, but off is much less common)
  - User-Agent? – Roughly 19 bits of entropy (OS → browser → version)
- Test yourself: https://panopticlick.eff.org/


1000
100
10
1
Frequency or Anonymity Set Size
1
10
100
1000
10000
100000
1000000
409,296 Distinct Fingerprints
83.6% of users have a unique fingerprint

- 60% of users have the same fingerprint over several days
- If you clear your cookies, they can precisely regenerated from your fingerprint



N = 4638

Canvas Fingerprinting

- Fingerprinting techniques are becoming more sophisticated
- Canvas fingerprinting
  - Leverages a hidden HTML5 <canvas>
  - Javascript renders text and drawing using various styles and fonts
  - The bitmap is then converted into a unique token
- Entropy is due to OS, browser, GPU, and graphics driver
  - Experiments observed 5.7 bits of entropy via canvas fingerprinting
  - True entropy is likely higher
- In 2014, many sites and ad trackers were caught using canvas fingerprinting

Canvas Fingerprinting Example

Mitigations Against Fingerprinting

- Adding more entropy into the browser
  - Example: uMatrix can randomize your User-Agent
  - Randomize the order of HTTP headers
- Reduce or restrict browser functionality
  - Cap the number of fonts a given page may query
  - Cap the number of plugins a given page may invoke
- Problem: some things cannot be randomized, removed, or restricted
  - Time zone and language cannot be randomized in general
  - Access to new Javascript APIs

History Stealing

CSS :visited

Timing Attacks

Story So Far

- Attacks thus far have been about inferring individual identity
  - Cookies and fingerprints
- What about attacks that try to infer your behavior
  - Specifically, your browsing history
  - Useful information for marketers and traditional attackers
    - E.g. do you have an account at BofA or a credit card with Chase?

Let's Talk About Hyperlinks

Visited Link

www.slashdot.org

www.reddit.com

www.webmd.com

www.chase.com

www.bankofamerica.com

Unvisited Link

```
var links =
document.querySelectorAll('a');

for (var x = 0; x < links.length; ++x) {
  console.log(

document.defaultView.getComputedStyle(
      link[x], null
    ).color
  );
}

>> rgb(85, 26, 139)    # Purple
>> rgb(0, 0, 238)      # Blue
>> rgb(85, 26, 139)    # Purple
>> rgb(85, 26, 139)    # Purple
>> rgb(0, 0, 238)      # Blue
```

History Stealing via CSS :visited

- Simple method to steal someone's browsing history

1. Send the victim to a page that includes malicious JavaScript *J*
   - Alternatively: use XSS to inject malicious JS into a benign website
2. *J* creates a list of <a> tags on the page
   - List is composed of links to well known sites
   - List can be hidden off-screen or using Javascript so the user is unware
3. *J* iterates through the list of anchors and examines their colors
   - Any purple links have been browsed by the victim

History Stealing via Timing Attack

- Observation: it takes browsers longer to render visited links than unvisited links
  - Unvisited: draw the link, has_link_been_visited() == false, move on
  - Visited: draw the link, has_link_been_visited() == true, draw the link again

1. Send the victim to a page that includes malicious JavaScript *J*
   - Alternatively: use XSS to inject malicious JS into a benign website
2. *J* injects <a> tags into the page one at a time
   - List is composed of links to well known sites
   - List can be hidden off-screen or using Javascript so the user is unware
3. *J* measures the time taken to draw each link
   - Calculate average draw-time by injecting links to non-existent pages
   - Links with draw-time significantly above the average have been visited

Mitigations Against History Stealing

- Basic approaches
  - Clear your history, or configure your browser to not store history
  - Disable styling of visited links
  - Disable Javascript
- Fixes implemented by Mozilla in 2010
  - CSS may only alter specific properties of :visited links versus :unvisited
    - Foreground and background color, outline, border, SVG stroke, and fill color
    - None of these properties impact the size or layout of surrounding elements
  - Javascript may no longer read certain style properties of links
    - All links appear to have unvisited colors
  - Changes to the rendering engine to make all links render in equal time

Final Thoughts

- Your personal information is valuable
  - Companies want it, attackers want it
- Your browser is a complex state machine that allows third-parties to run (somewhat) arbitrary code
  - Obvious and non-obvious mechanisms for tracking you personally…
  - … as well as your browsing history
- There is no silver bullet for privacy on the Web
  - Technological measures can help (modified browsers + extensions)
  - Eventually, regulatory mechanisms will also be necessary

Sources

1. Evercookies: http://samy.pl/evercookie/
2. Panopticlick (browser fingerprinting): https://panopticlick.eff.org/
3. Canvas fingerprinting examples: https://securehomes.esat.kuleuven.be/~gacar/persistent/index.html
4. History stealing example: http://www.dicabrio.com/javascript/steal-history.php
5. Plugging the CSS history leak: https://blog.mozilla.org/security/2010/03/31/plugging-the-css-history-leak/