



Third-Party Web Tracking

Technology, Policy, and Politics

Jonathan Mayer

jmayer@stanford.edu

stanford.edu/~jmayer

Obligatory Disclaimers



Stanford Security Laboratory



This is not legal advice.

History and Status Quo

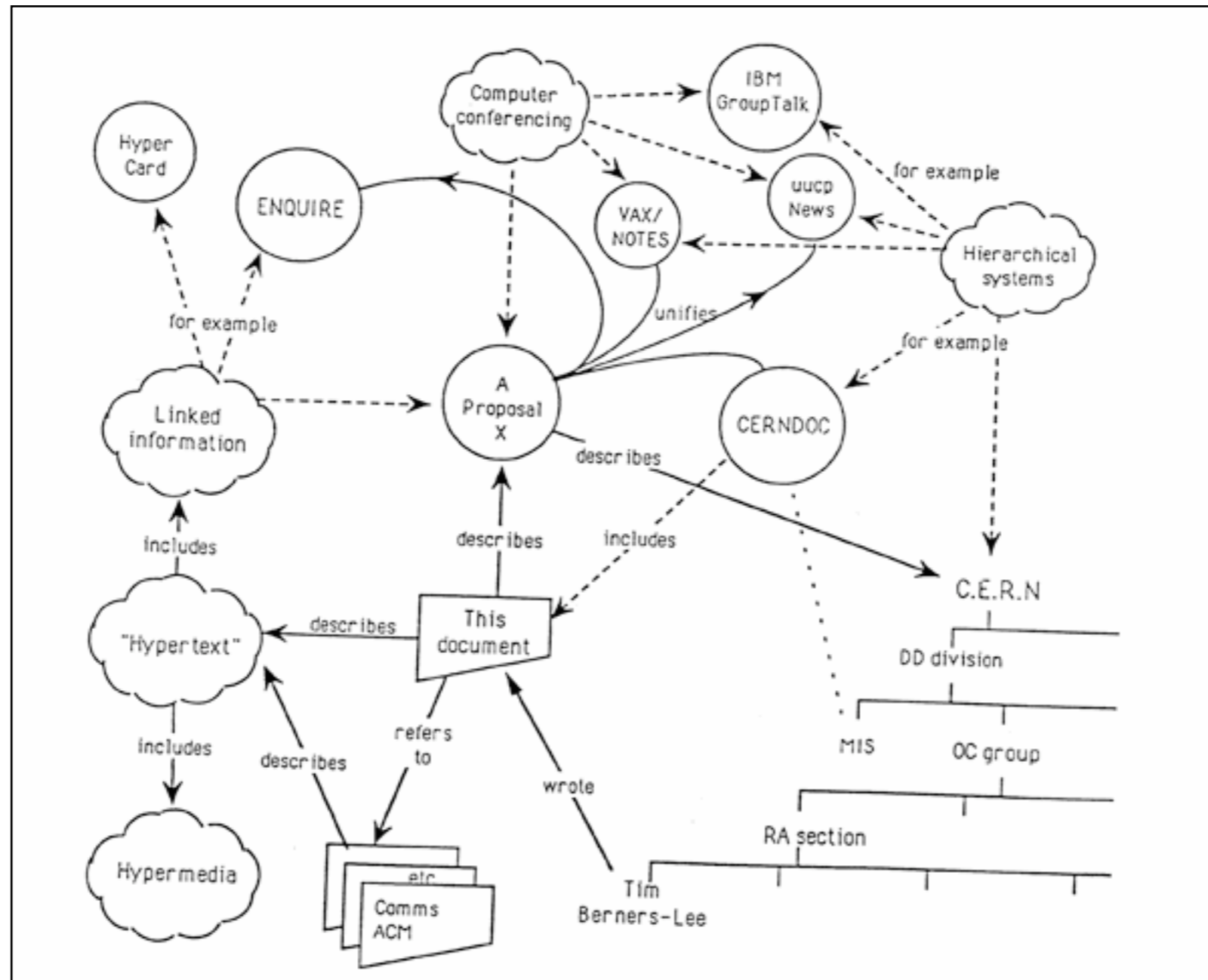
Tracking Methods

Signaling Mechanisms

Technical Countermeasures

Government Intervention

Privacy-Preserving Advertising



1989

World Wide Web

The WorldWideWeb (W3) is a wide-area [hypermedia](#) information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an [executive summary](#) of the project, [Mailing lists](#) , [Policy](#) , November's [W3 news](#) , [Frequently Asked Questions](#) .

[What's out there?](#)

Pointers to the world's online information, [subjects](#) , [W3 servers](#), etc.

[Help](#)

on the browser you are using

[Software Products](#)

A list of W3 project components and their current state. (e.g. [Line Mode](#) ,X11 [Viola](#) , [NeXTStep](#) , [Servers](#) , [Tools](#) , [Mail robot](#) , [Library](#))

[Technical](#)

Details of protocols, formats, program internals etc

[Bibliography](#)

Paper documentation on W3 and references.

[People](#)

A list of some people involved in the project.

[History](#)

A summary of the history of the project.

[How can I help ?](#)

If you would like to support the web..

[Getting code](#)

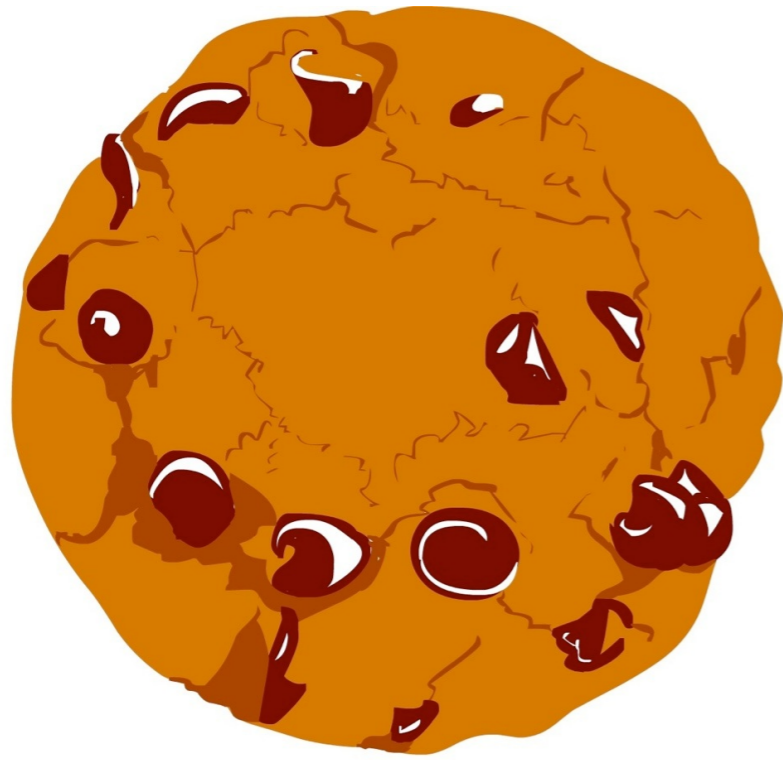
Getting the code by [anonymous FTP](#) , etc.

1992

<marquee>

<blink>

That didn't last long.



Cookies

1994



JavaScript

1995

“A user agent should make ~~every~~ ^{no} attempt to prevent the sharing of session information between hosts that are in different domains.”

-IETF RFC 2109

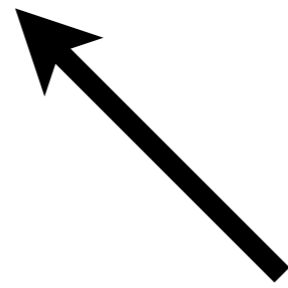
1997

DoubleClick
media. technology. people.

1997



Current Page



DISPLAY LUMAscape



Denotes acquired company

Denotes shuttered company

© LUMA Partners LLC 2013

<script>

<iframe>

GET http://advertising.com/ HTTP/1.1
Referer: http://news.com/
Cookie: id=12345

| User ID | Time | URL | Page Title |
|---------|--------------------|---|------------------------------|
| 12345 | 6/18/12
10:01am | <u>http://
foxnews.com/...</u> | Why Liberals
Hate America |
| 12345 | 6/18/12
10:02am | <u>http://
youtube.com/...</u> | Squirrels
Waterskiing?! |

...

HELLO
MY NAME IS

Browsing History

Health Information

Financial Information

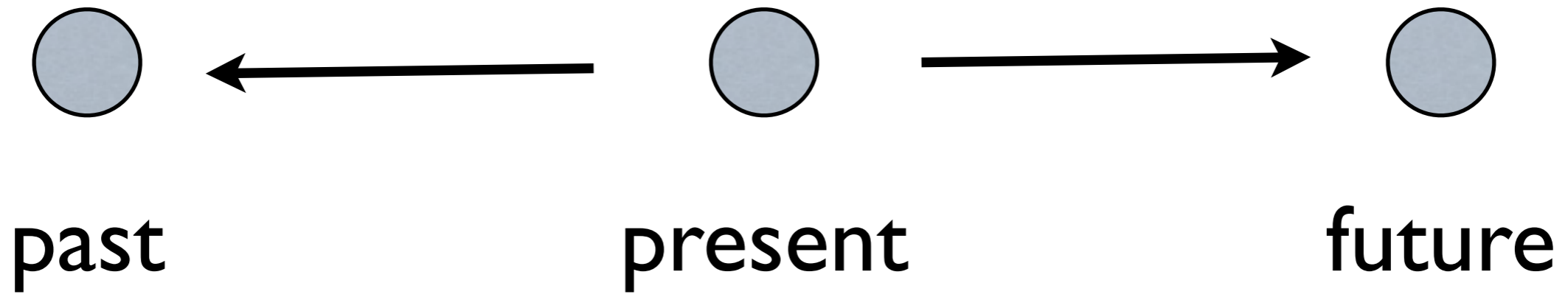
Shopping History

...



“it’s all anonymous”

actually, it's all pseudonymous



- social network or other first party
- intentional leakage
- unintentional leakage
- security exploit
- deanonymization



1. Scorecard Research, 81 sites (44%)
2. Google Analytics, 78 sites (42%)
3. Quantcast, 63 sites (34%)
4. Google Advertising, 62 sites (34%)
5. Facebook, 45 sites (24%)

(signed up and interacted with 185 sites)



[Learn About Mitt](#)

[In Your Community](#)

[News & Media](#)

[Shop](#)

Leland Stanford | [Logout](#)

[CONTRIBUTE](#)

Americans deserve more jobs and more take-home pay.

[Read Mitt's Five Point Plan](#)

MyMitt Member Account

[EDIT ACCOUNT](#)

Leland Stanford



Member for: 26 sec

[CREATE A FUNDRAISER](#)

TOTAL AMOUNT RAISED
\$0

PEOPLE WHO DONATED
0

FUNDRAISING CAMPAIGNS
0

POINTS
5,000

INVITE A FRIEND ?

[SEND INVITE](#)

[See your invites](#)

 **GEAR UP**
Shop for official campaign gear.

[SHOP NOW](#)

 **DONATE**
Make a contribution.

 **MAKE CALLS FOR MITT**

LELAND STANFORD

Supporter for Stanford Faculty for Obama

[Help](#)



Your profile is not public.
Update your sharing settings to share your page and the work you're doing to help re-elect the President.

Edit profile

Edit account

You're currently a supporter of Stanford Faculty for Obama.
Want to get more involved? [Talk to your team manager](#) about becoming a Team Member.

Why I'm in

[edit](#)

tracking |'trakiNG|

noun

1 the collection of a user's browsing history

- sensitive, identifiable information
- lack of transparency
- lack of usable, effective controls
- inadequate market incentives

History and Status Quo

Tracking Methods

Signaling Mechanisms

Technical Countermeasures

Government Intervention

Privacy-Preserving Advertising

stateful tracking

tagging

stateless tracking

fingerprinting

HTTP cookies

HTTP authentication

HTTP ETags

content cache

IE userData

Flash Local Shared Objects

Silverlight Isolated Storage

TLS session ID & resume

browsing history

window.name

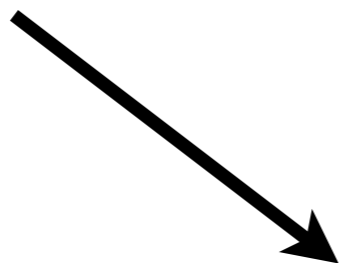
HTML5 protocol & content handlers

HTTP STS

HTML5 session/local/global/database storage

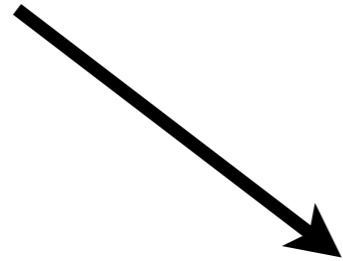
DNS cache

this is blue



link

this is purple



link

User-Agent

installed fonts

HTTP ACCEPT Headers

cookies enabled?

browser plug-ins

browser add-ons

MIME support

screen resolution

clock skew

the technologies that make the web awesome
are the very technologies that enable tracking

Lots of empirical questions.

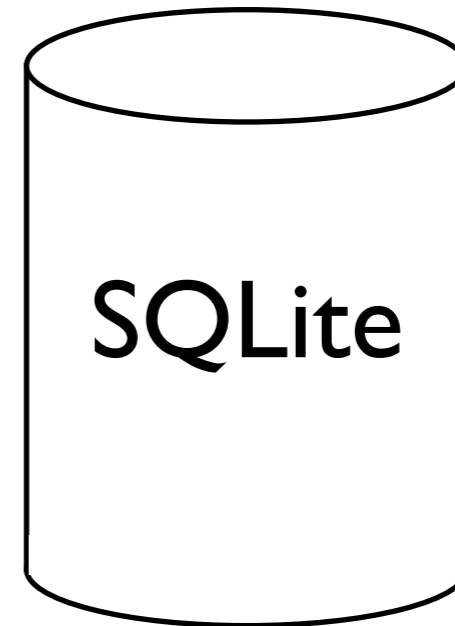
Many Research Designs

1. build custom platform for experiment
2. run experiment
3. write paper
4. goto 1

FourthParty Design

1. build one platform
2. collect as much data as possible
3. run many experiments
4. write many papers

FourthParty Architecture



- easy to use
- shared data, historical data
- works with existing extensions (crawling and more)
- multiplatform

History and Status Quo

Tracking Methods

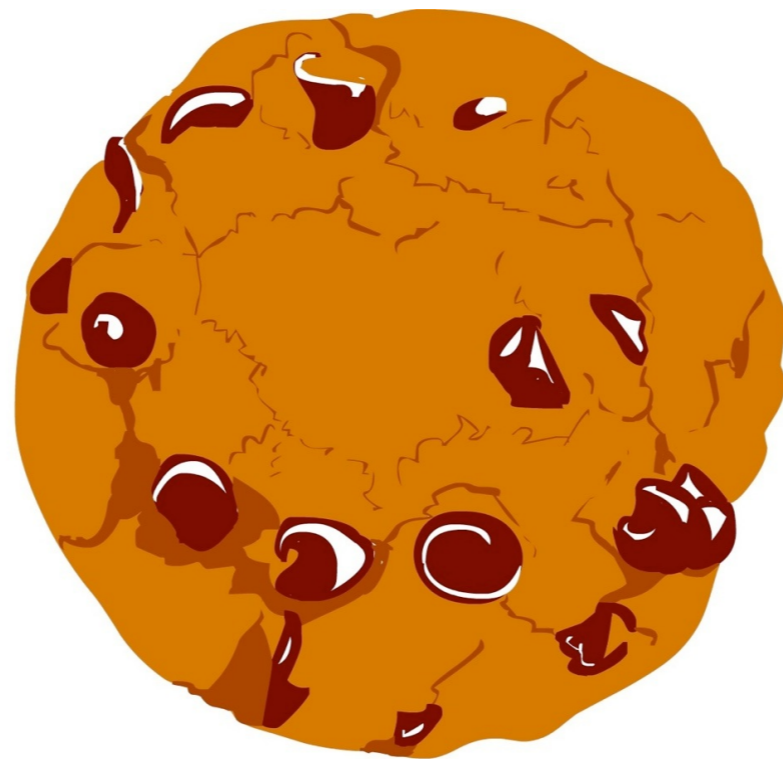
Signaling Mechanisms

Technical Countermeasures

Government Intervention

Privacy-Preserving Advertising

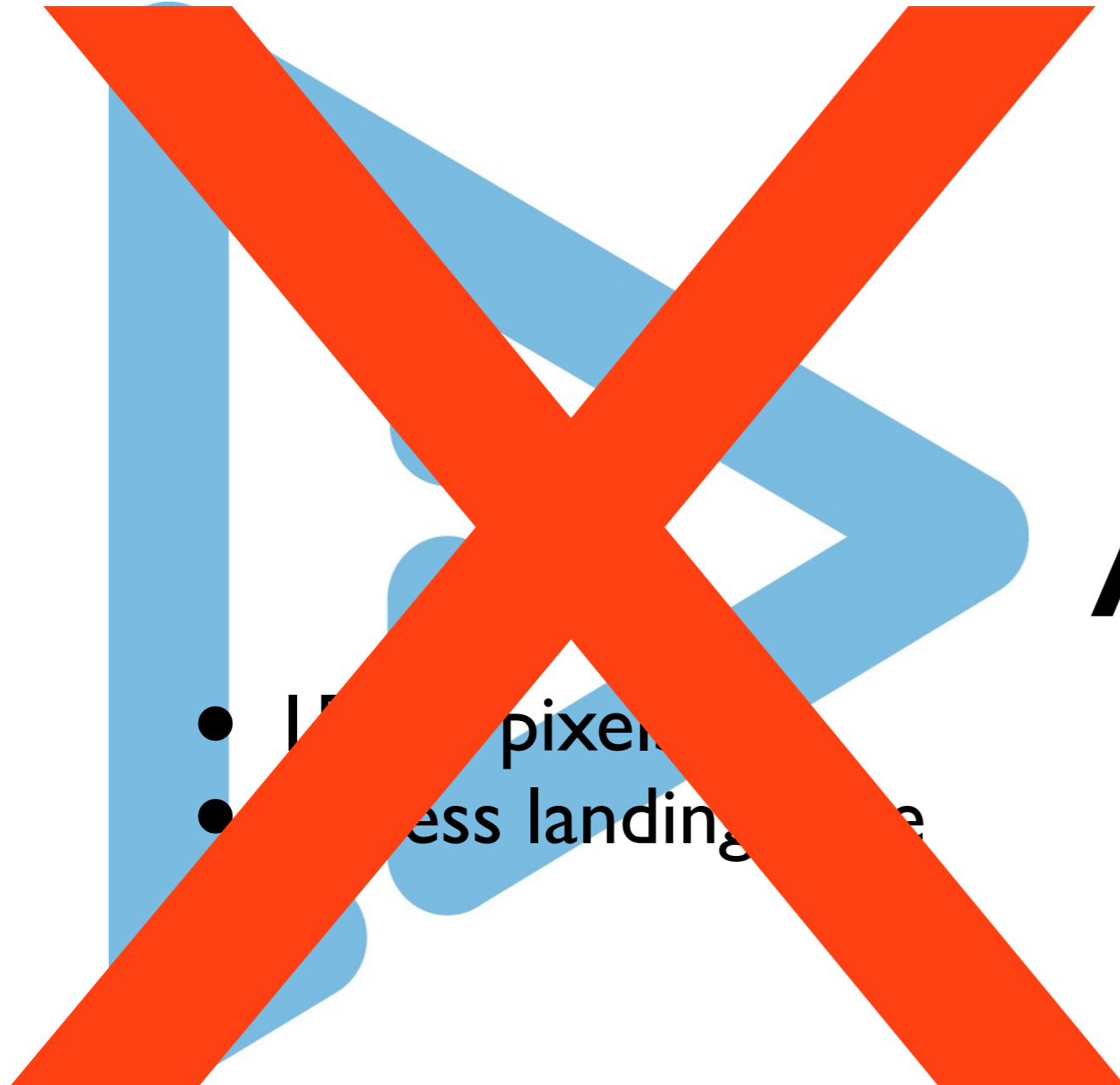
opt out





≈ 75 companies

- not comprehensive
 - not all third-party trackers offer
 - vast majority do not participate in NAI
- requires updating*
- can accidentally clear*



- 100 pixels
- Less landing

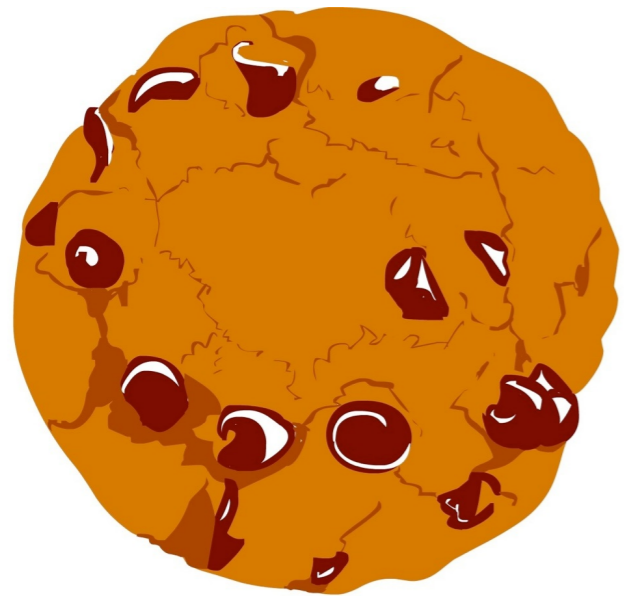
10% with icon

AdChoices

5% with icon + text

opt out

DECEPTIVE



= Do Not Target Ads

Do Not Track

- universal
- no updating
- one click
- in the browser

HTTP header

DNT : 1



4+



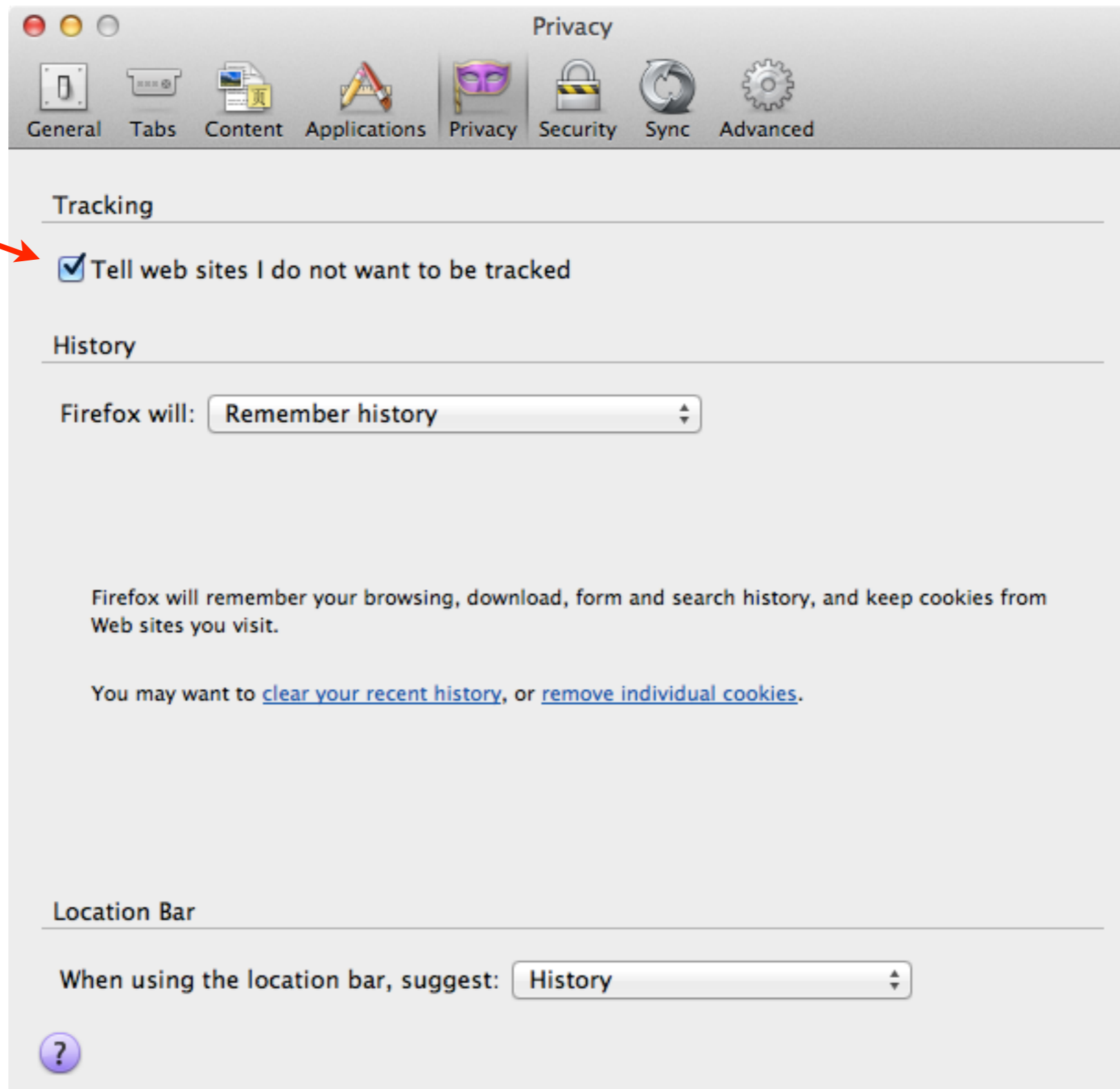
9+

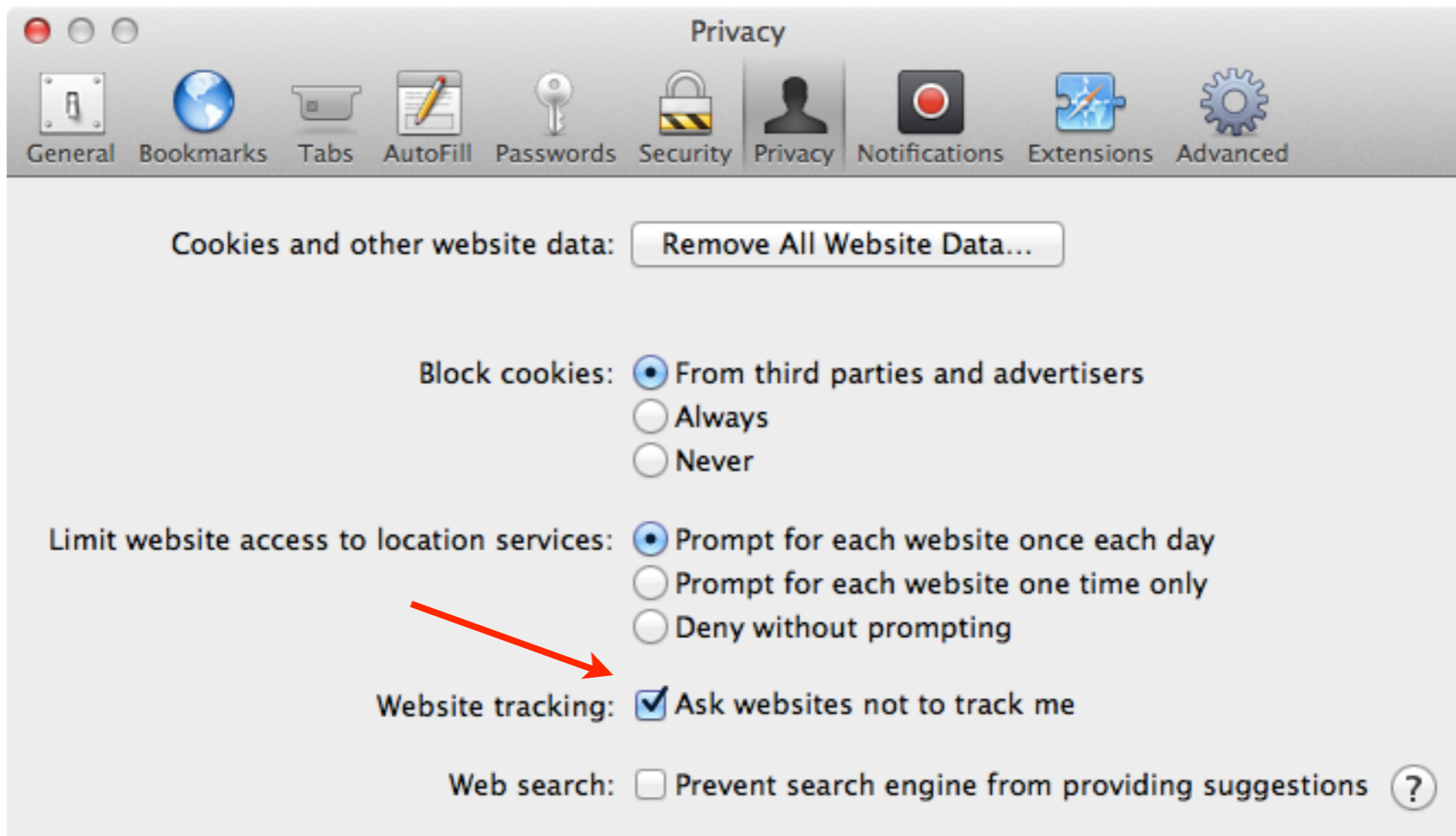


5.1+



25+





chrome://settings

Chrome Settings

Search settings

History

Extensions

Settings

Help

Sign in to Chrome

On startup

- Open the New Tab page
- Continue where I left off.
- Open a specific page or set of pages. [Set pages](#)

Appearance

[Get themes](#) [Reset to default theme](#)

- Show Home button
- Always show the bookmarks bar

Search

Set which search engine is used when searching from the [omnibox](#).

[Google](#) [Manage search engines...](#)

- Enable Instant for faster searching (omnibox input may be [logged](#))

Users


You are currently the only Google Chrome user.

[Add new user...](#) [Delete this user](#) [Import bookmarks and settings...](#)

Default browser

The default browser is currently Google Chrome.

[Show advanced settings...](#)



chrome://settings

Chrome Settings

Default browser

The default browser is currently Google Chrome.

Privacy

Content settings... Clear browsing data...

Google Chrome may use web services to improve your browsing experience. You may optionally disable these services. [Learn more](#)

- Use a web service to help resolve navigation errors
- Use a prediction service to help complete searches and URLs typed in the address bar
- Predict network actions to improve page load performance
- Enable phishing and malware protection
- Use a web service to help resolve spelling errors
- Automatically send usage statistics and crash reports to Google
- Send a 'Do Not Track' request with your browsing traffic

Passwords and forms


- Enable Autofill to fill out web forms in a single click. [Manage Autofill settings](#)
- Offer to save passwords I enter on the web. [Manage saved passwords](#)

Web content

Font size: Medium Customize fonts...

Page zoom: 100%

- Pressing Tab on a webpage highlights links, as well as form fields



Chrome

Settings

Search settings

History

Extensions

Settings

Help

Default browser

The default browser is currently Google Chrome.

Privacy

Content settings...

Clear browsing data...

Google Chrome may use web services to improve your browsing experience. You may optionally disable these services. [Learn more](#)

- Use a web service to help improve search results
- Use a prediction service to help improve search results
- Predict network connectivity to help improve search results
- Enable phishing protection
- Use a web service to help improve search results
- Automatically update Chrome
- Send a 'Do Not Track' header with my browsing requests

Do Not Track

Enabling 'Do Not Track' means that a request will be included with your browsing traffic. Any effect depends on whether a website responds to the request, and how the request is interpreted. For example, some websites may respond to this request by showing you ads that aren't based on other websites you've visited. Many websites will still collect and use your browsing data - for example to improve security, to provide content, services, ads and recommendations on their websites, and to generate reporting statistics. [Learn more](#)

OK Cancel

Passwords and forms

- Enable Autofill to fill out web forms in a single click. [Manage Autofill settings](#)
- Offer to save passwords I enter on the web. [Manage saved passwords](#)

Web content

Font size: Medium Customize fonts...

Page zoom: 100%

- Pressing Tab on a webpage highlights links, as well as form fields



Settings

Express settings

You can customize the following settings or choose express settings. If you choose express settings, your PC will occasionally send info to Microsoft and will:

- Automatically install important and recommended updates.
- Help protect your PC from unsafe files and websites.
- Turn on Do Not Track in Internet Explorer.
- Help improve Microsoft software, services, and location services by sending us info.
- Check online for solutions to problems.
- Let apps give you personalized content based on your PC's location, name, and account picture.
- Turn on sharing and connect to devices on this network.

[Learn more about express settings](#)

[Privacy statement](#)

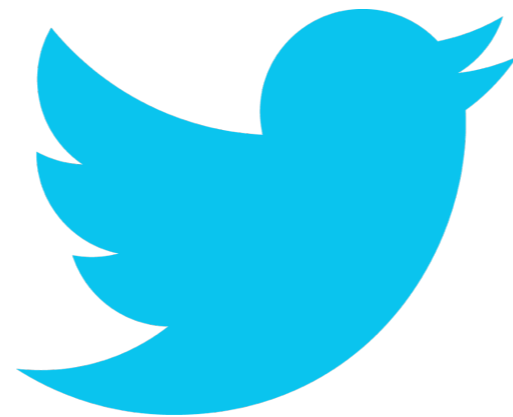


Use express settings

Customize

tens of millions of users

≈ 25 third parties honor DNT



Advertising industry White House pledge:
Implement by end of 2012

W3C

Negotiations for 2+ years, no agreement.

Enforcing a Signaling Mechanism

- ✓ ● observe suspicious behavior
- monitor ad distributions

History and Status Quo

Tracking Methods

Signaling Mechanisms

Technical Countermeasures

Government Intervention

Privacy-Preserving Advertising

old view: privacy \approx security

anti-tracking technology

blocking



Internet Explorer 9



| forbes.com^*/track.
| fresh.techdirt.com^
| frstatic.net^*/tracking.js
| ft.com^*/ft-tracking.js
| ft.com^*/fttrack2.js
| ft.com^*/si-tracking.js
| g.msn.com^
| gamerevolution.com^*/gn_analytics.min.js
| gamesradar.com^*/clacking.js
| gametrailers.com/neo/stats/
| gamezone.com/?act=
| gamezone.com/site/linktracker.js
| geo.perezhilton.com^
| geo.yahoo.com^
| geoip.mlive.com^
| geoip.nola.com^
| geoiplookup.wikimedia.org^
| ghostery.com^*/clicky.js
| go.com/stat/
| goauto.com.au^*/ecblank.gif?
| godaddy.com/image.aspx?
| google.*/gwt/x/ts?
| google.*/stats?ev=
| google.com/lh/ajaxlog?
| google.com/uds/stats?
| greatschools.org/res/js/trackit.js
| guim.co.uk^*/sophusthree-tracking.js
| harrisbank.com^*/zig.js
| heraldm.com/tracker.tsp?
| hitcount.herald.com^
| holiday-rentals.co.uk/thirdparty/tag
| holiday-rentals.co.uk^*/hrtrackjs.gif?
| hostels.com/includes/lb.php?
| hostels.com/includes/thing.php?
| hostels.com/includes/vtracker.php?

- not comprehensive
- requires updating
- breaks stuff
- requires user knowledge about providers

“blocks many ... technologies that can track and profile you as you browse the Web ... updated weekly”

-Abine TPL

“completely removes all forms of tracking from the internet”

-EasyPrivacy ABP Subscription

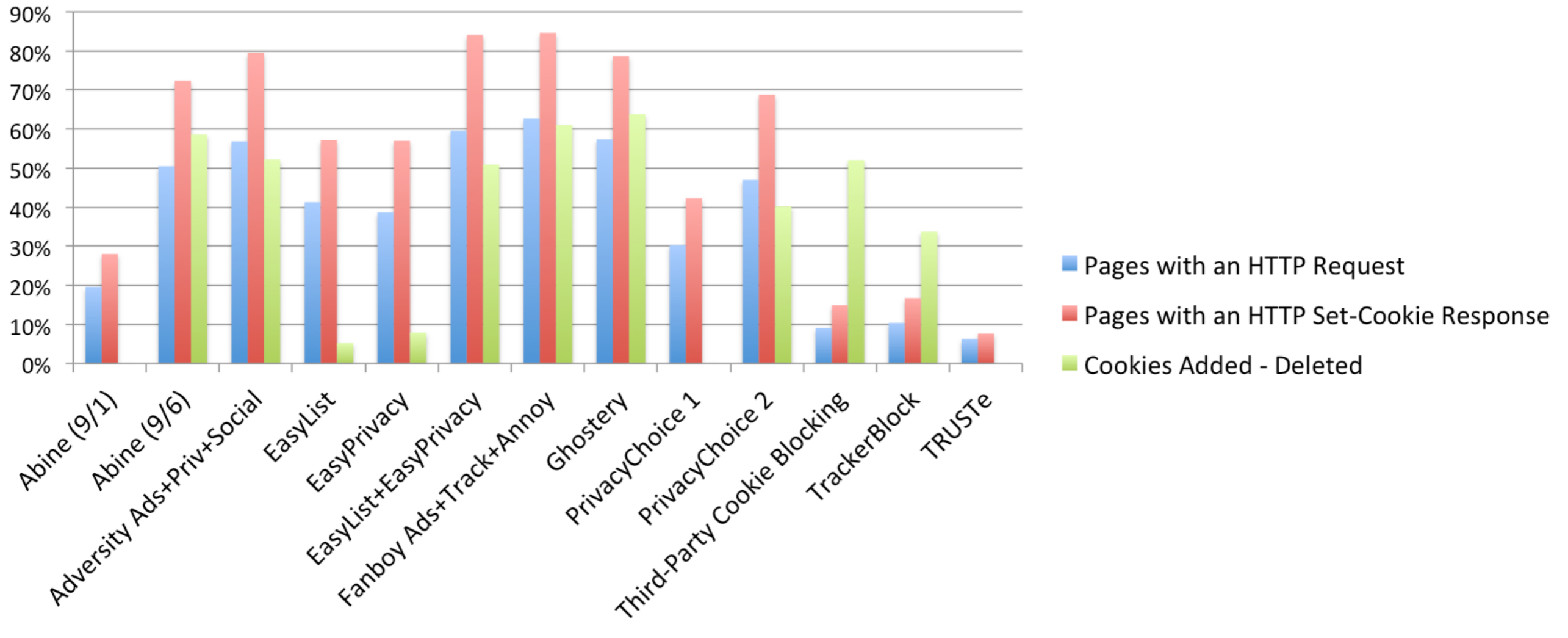
“complete control over online tracking”

-PrivacyChoice TrackerBlock

“helps users get good ads, without compromising personal privacy”

-TRUSTe TPL

Average Decrease in Tracking

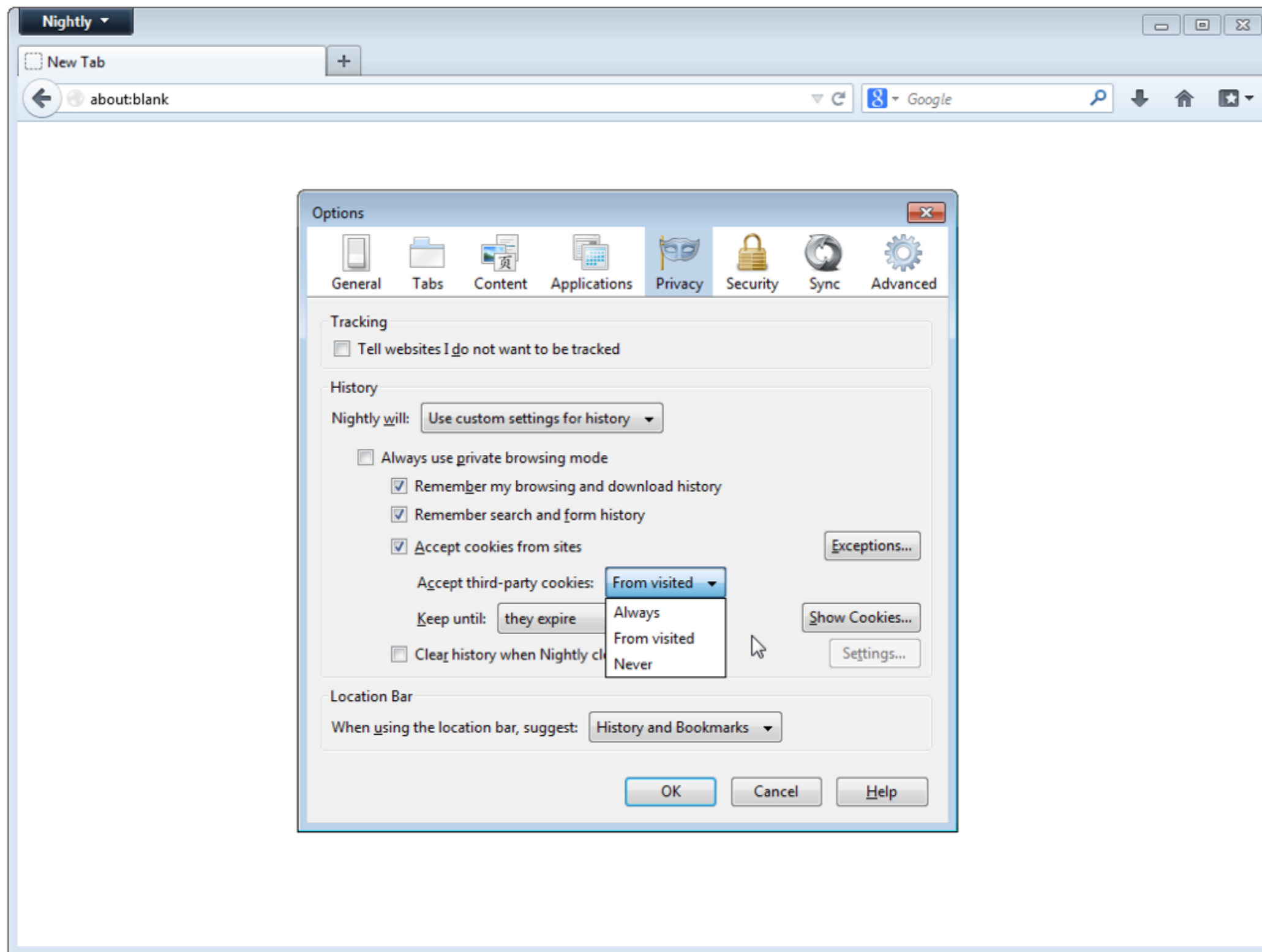


(we can't reasonably expect the average user to sort this out)

new view: privacy \neq security



relaxed third-party cookie blocking



study: measuring browsers


file:///Users/j...easurement.html

Go to a Website

Google

STANFORD UNIVERSITY

Stanford Center for Professional Development



Adapt to new opportunities.

Request Information

First Name

Last Name

Email

SUBMIT

[f](#) [t](#) [in](#) [You Tube](#) [g+](#)

Online Engineering Courses, Certificates, and Degrees from Stanford University

Participate in Stanford courses, seminars, and webinars **online** and be part of the Stanford University spirit of innovation and openness. Our online engineering courses reflect the exciting research and world-class teaching of Stanford faculty from the School of Engineering.

Engineering Courses | Leadership & Management Courses | Why Stanford? | Free Content


Earn world-class credentials.

Through the Stanford Center for Professional Development, you may take online individual graduate and professional courses in engineering and related fields. Graduate courses may also be audited. Earn a graduate or professional certificate or a master of science degree in engineering part-time, at a pace that works for you. Listed here are some areas in which you may take online courses and earn a certificate or an engineering degree:

Computer Science and Information Technology

- Computer Science MS Degree
- Advanced Computer Security Professional Certificate
- Advanced Software Systems Graduate Certificate
- Artificial Intelligence Graduate Certificate

Student Spotlight



FLEX INTERCONNECT TECHNOLOGIES

"The coursework and lessons learned at Stanford have enabled me to work more closely with customers and be on the forefront of product development."

60k impressions
13 countries

Result 1: Advertising Opt Out Usage (U.S.)



AdChoices

Inspiron™ 15R

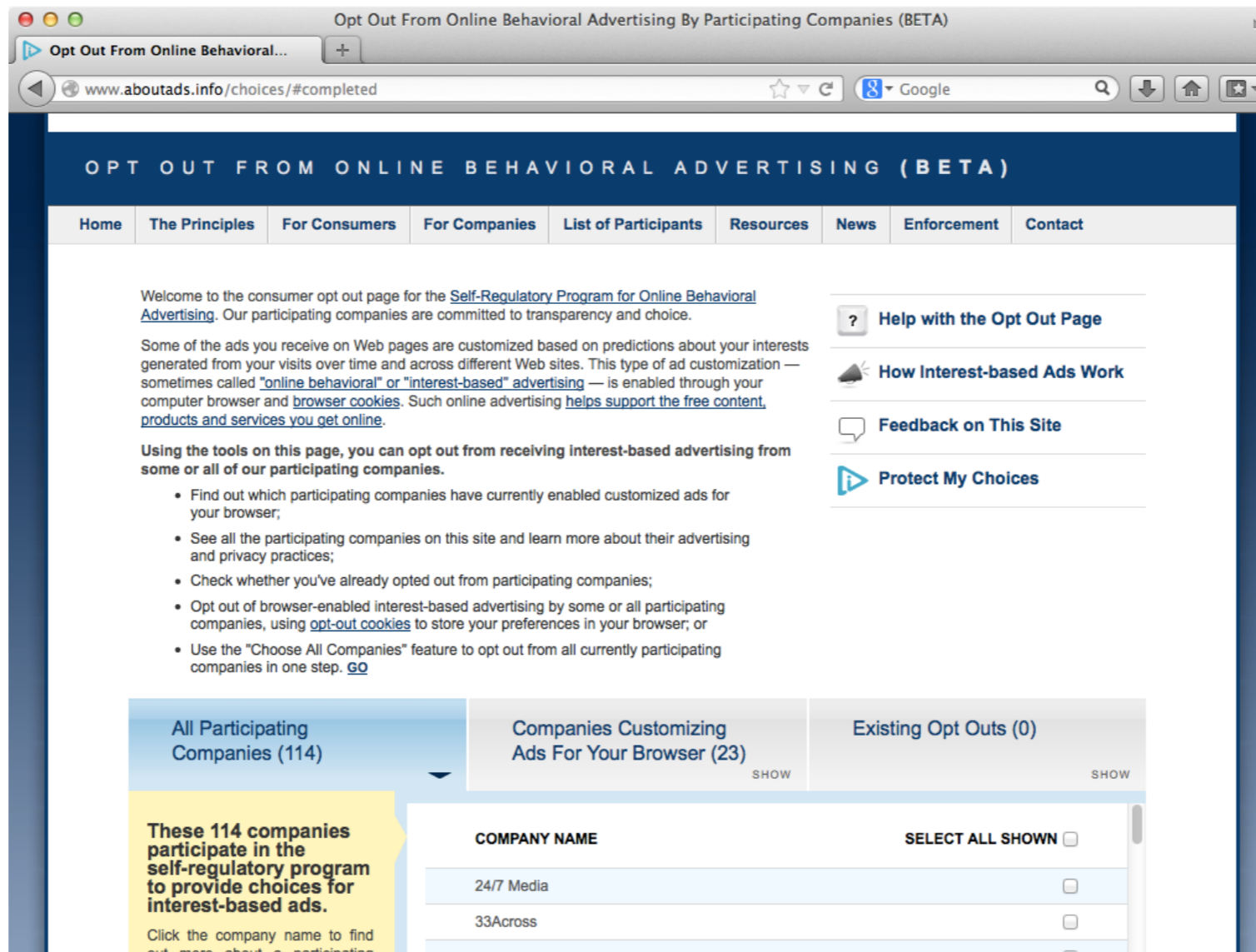
It helps me email, study, chat,
flirt, and focus all at once



Select Dell systems are
available with 2nd
generation Intel® Core™ i7
processors

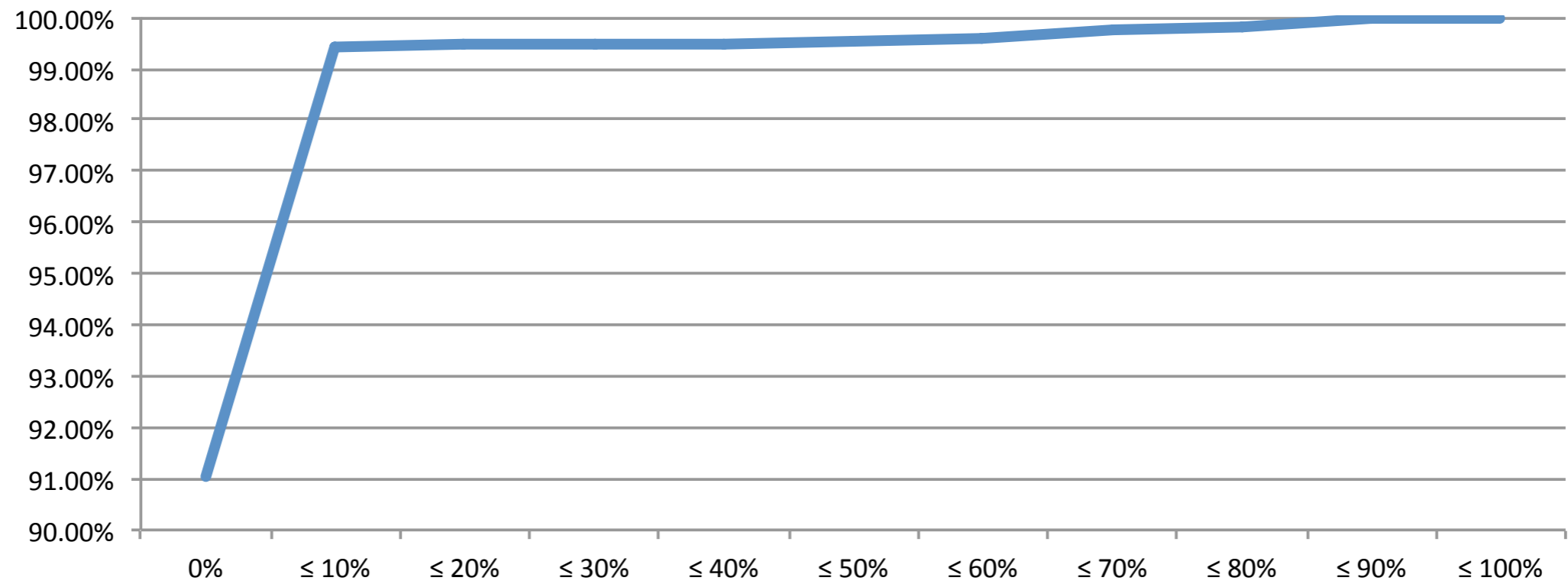
Shop Now >





Queried \approx 175 advertising status mechanisms

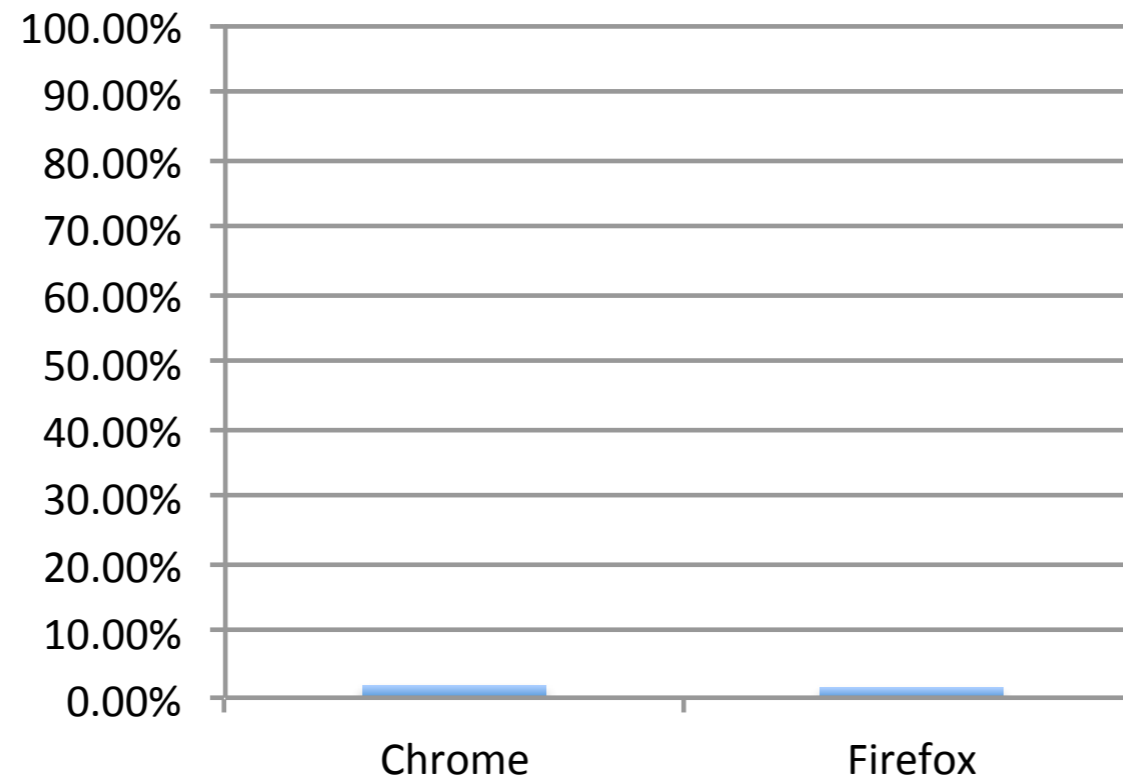
Cumulative Share of Browsers by Proportion of Opt Outs



Q: revealed preference or usability?

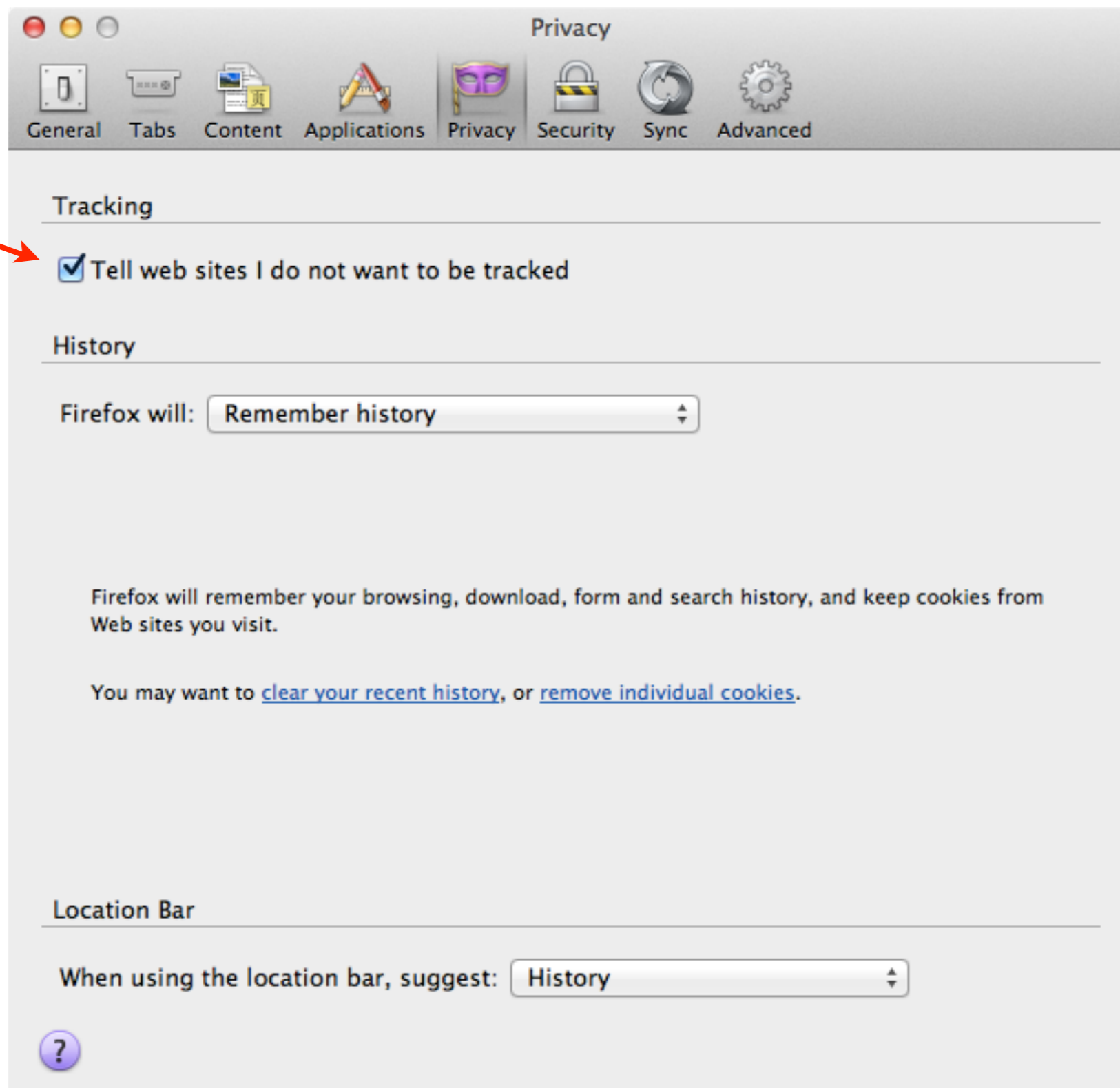
Result 2: Third-Party Cookie Blocking (U.S.)

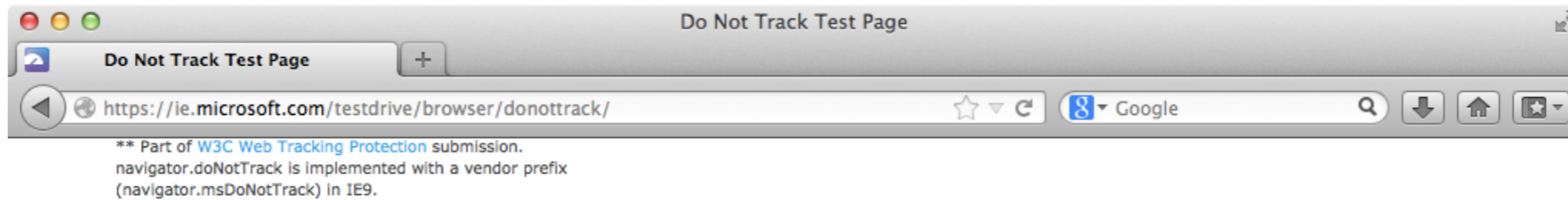
Third-Party Cookie Blocking



Supports the usability hypothesis

Result 3: Do Not Track (U.S.)





Technical Details

The DOM properties and methods below can be tested from Javascript, for example:

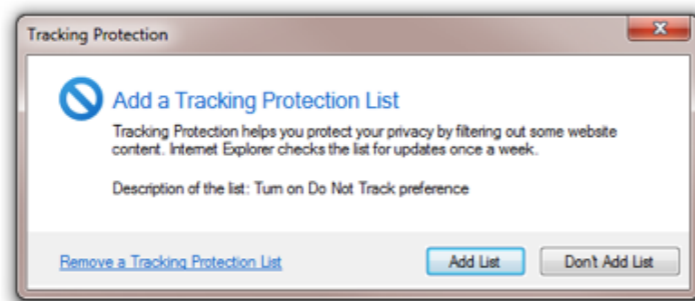
```
var tpEnabled = window.external.msTrackingProtectionEnabled();
```

The DNT HTTP header requires a small bit of server code in order to be accessible from script:

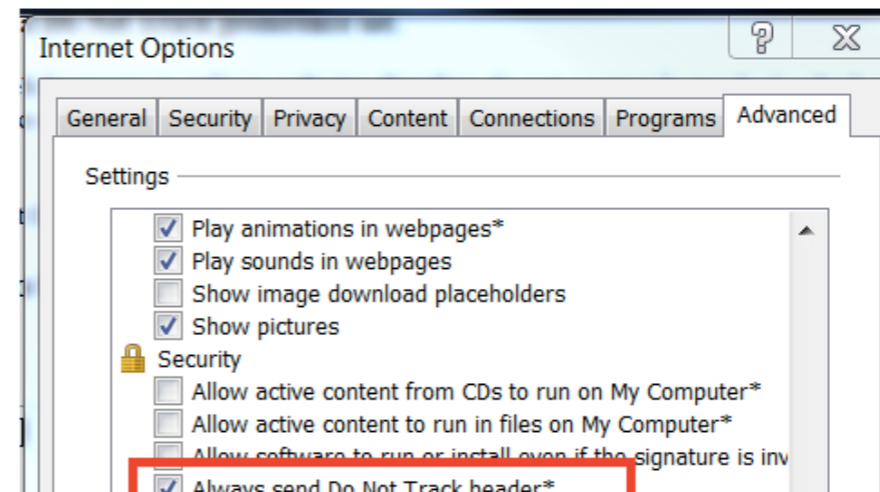
```
var dntValue = "<%= Request.ServerVariables["HTTP_DNT"] %>";
```

To express your preference not to be tracked in IE9

1. [Click here to add an empty Tracking Protection list](#)
2. When prompted, click the "Add List" button:

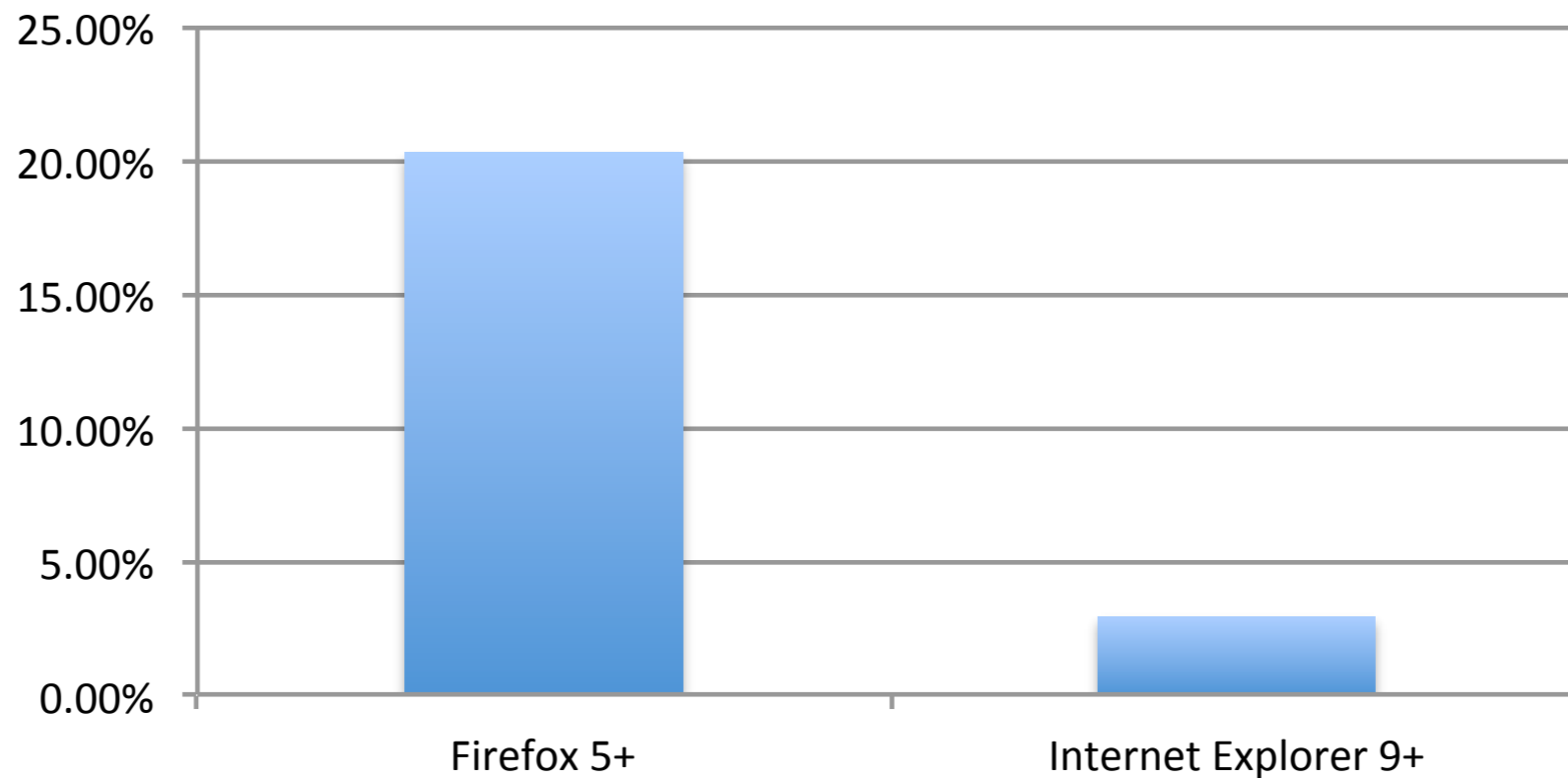


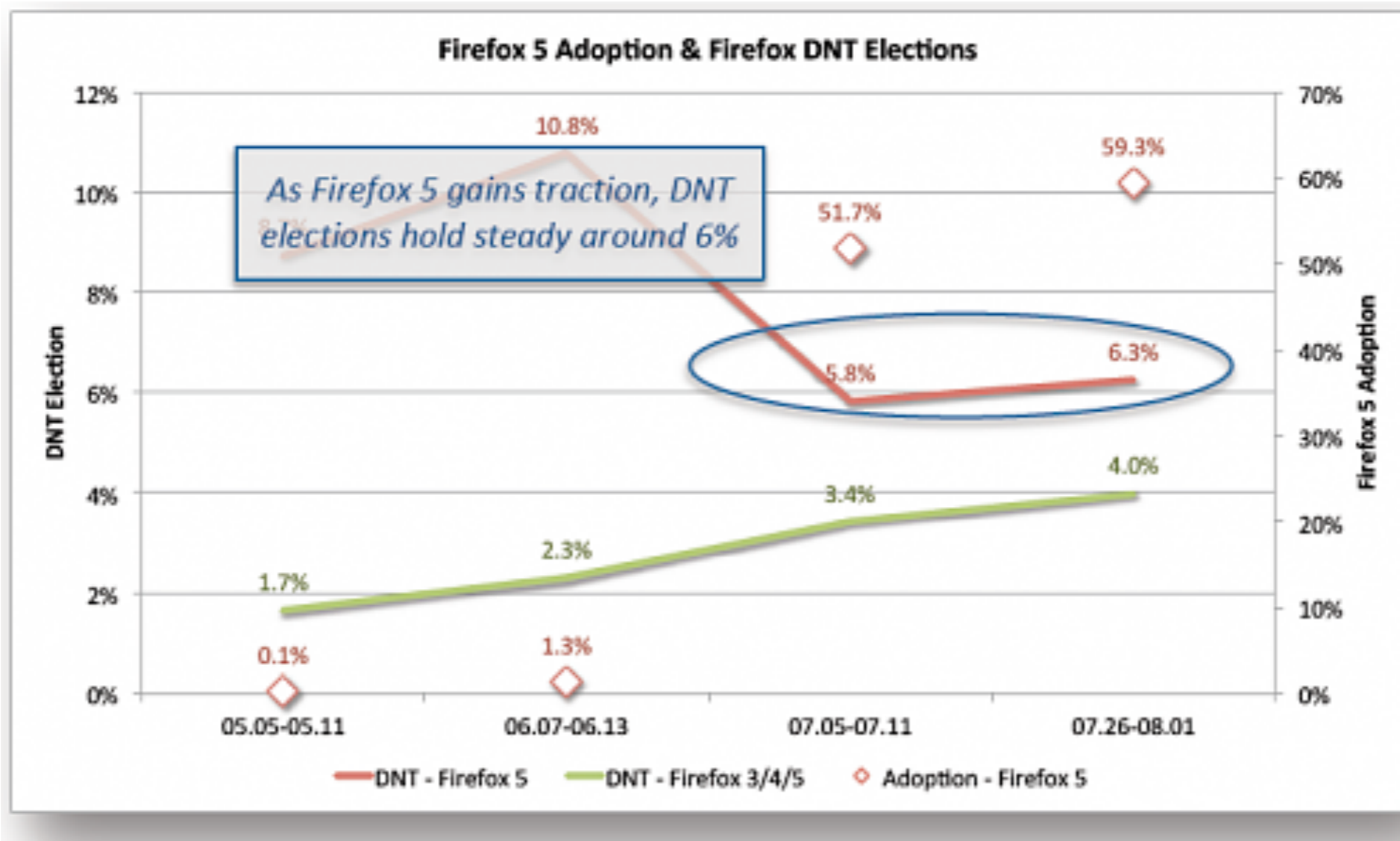
1. This preference is set by default for you in IE10
2. You can check by going to the Advanced tab under Internet Options and making sure the "Always send Do Not Track header" option is set



To
Fi

Do Not Track





Supports the usability hypothesis

History and Status Quo

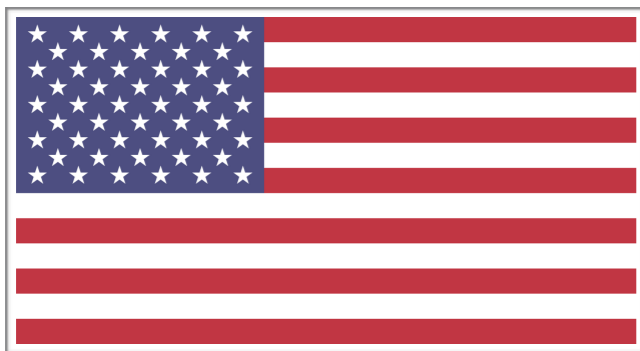
Tracking Methods

Signaling Mechanisms

Technical Countermeasures

Government Intervention

Privacy-Preserving Advertising



- FTC enforcement
- FTC proposal
- White House proposal
- Pending legislation



- Draft legislation



- ePrivacy Directive
- Article 29 Working Party opinions
- Pending legislation



- PIPEDA

History and Status Quo

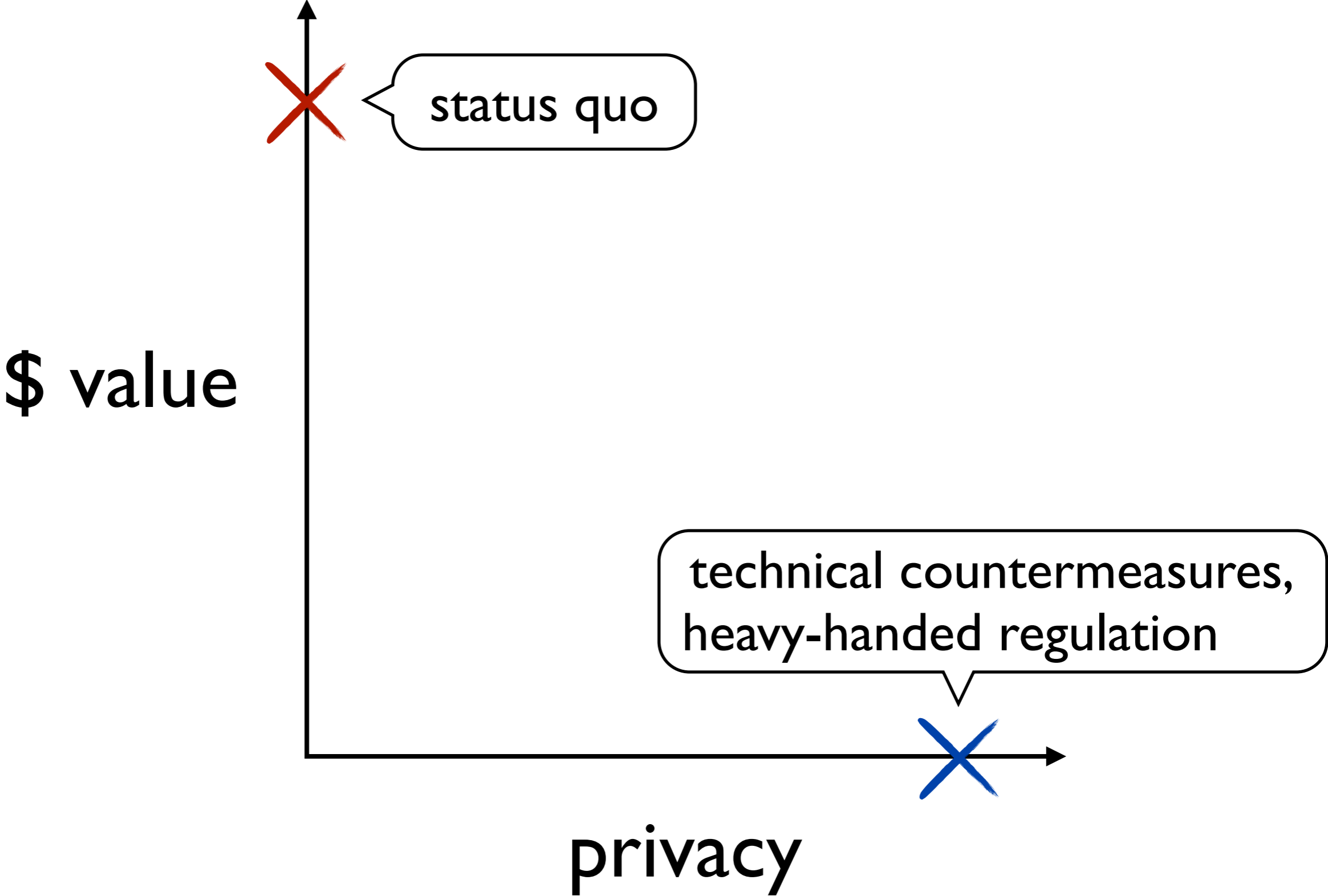
Tracking Methods

Signaling Mechanisms

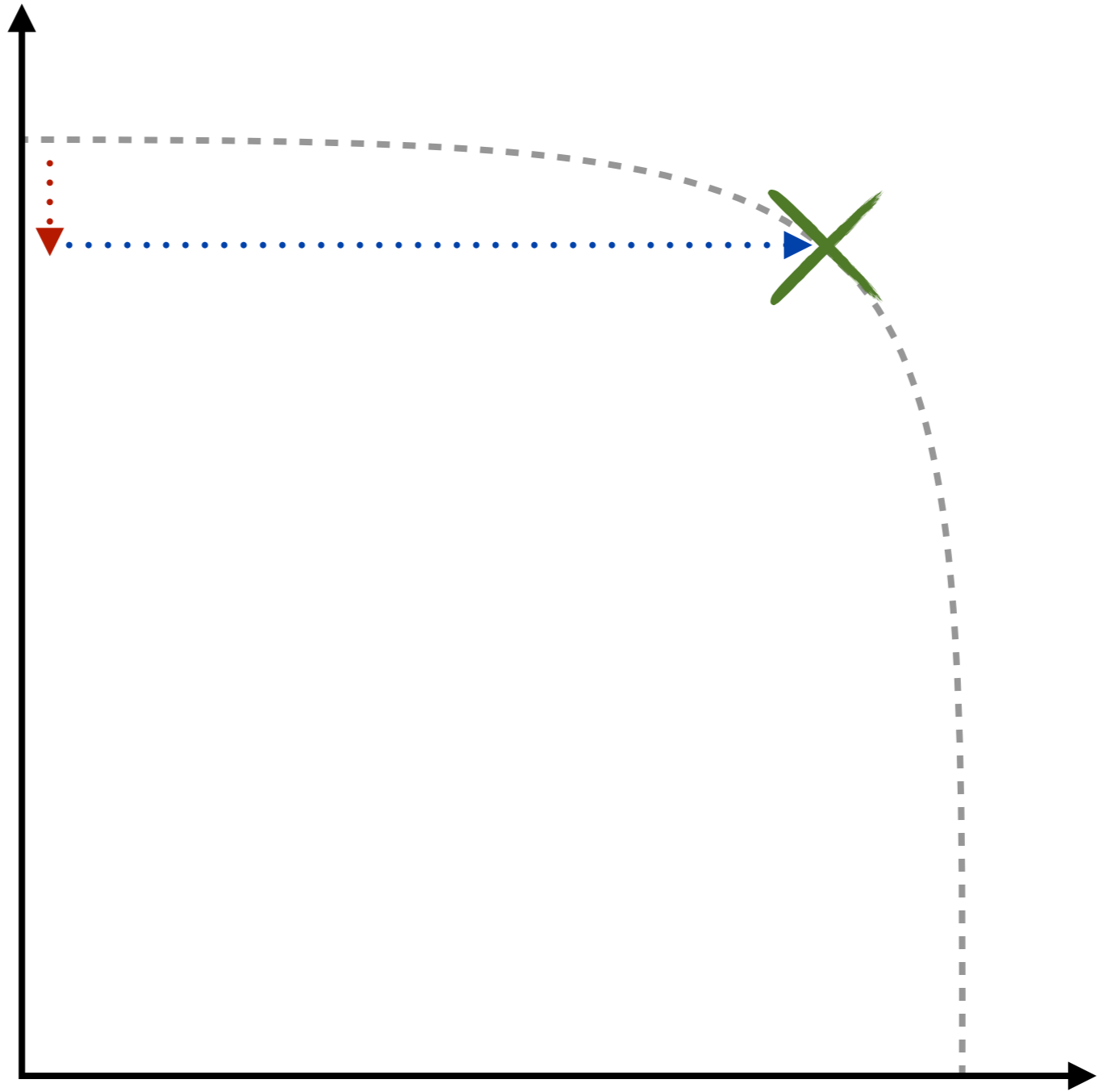
Technical Countermeasures

Government Intervention

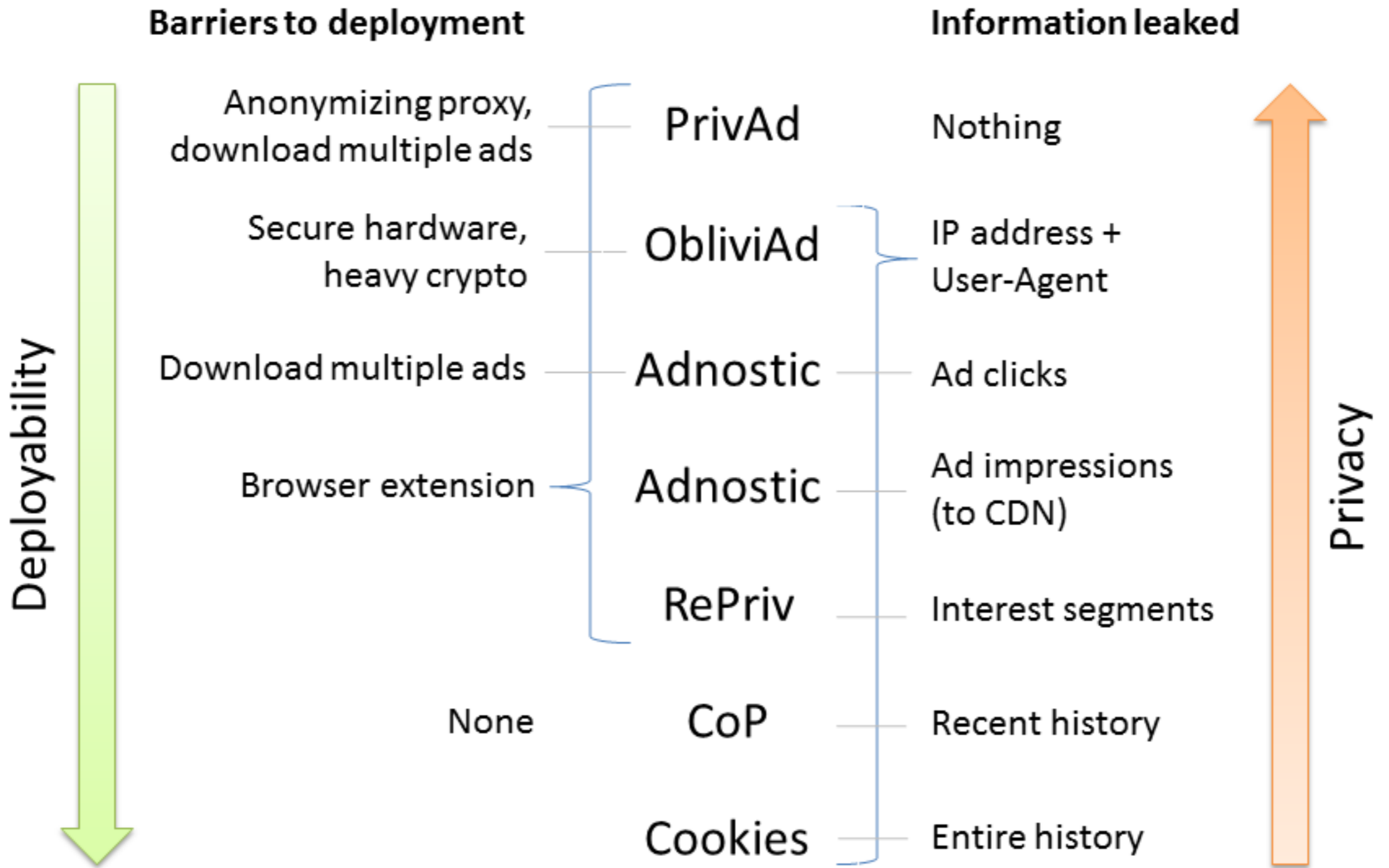
Privacy-Preserving Advertising



\$ value



privacy



engineering conventions

information asymmetries

implementation and switching costs

diminished private utility

inability to internalize

competition barriers

regulatory responses?

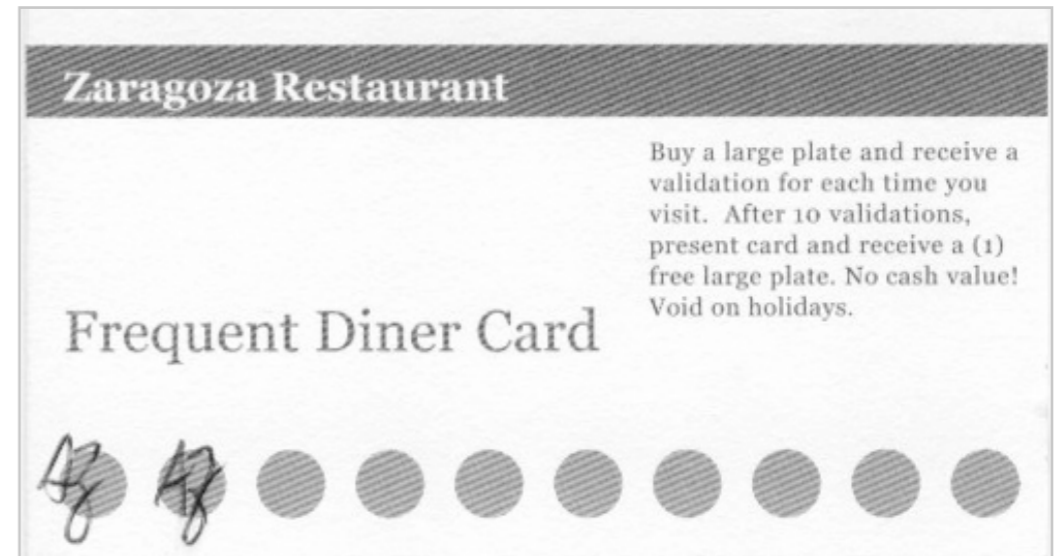
History and Status Quo
Signaling Mechanisms
Technical Countermeasures
Government Intervention
Privacy-Preserving Advertising

Questions?

jmayer@stanford.edu
stanford.edu/~jmayer

Tracking Not Required

Privacy-Preserving Third-Party Web Services



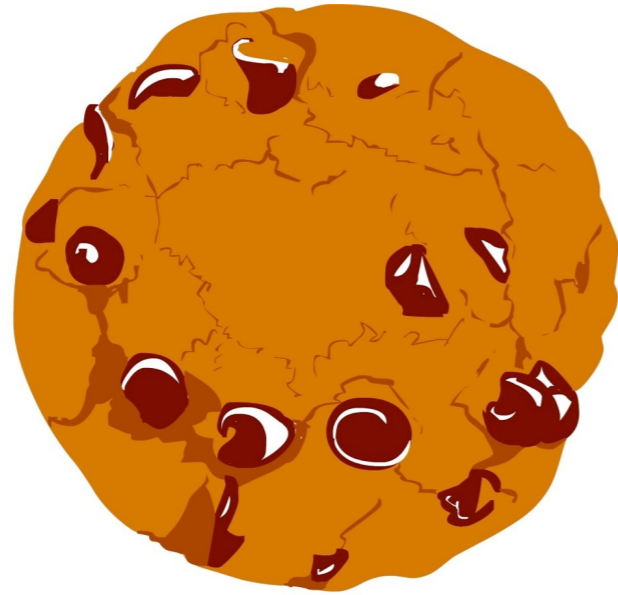


Tell me about yourself..

A black-outlined speech bubble with a tail pointing towards the bottom-left. Inside the bubble, the text "Do you come here often?" is written in a black, sans-serif font.

Do you come here often?

1. client-side storage

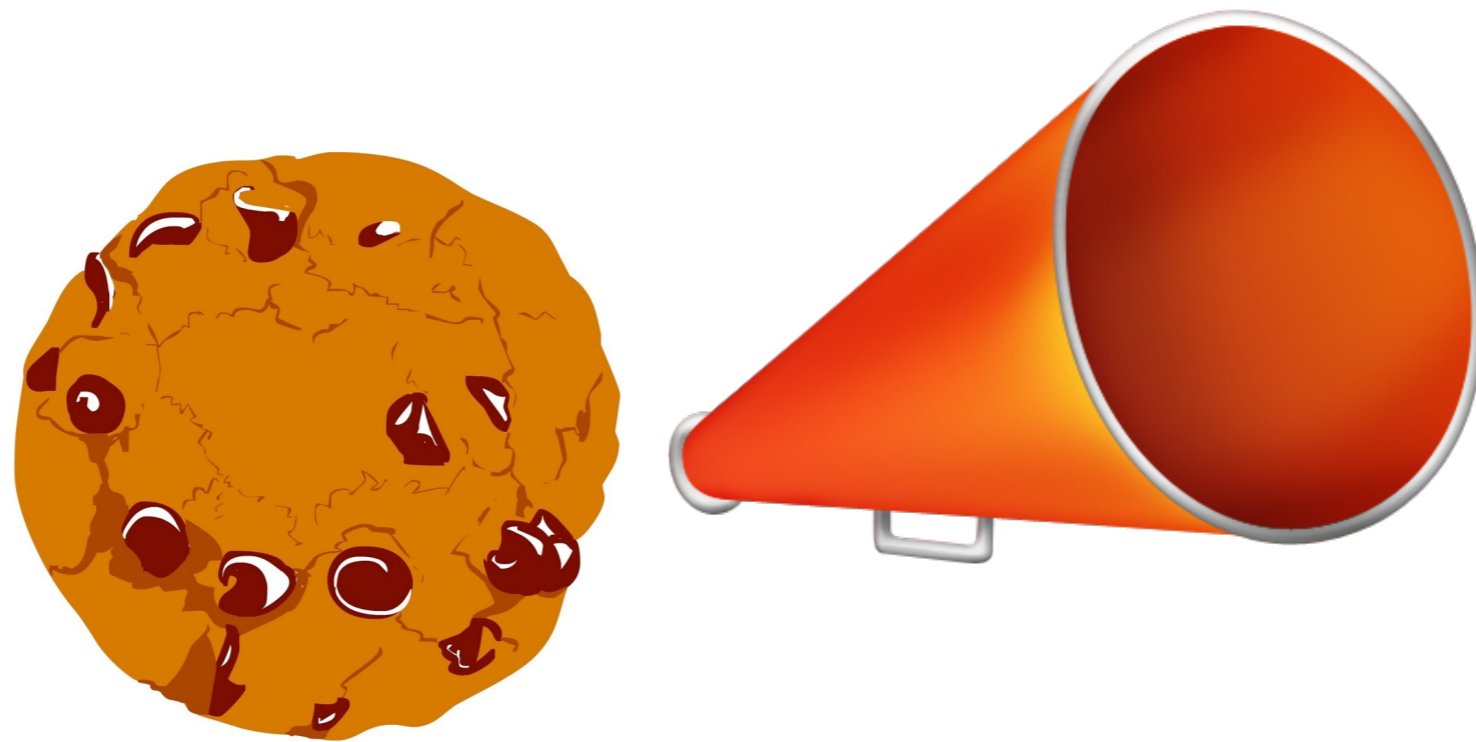


≈ 4 KB

HTML

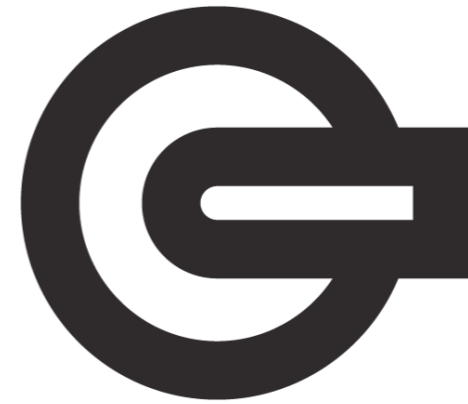


≈ 5 MB



(but domains, paths)

HTML



2. client-side logic

JavaScript

3. server-side preprocessing

any web platform

example:
advertisement frequency capping

today: ID cookies

step 1: client-side storage

| <u>Campaign</u> | <u>Count</u> | <u>Last</u> |
|----------------------|--------------|-------------|
| CarCo | 1 | today |
| Toothpaste Unlimited | 4 | yesterday |
| ... | ... | ... |

“tell me everything”

(gets unique very quickly)

step 2: client-side secrecy

| <u>Campaign</u> | <u>Count</u> | <u>Last</u> |
|----------------------|--------------|-------------|
| CarCo | 1 | today |
| Toothpaste Unlimited | 4 | yesterday |
| ... | ... | ... |

“tell me about impressions”

step 3: server-side preprocessing

| <u>Campaign</u> | <u>Count</u> | <u>Last</u> |
|----------------------|--------------|-------------|
| CarCo | 1 | today |
| Toothpaste Unlimited | 4 | yesterday |
| ... | ... | ... |

“tell me about impressions for these ads”

step 4: client-side logic

| <u>Campaign</u> | <u>Cap</u> |
|----------------------|------------|
| CarCo | 1 |
| Toothpaste Unlimited | 5 |

+

| <u>Campaign</u> | <u>Count</u> | <u>Last</u> |
|----------------------|--------------|-------------|
| CarCo | 1 | today |
| Toothpaste Unlimited | 4 | yesterday |
| ... | ... | ... |

=

| <u>Campaign</u> | <u>Capped</u> |
|----------------------|---------------|
| CarCo | yes |
| Toothpaste Unlimited | no |

“tell me whether these ads are capped”

step 5: client-side logic

| <u>Campaign</u> | <u>Cap</u> | <u>Preference</u> |
|----------------------|------------|-------------------|
| CarCo | 1 | 1 |
| Toothpaste Unlimited | 5 | 2 |

+

| <u>Campaign</u> | <u>Count</u> | <u>Last</u> |
|----------------------|--------------|-------------|
| CarCo | 1 | today |
| Toothpaste Unlimited | 4 | yesterday |
| ... | ... | ... |

=

| |
|----------------------|
| <u>Campaign</u> |
| Toothpaste Unlimited |

“tell me my most preferred uncapped ad”

state space = # of candidate campaigns

Website

- list of preference-ranked ads
- arbitrary data



Browser

arbitrary computation

- list of ads to display



- frequency capping
- frequency targeting
- behavioral targeting
- retargeting

example:
advertising conversion measurement

Campaign

Origin

Time

CarCo

news.com

this week

Toothpaste Unlimited

sports.com

last week

...

...

...

social sharing widget personalization

URL

Friend

http://gothamnews.com/batman_strikes_again

Bruce

...

...

+

Friend

Photo

Alfred



Bruce



...