Third Party Web Tracking

Amit Datta

CMU

Fall 2016

Outline

- **History and Status Quo**

- How web tracking works and ensuing concerns

- What can we do?

  - User Choice Mechanisms

    - Opt-out cookies + AdChoices

    - Do Not Track

    - Blocking trackers

  - Government Regulation

  - Privacy Preserving Advertising

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.
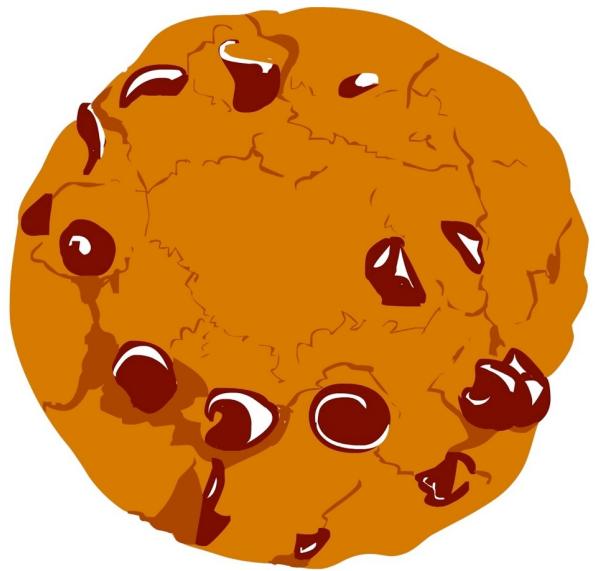
1992

That didn't last long.

Cookies

1994

JavaScript

1995
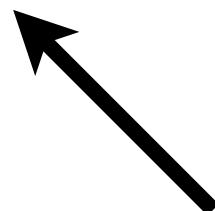
"A user agent should make ~~every~~ no attempt to prevent the sharing of session information between hosts that are in different domains."

-IETF RFC 2109

1997

Current Page

DISPLAY LUMAscape

MARKETER

PUBLISHER

PEOPLE

Agencies

WPP, OmnicomGroup, PUBLICIS GROUPE, IPG, dentsu, HAVAS, MERKLE, MDC Partners, AEGIS, AKQA, LBi, ROSETTA, icrossing, Innovation INTERACTIVE, camelot, Mediasmith, at&t Interactive, IBM, essence, yodle, REACHLOCAL, WebVisible

Agency Trading Desks

XAXIS, accuen, vivaki, RUN, CADREON, amnet, accordant media, AFFIPERF, IMPAQT, VARICK MEDIA MANAGEMENT

DSPs

MediaMath, invitemedia, TURN, DataXu, EfficientFrontier, theTradeDesk, Chango, Simpli.fi, sitescout, DIGILANT, acuity, AdBuyer.com, CTRL/SHIFT, BRANDSCREEN, choozle

AppNexus

Exchanges

doubleclick by Google, RIGHTMEDIA, Facebook Exchange, AdECN, OpenX, ZENOVIA, PLYmedia, PULSEPOINT, adBrite, adjug

Ad Networks

Horizontal
YAHOO!, CONVERSANT, Microsoft Media Network, Google, adconion MEDIA GROUP, Undertone, amazon media group, Aol., 24/7 MEDIA, audience network, xertive media, Tribal FUSION, CPX, INDEX EXCHANGE, SOLVE MEDIA, AT&T AdWorks, ADTEGRITY.COM, Brand.net, BurstMedia, ENGAGE, B, traffic marketplace, 5:1, UNDERDOG Media

Video / Rich Media
SAY, TREMOR VIDEO, videology, YuMe, BBE, TubeMogul, BrightRoll, spotXchange, AdoTube, smartclip, gumgum, BREAK Media, Jambo, meebo, ScanScout, OGGIFINOGI, MATOMY MEDIA GROUP

Vertical / Custom
ravora, ADIFY, EVOLVE MEDIA, GourmetAds, MOB SOC, LONGBOARD MEDIA, RGM ALLIANCE, jumpstart, Sportgenic, NetShelter, MODE MEDIA, MAXPOINT, martini media, quadrantONE, Q1Media, InflectionPointMedia, IDGTechNetwork, SOJERN, intergi entertainment, technorati, Ideas People, brightcom, TongFuse, FEDERATED MEDIA

Targeted Networks / AMPs
rocketfuel, Adgorithms, interclick, Collective, dstillery, Taykey, precisiondemand, cs, 33across, VIANT, radiumone, ownerIQ, resonate networks, true[x] media, Adblade, adadyn, NetSeer, yp, pulse 360, adaramedia, pearset, inadco, IGNITIONONE, XGRAPH, CONTEXTWEB, rockerbox

Performance
epic, vantagemedia, MATOMY MEDIA GROUP, advertise.com, convertmedia, moxy media, commission junction, VigLink, Kontera, Madison Logic, VIBRANT, LinkShare, phg, adknowledge, MARCHEX, Fluent, hydra, neverblue, Datran Media, MediaTrust, affinity, TattoMedia, TheUseful, sourceknowledge

Mobile
AdMob, crossboard, millennialmedia, Quattro WIRELESS, Jumptap, mocean mobile, GREYSTRIPE, rhythm, INMOBI, transpera

Sharing Data / Social Tools
facebook, Twitter, ShareThis, AddThis, tynt, GIGYA

SSPs
Admeld, rubicon project, PubMatic, LIFT DNA, Beanstock, ReviNet, ALTITUDE DIGITAL, sonobi

Publisher Tools
yieldex, fatTail, ShortTail, iSOCKET, SHINY ADS, npi, komoona, sourn, BuySellAds, adomik, Adslot media, yieldbot, BuzzFeed, DailyMe, scout analytics, daylife, Mochila, MAXIFIER, PerfectMarket, Parse.ly, Outbrain, Jump Time, Taboola, Gravity, Luminate

Ad Servers (publisher)
doubleclick by Google, 24/7 MEDIA, apt from Yahoo!, LiveIntent, adgear, ADTECH, OpenX, ZEDO, AdJuggler, adzerk, aiMatch

Creative Optimization
teracent, TUMRI, dapper, AdExtent, Impossible Software, adroit, struq, MIXPO, paper g, SPONGECELL, engageclick, ipromote, adacado, Dynamix, snapads, CERTONA, AdReady, richrelevance, ADISN

Media Planning and Attribution
MarketShare, COGNITIVE MATCH, DataSong, M FACTOR, NETWORKED INSIGHTS, CONVERTRO, colspace, SQAD, TUNE, visual iq, adometry, C3 Metrics, OPTIMINE, MARKETING EVOLUTION, ThinkVine, Encore, ConversionLogic, Korrelate

Tag Mgmt
TagMan, ensighten, TEALIUM, SIGNAL

Measurement and Analytics
OMNITURE, unica, Coremetrics, tracksimple, webtrends, Google Analytics, nielsen, METAMARKETS, comSCORE, bi:science, MOAT, WHAT'SWHERE!, vizu, PATHMATICS, ijento, LUCID, aprimo

DMPs and Data Aggregators
bluekai, x+1, aka, Demdex, nPARIO, exelate, brilig, quantcast, krux, AUDIENCE SCIENCE, Red Aril, bizo, LOTAME, datonics, jovianDATA, upfront, LiveRamp, acerno, CIRCULATE, grapeshot, proximic, bombora, PEER39, ADmantX

Data Suppliers
Experian, acxiom, datalogix, NUUVA, EPSILON, TARGUSinfo, TruSignal, NETEZZA, MasterCard, WhatRunsWhere, bluecava, MERITDIRECT, Xi group, infogroup, FGK, V12 group, DATALINE, IRi

Ad Servers (marketer)
doubleclick by Google, atlas, Mediaplex, FLITE, mediamind, Sizmek, pictela, pointroll, eyeWonder, flashtalking, trueffect, adform, adgear

Verification / Privacy
DoubleVerify, GHOSTERY, Integral Ad Science, Ad-Juster, TRUST METRICS, THE MEDIA TRUST, ADXPOSE, privacychoice, TRUSTe

Retargeting
dotomi, criteo, MAGNE+IC, AdRoll, TellApart, TRIGGIT, sociomantic, crosspixelmedia, MYBUYS, FetchBack, buysight, PERFECT AUDIENCE, ReTargeter, STEELHOUSE, myThings, mediaFORGE

Media Mgmt Systems and Operations
DDS, operative, telmar, Centro, media x, theorem, MEDIABANK, facilitate, TRAFFIQ, BIONIC, Solbright, AdPro, SQAD

Denotes acquired company   Denotes shuttered company   © LUMA Partners LLC 2016

LUMA

Outline

- History and Status Quo
- **How web tracking works and ensuing concerns**
- What can we do?
  - User Choice Mechanisms
    - Opt-out cookies + AdChoices
    - Do Not Track
    - Blocking trackers
  - Government Regulation
  - Privacy Preserving Advertising

stateful tracking

tagging

stateless tracking

fingerprinting

HTTP cookies

HTTP authentication

Flash Local Shared Objects

Silverlight Isolated Storage

HTTP ETags

TLS session ID & resume

content cache

browsing history

IE userData

window.name

HTML5 protocol & content handlers

HTTP STS

HTML5 session/local/global/database storage

DNS cache

Source: [Aggrawal10]
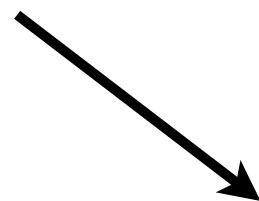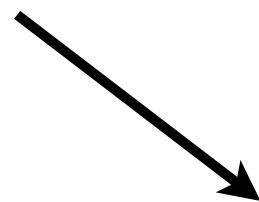
this is blue
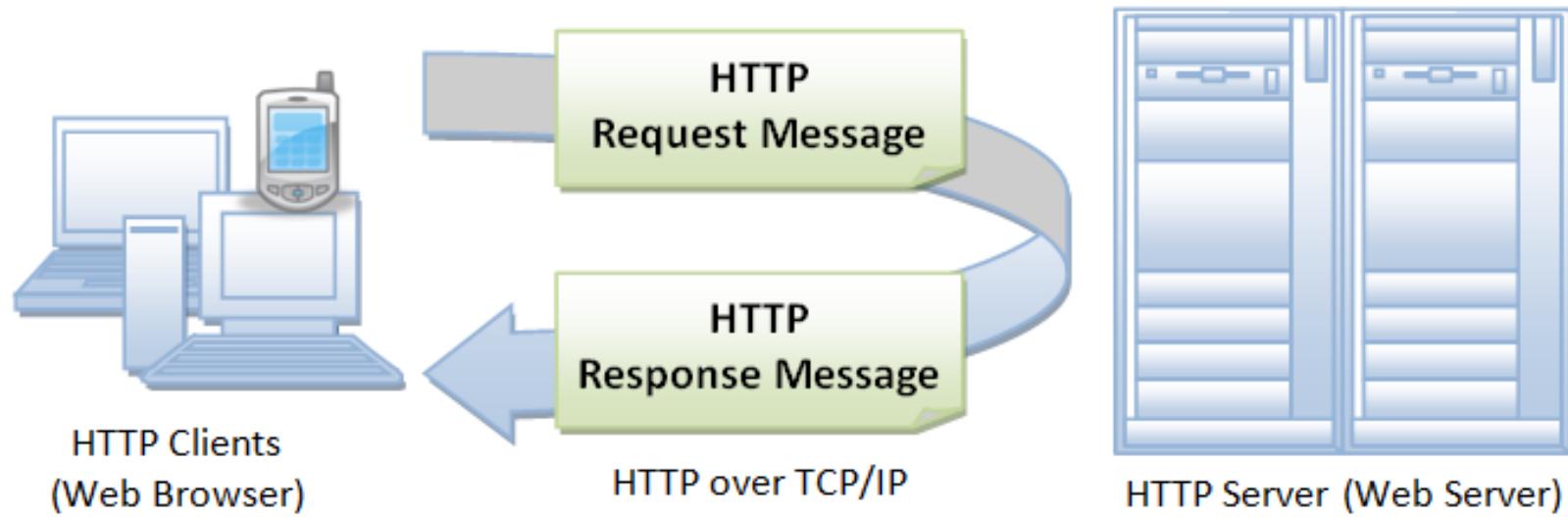
link

this is purple
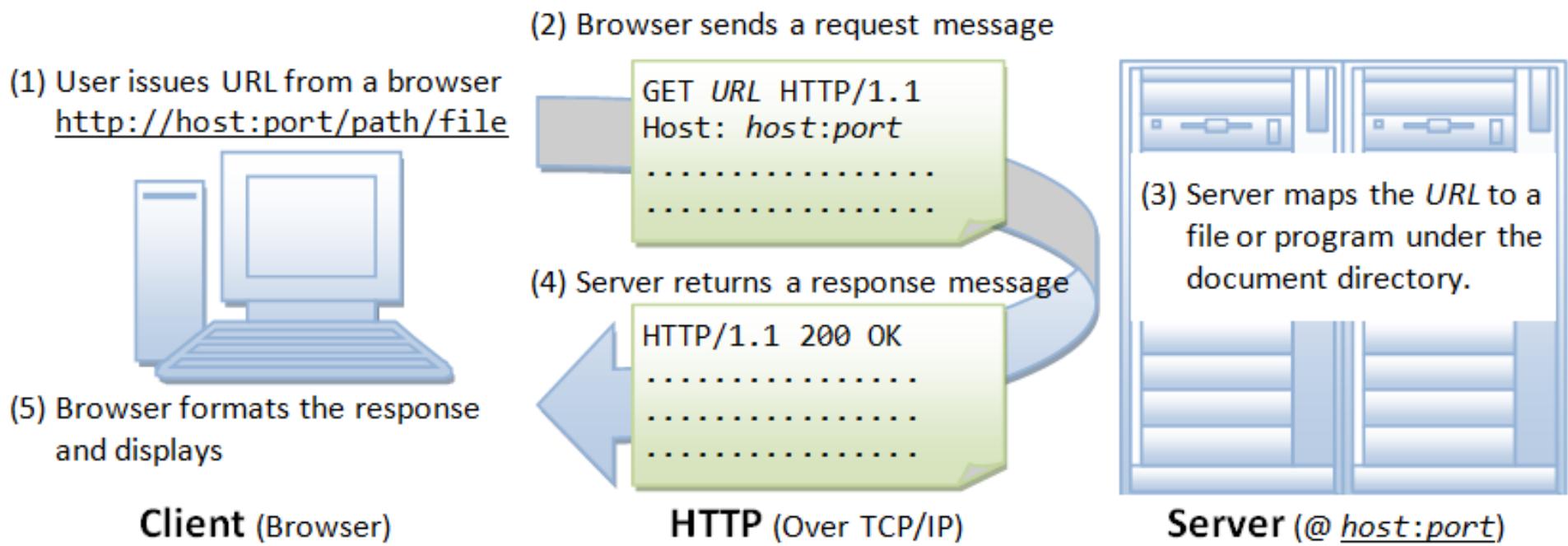
link

HTTP Basics

HTTP Basics

(1) User issues URL from a browser
http://host:port/path/file

(2) Browser sends a request message

```
GET URL HTTP/1.1
Host: host:port
.................
.................
```

(4) Server returns a response message

```
HTTP/1.1 200 OK
.................
.................
.................
```

(5) Browser formats the response and displays

(3) Server maps the URL to a file or program under the document directory.

Client (Browser)

HTTP (Over TCP/IP)

Server (@ host:port)

Set-Cookie

- First interaction between user and server:

User request:

GET /index.html HTTP/1.1
Host: www.example.org

Server response : places cookie in user's browser

HTTP/1.0 200 OK
Content-type: text/html
Set-Cookie: theme=light
Set-Cookie: sessionToken=abc123; Expires=Wed, 09 Jun 2021 10:18:14 GMT

Set-Cookie

- Subsequent interactions between browser and server:

    Browser sends cookie + other info to server

GET /spec.html HTTP/1.1

Host: www.example.org

Cookie: theme=light; sessionToken=abc123

Third Party Cookies

- There is **no intrinsic difference** between a first-party cookie and a third-party cookie.
- The distinction only exists within the context of a particular visit.

🍪

website: news.com
cookie: news.com

First party cookie

🍪

website: news.com
cookie: advertising.com

Third party cookie

<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	http:// foxnews.com/...	Why Liberals Hate America
12345	6/18/12 10:02am	http:// youtube.com/...	Squirrels Waterskiing?!

...

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)

HTTP cookies

HTTP authentication

Flash Local Shared Objects

Silverlight Isolated Storage

HTTP ETags

TLS session ID & resume

content cache

browsing history

IE userData

window.name

HTML5 protocol & content handlers

HTTP STS

HTML5 session/local/global/database storage

DNS cache

Source: [Aggrawal10]

Super Cookies

- Evercookie library

http://samy.pl/evercookie/

Stateless Tracking

User-Agent

installed fonts

HTTP ACCEPT Headers

cookies enabled?

browser plug-ins

browser add-ons

MIME support

screen resolution

clock skew

Sources: [Eckersley10], [Mayer09]
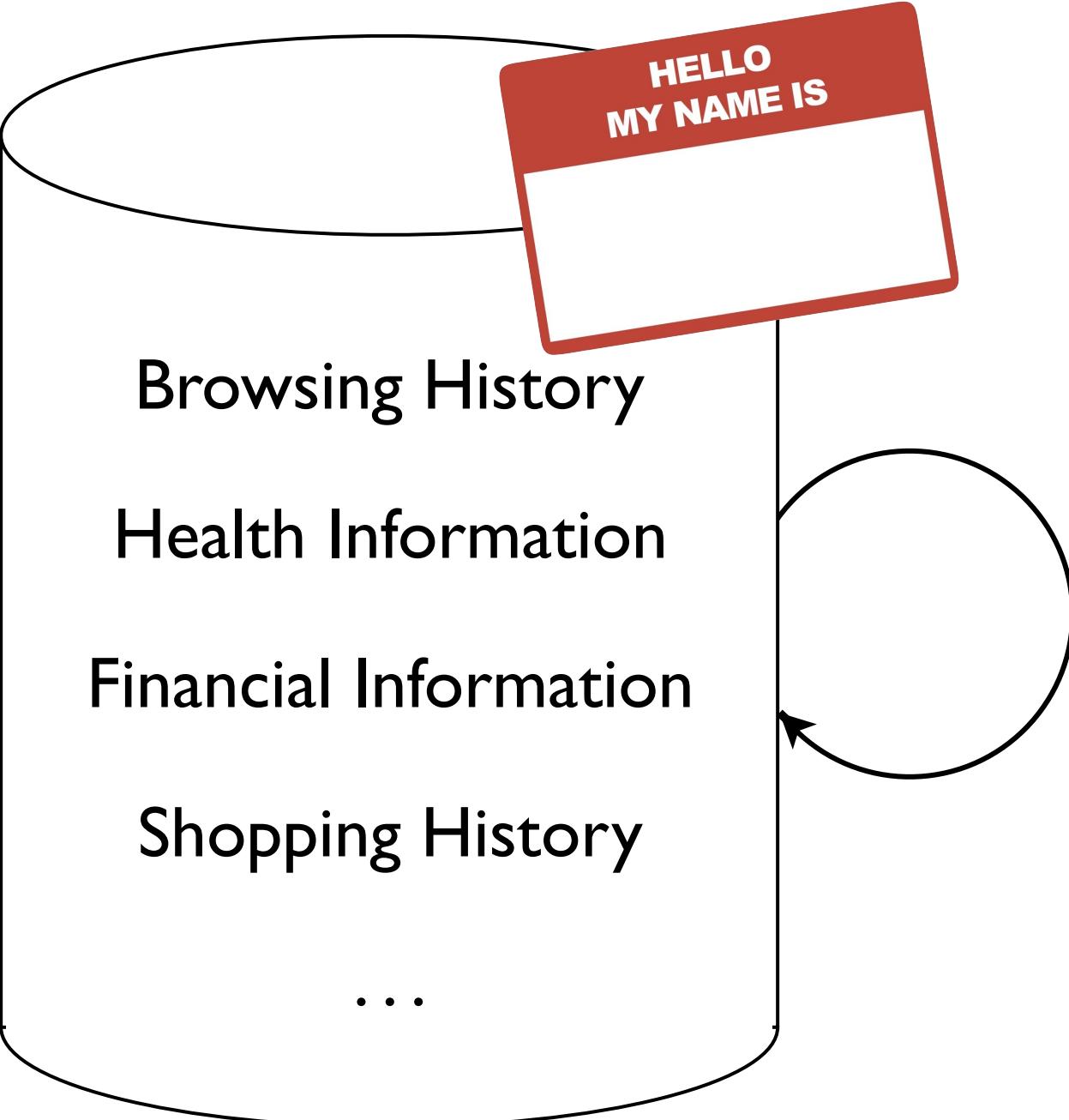
Browser Fingerprinting

- EFF's Panopticlick

https://panopticlick.eff.org/

Peter Eckersley on browser fingerprinting:

https://panopticlick.eff.org/browser-uniqueness.pdf

Browsing History

Health Information

Financial Information

Shopping History

…

HELLO
MY NAME IS

"it's all anonymous"
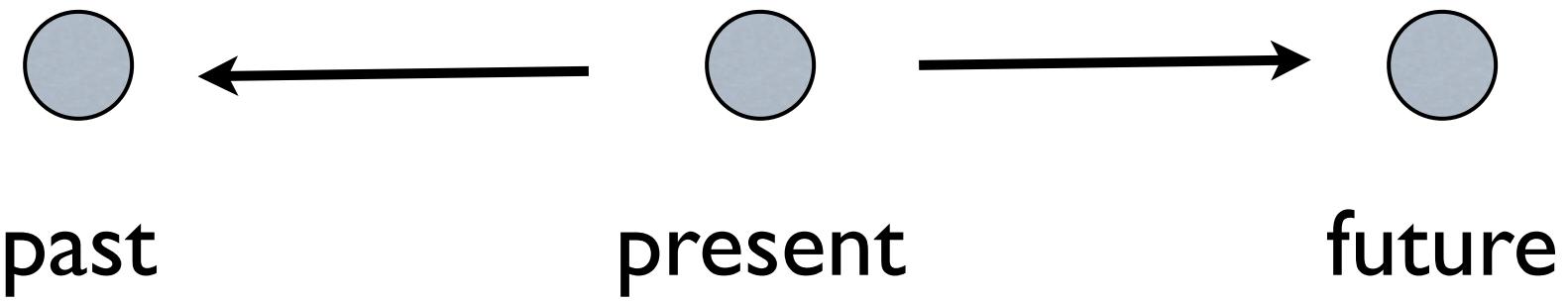
actually, it's all pseudonymous

The Rock

Dwayne Johnson

past   present   future

Concerns

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

Cross Device Tracking

- Deterministic
  - Based on user logins
  - Facebook, Google

- Probabilistic
  - Based on location, browsing habits, etc.
  - Drawbridge, Tapad, BlueCava

Outline

- History and Status Quo
- How web tracking works and ensuing concerns
- **What can we do?**
  - User Choice Mechanisms
    - Opt-out cookies + AdChoices
    - Do Not Track
    - Blocking trackers
  - Government Regulation
  - Privacy Preserving Advertising

Opt-out Cookies
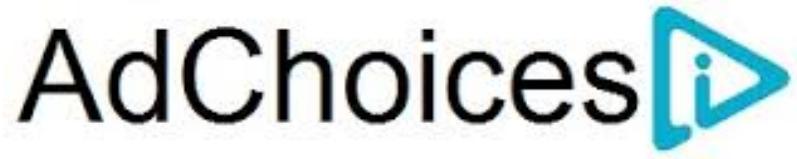
- Users opt out by installing opt-out cookies
- Anecdotal reports: usage < 1% in browsers
- Usability: low

Network Advertising Initiative

102 companies

http://www.networkadvertising.org/choices

AdChoices

- Icon + text in display ads
- Explanation for why ad was served
- Landing page for user to set opt-out cookies
- Opt out of behavioral targeting not tracking
- Alexa US top 500 web sites:
    Icon in 9.9% of ads; icon + text in 5.1% of ads

DIGITAL ADVERTISING ALLIANCE

128 companies

http://www.aboutads.info/choices/

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

Do Not Track

- Users specify via browser settings

- Implemented through an HTTP header field
  - DNT: 1

- W3C standardization effort ongoing

4+   9+   5.1+   25+

Privacy

General | Bookmarks | Tabs | AutoFill | Passwords | Security | Privacy | Notifications | Extensions | Advanced

Cookies and other website data: [Remove All Website Data...]

Block cookies:
- ⦿ From third parties and advertisers
- ○ Always
- ○ Never

Limit website access to location services:
- ⦿ Prompt for each website once each day
- ○ Prompt for each website one time only
- ○ Deny without prompting

Website tracking: ☑ Ask websites not to track me

Web search: ☐ Prevent search engine from providing suggestions ⑦

Mozilla Firefox Start Page   Preferences   +

Firefox | about:preferences#privacy

Most Visited   Getting Started

Privacy   ?

Tracking

☑ Use Tracking Protection in Private Windows   Learn more   Change Block List

You can also manage your Do Not Track settings

History

Firefox will:

Firefox will

You may wa

Do Not Track   ✕

☑ Always apply Do Not Track

Firefox will send a signal that you don't want to be tracked whenever Tracking Protection is on. Learn More

Cancel   OK

Location Ba

When using

☑ History

☑ Bookmarks

☑ Open tabs

Change preferences for search engine suggestions...

Sidebar
General
Search
Content
Applications
Privacy
Security
Sync
Advanced

Settings | x

chrome://settings

Chrome

History

Extensions

Settings

Help

Settings

Search 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

Chrome

Settings

Search 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 s
☑ Use a predic
☑ Predict netw
☑ Enable phis
☐ Use a web s
☐ Automatica
☑ Send a 'Do

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...
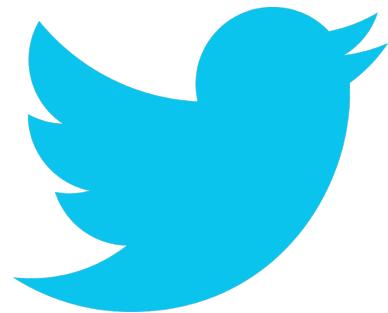
Page zoom: 100% ▾

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

tens of millions of users

≈25 third parties honor DNT

W3C

Negotiations for 2+ years, no agreement.

Enforcing these signaling mechanisms

✓ ● observe suspicious behavior

● monitor ad distributions

Sources: [Jang10], [Guha10]

anti-tracking technology

blocking

```
||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.heraldm.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?
```

Ghostery

# Privacy Badger

https://www.eff.org/privacybadger
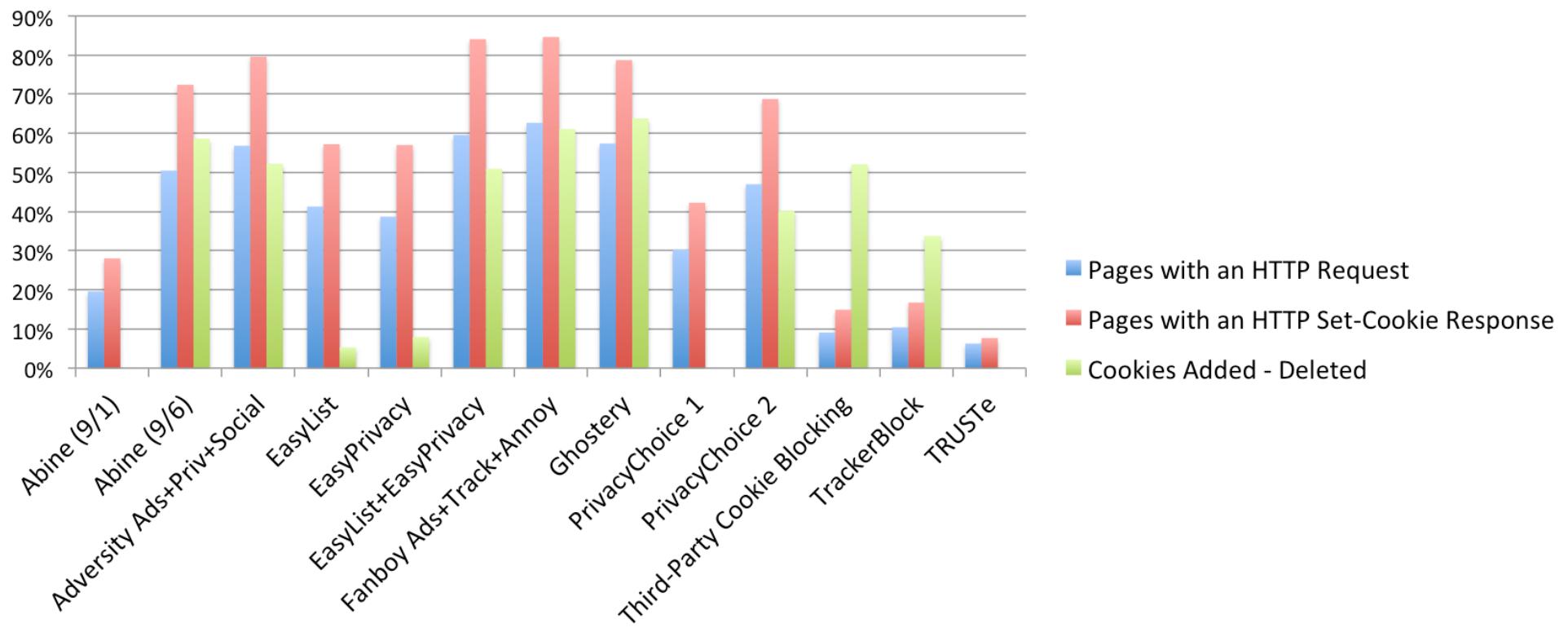
- not comprehensive

- requires updating

- breaks stuff

- requires user knowledge about providers

Evaluation of Blocking tools

Study with FourthParty of 11 blocking tools

- Crawled Alexa US top 500 sites

- For each tool, measured average across all trackers relative to baseline

  - Pages with HTTP request

 - Pages with HTTP Set-Cookie response

  - Cookies added - deleted
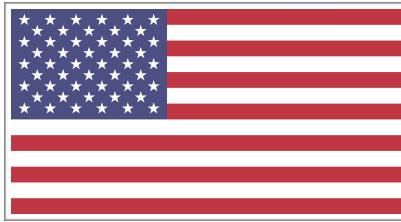
Average Decrease in Tracking

Legend:
- ■ Pages with an HTTP Request
- ■ Pages with an HTTP Set-Cookie Response
- ■ Cookies Added - Deleted

Y-axis: 0% to 90%

X-axis categories: Abine (9/1), Abine (9/6), Adversity Ads+Priv+Social, EasyList, EasyPrivacy, EasyList+EasyPrivacy, Fanboy Ads+Track+Annoy, Ghostery, PrivacyChoice 1, PrivacyChoice 2, Third-Party Cookie Blocking, TrackerBlock, TRUSTe

Evaluation of Blocking tools

- Findings
  - Most effective tool: community-maintained Fanboy's lists
  - All top performing tools blocked third-party advertising
  - Block list from TRUSTe least effective + overrides other lists to allow tracking by several sizable third parties

Government Intervention

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

- Draft legislation
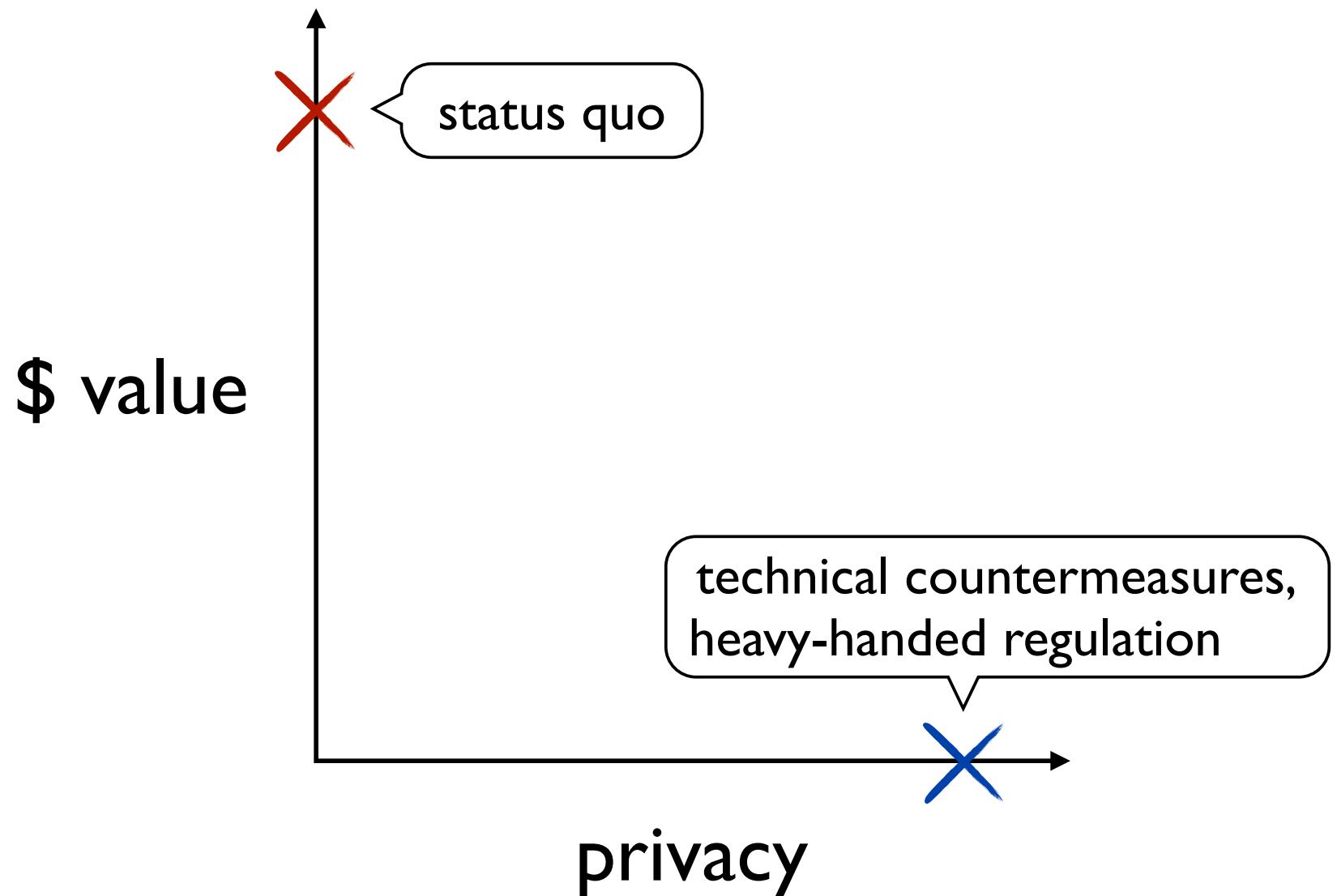
- ePrivacy Directive
- Article 29 Working Party opinions
- **GDPR**
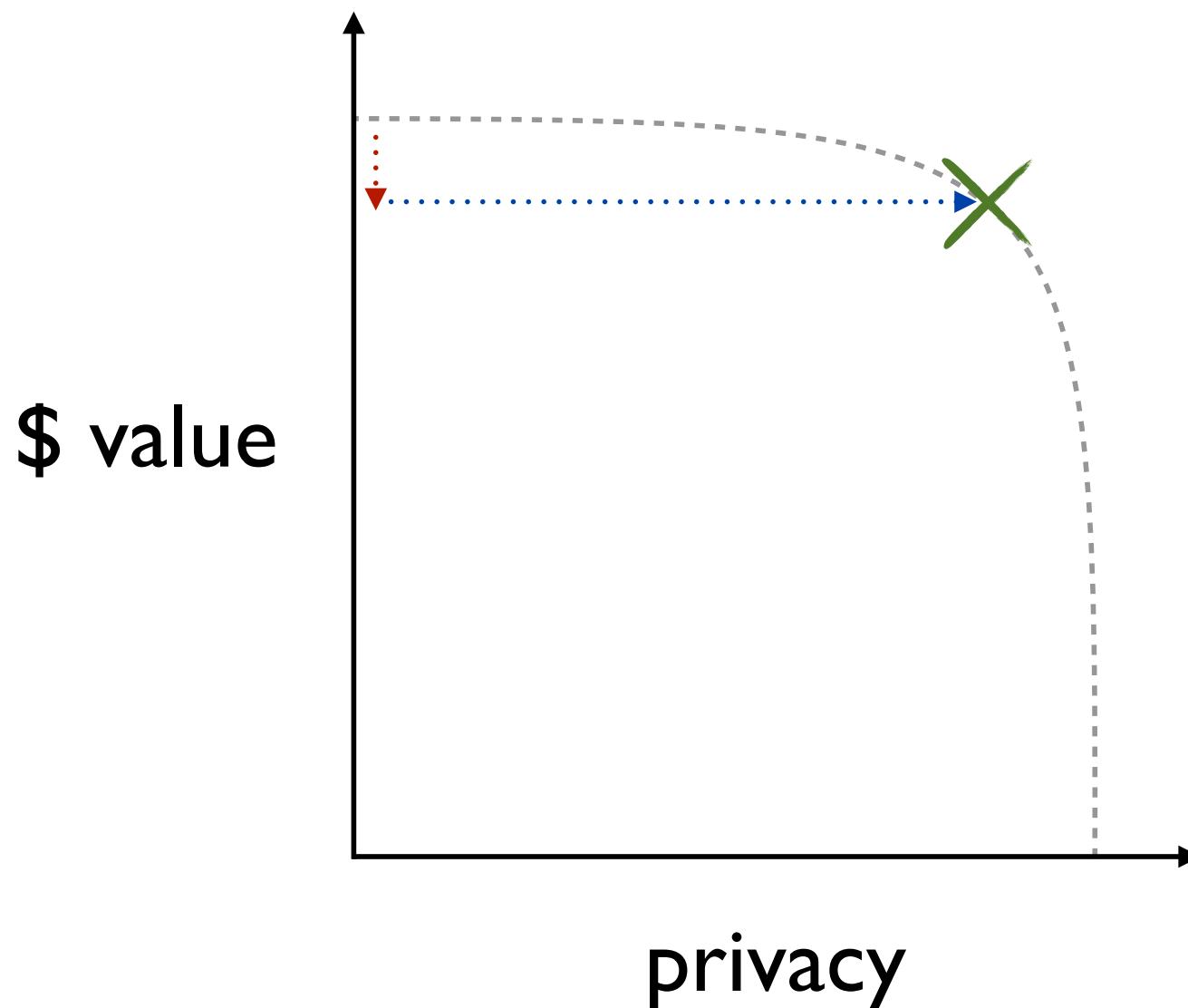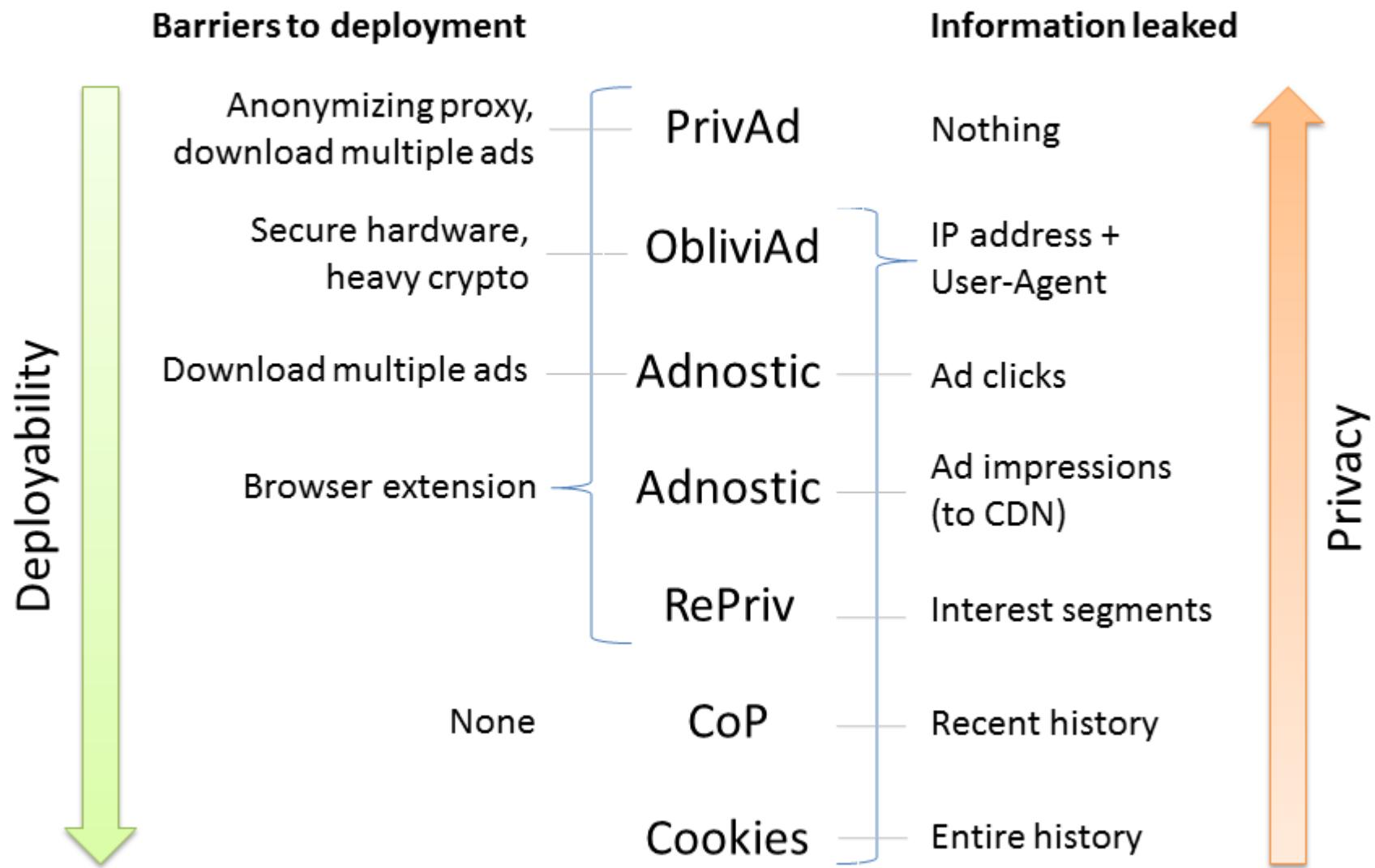
- PIPEDA

Privacy Preserving Advertising

Barriers to deployment

Information leaked

Deployability

Anonymizing proxy,
download multiple ads — PrivAd   Nothing

Secure hardware,
heavy crypto — ObliviAd   IP address +
User-Agent

Download multiple ads — Adnostic   Ad clicks

Browser extension — Adnostic   Ad impressions
(to CDN)

RePriv   Interest segments

None   CoP   Recent history

Cookies — Entire history
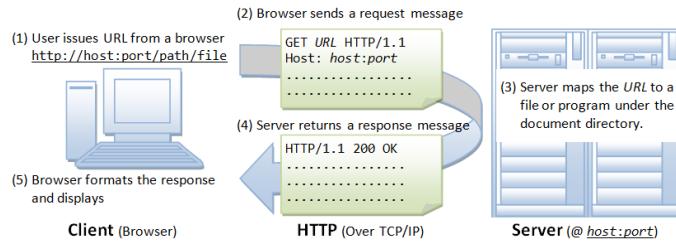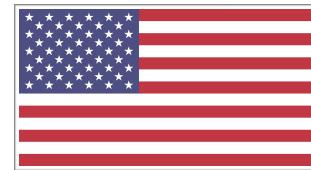
Privacy

What did we learn today?

- How web tracking works

- Concerns with tracking; how to address them

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

18734: Foundations of Privacy

Third Party Web Tracking

Amit Datta

CMU

Fall 2016