Session 4: Online Behavioral Advertising

Anonymity & Autonomy: Evidence from Google's 2012 Privacy Policy Change

James C. Cooper*
Antonin Scalia Law School
George Mason University

There was no financial support for this research. Cooper is affiliated with the Program on Economics & Privacy, a division of the Law & Economics Center, which receives support from numerous donors, including Google

Anonymity and Autonomy

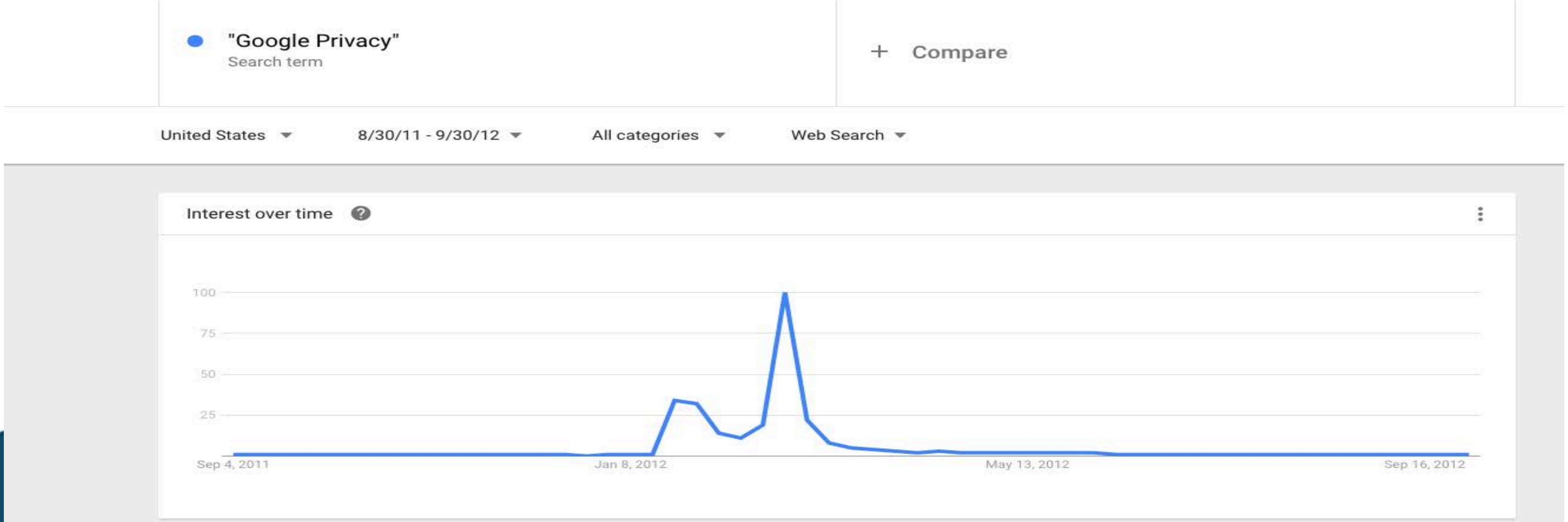
- Anonymity as an important dimension of privacy:
  - Engage in the world without actions being traced to identity
- Autonomy:
  - Zone to make private decisions free from observation or interference
- Reduction in ability to remain anonymous can reduce autonomy
- Harms:
  - Dignity
  - Personal development
  - Society
  - Privacy protective behavior

Hypothesis

- 2012 Google announced it would combine data across platforms
- At the margin, this increased view into one’s life will deter engagement in search behavior that one may want to keep private
- Measuring a reduction in autonomy due to loss in anonymity

Hypothesis

- Did people know and care?



● "Google Privacy"
Search term

+ Compare

United States ▾ 8/30/11 - 9/30/12 ▾ All categories ▾ Web Search ▾

Interest over time

- Examir arch as a result of the 201
- Google
- "Differe
  - Se
  - No
  - Loc ore and after the cha
- If mingl see an increase in the c

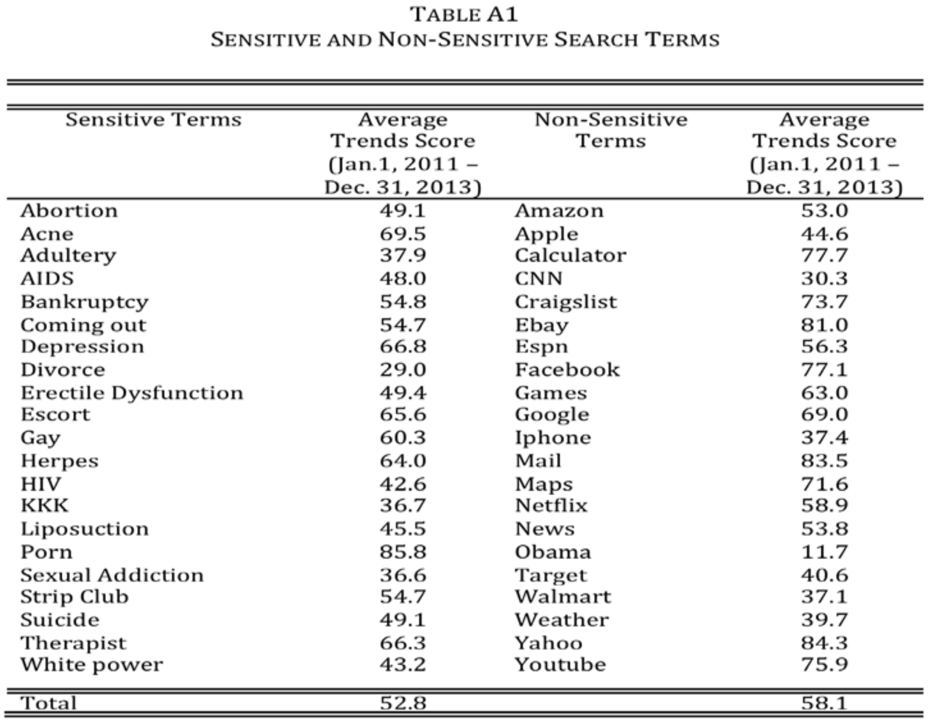


TABLE A1
SENSITIVE AND NON-SENSITIVE SEARCH TERMS

Sensitive Terms	Average Trends Score (Jan.1, 2011 – Dec. 31, 2013)	Non-Sensitive Terms	Average Trends Score (Jan.1, 2011 – Dec. 31, 2013)
Abortion	49.1	Amazon	53.0
Acne	69.5	Apple	44.6
Adultery	37.9	Calculator	77.7
AIDS	48.0	CNN	30.3
Bankruptcy	54.8	Craigslist	73.7
Coming out	54.7	Ebay	81.0
Depression	66.8	Espn	56.3
Divorce	29.0	Facebook	77.1
Erectile Dysfunction	49.4	Games	63.0
Escort	65.6	Google	69.0
Gay	60.3	Iphone	37.4
Herpes	64.0	Mail	83.5
HIV	42.6	Maps	71.6
KKK	36.7	Netflix	58.9
Liposuction	45.5	News	53.8
Porn	85.8	Obama	11.7
Sexual Addiction	36.6	Target	40.6
Strip Club	54.7	Walmart	37.1
Suicide	49.1	Weather	39.7
Therapist	66.3	Yahoo	84.3
White power	43.2	Youtube	75.9
Total	52.8		58.1

Results

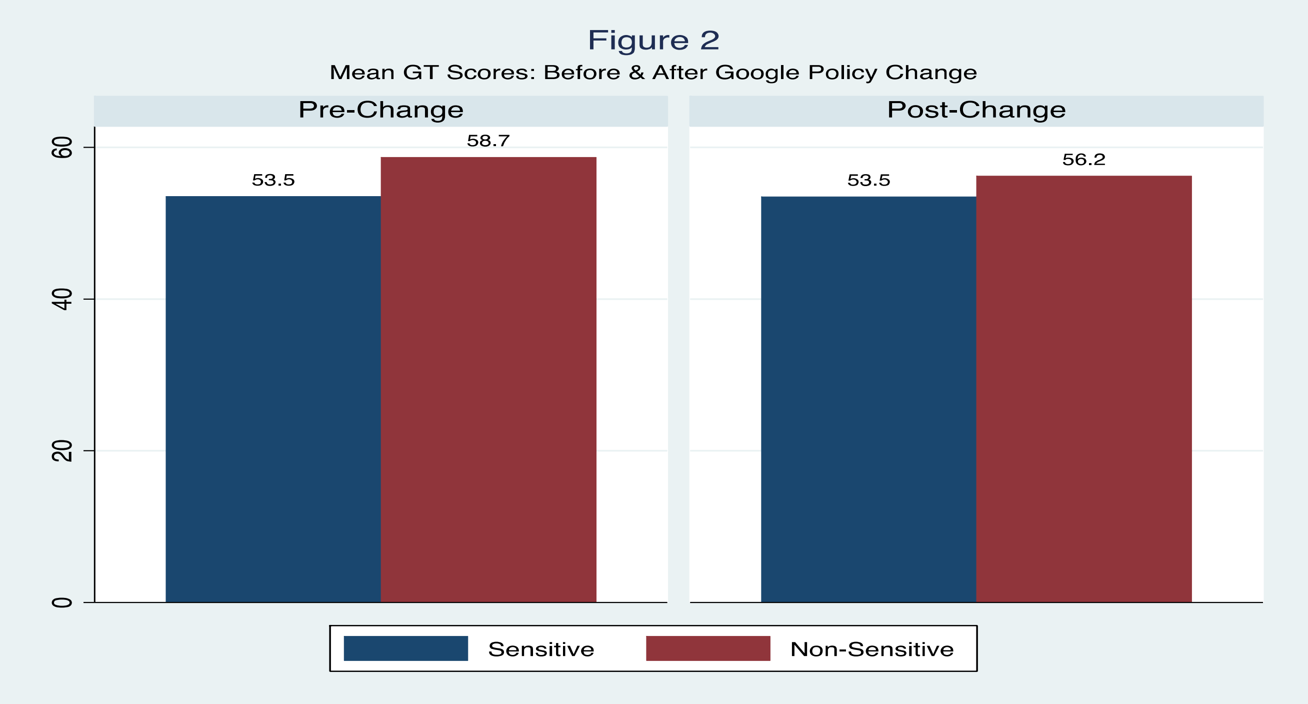


Figure 2
Mean GT Scores: Before & After Google Policy Change
Pre-Change
Post-Change
60
40
20
0
53.5
58.7
53.5
56.2
Sensitive
Non-Sensitive

Results

- Regression analysis
  - Setup:
    - Unit of observation: GT score for search *i*, during week *t*, in state *j*
    - 13k – 108k observations, depending on window
    - Week, term, and state effects in all specifications
  - Main Findings:
    - 5-10% reduction in sensitive search with +/-1 and +/- 3 month windows
    - No measurable impact with +/- 6 month window
    - Impact does not vary by state-level privacy demand

Results

- Results robust to different mixes of sensitive terms
- Falsification check for 2011 fails, but check for 2013 OK

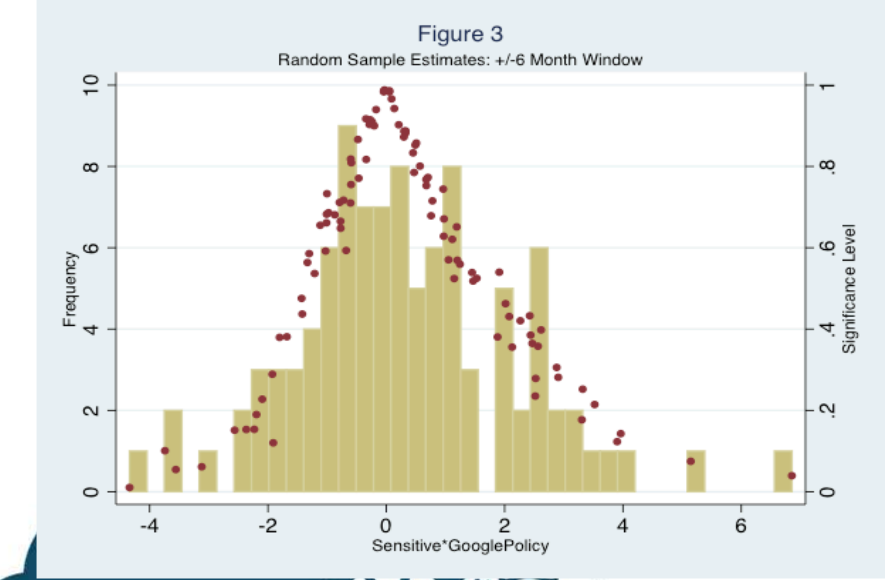


Figure 3

Random Sample Estimates: +/-6 Month Window

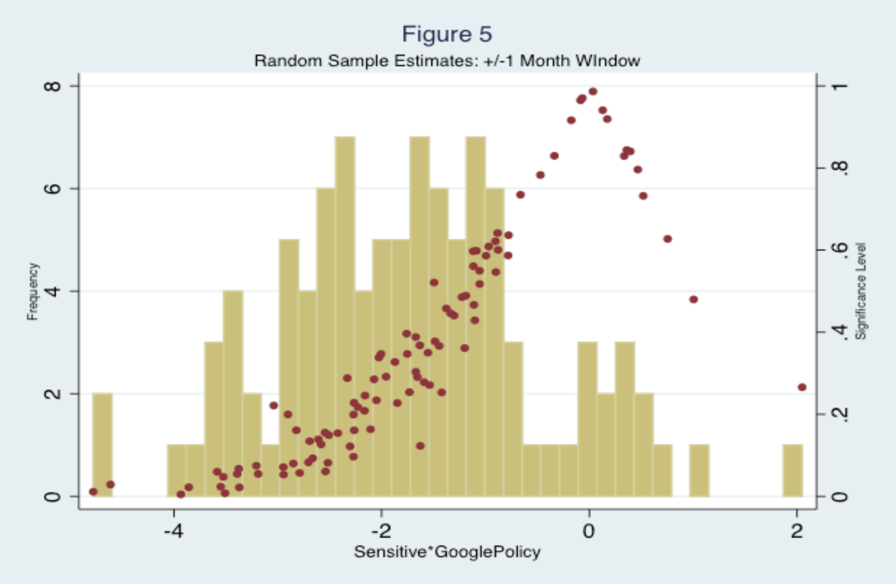


Figure 5

Random Sample Estimates: +/-1 Month WIndow

Conclusion

- Change may have induced a small drop in sensitive search, but this faded quickly
- Limitations/Future Work:
  - Trends not volume
  - Don’t have universe of sensitive search
  - Other unmeasured margins may be more important:
    - E.g., content in Gmail; viewing in YouTube

PRINCETON UNIVERSITY

CENTER FOR INFORMATION TECHNOLOGY POLICY
PRINCETON UNIVERSITY

Insights from a 1-million-site Measurement of Online Tracking

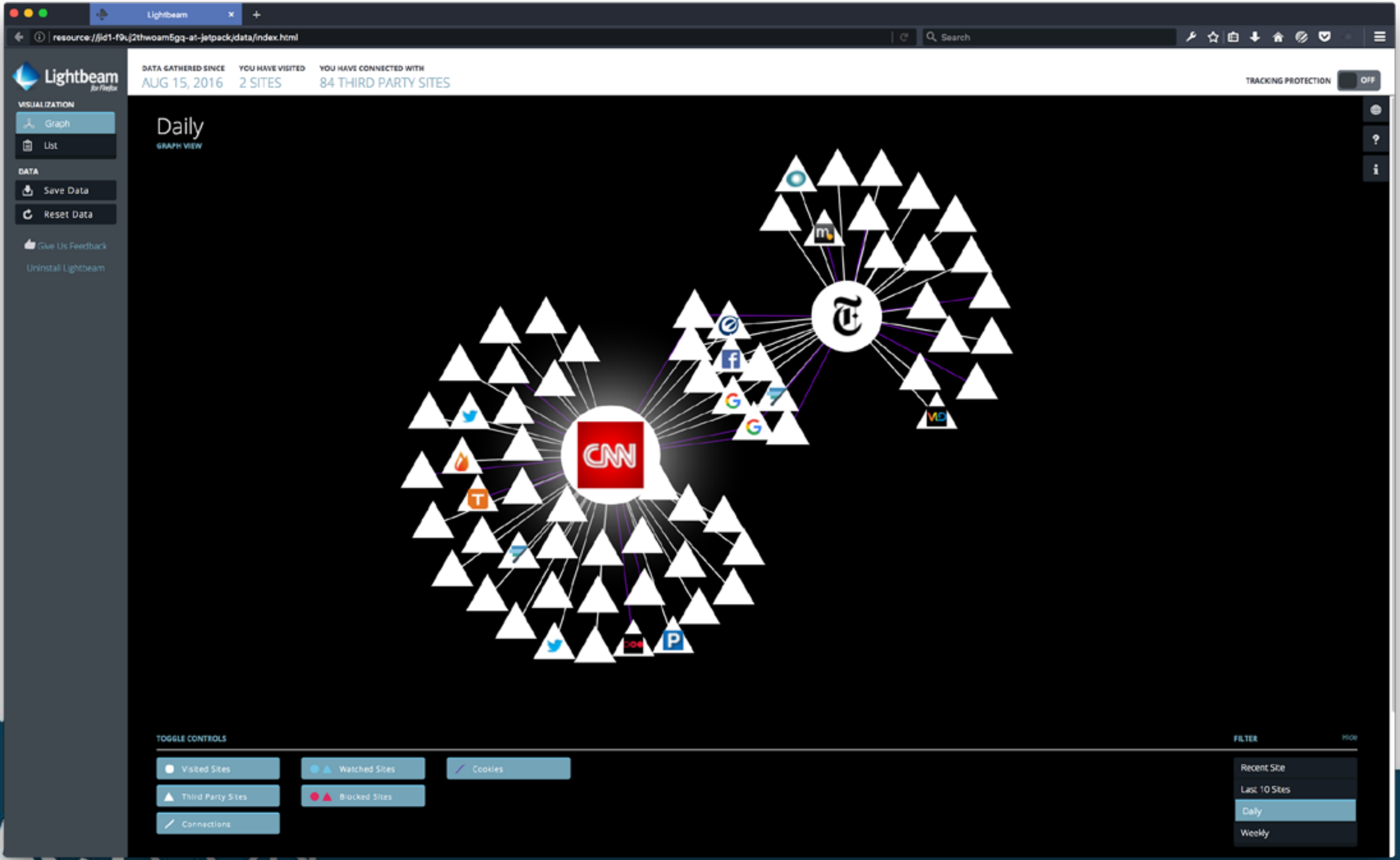
Steven Englehardt
@s_englehardt

Dillon Reisman
@dillonthehuman

Arvind Narayanan
@random_walker

This research was funded by NSF award CNS 1526353, a grant from the Data Transparency Lab, and by a cloud credits for research grant from Amazon Web Services.

Visiting 2 websites results in 84 third parties contacted

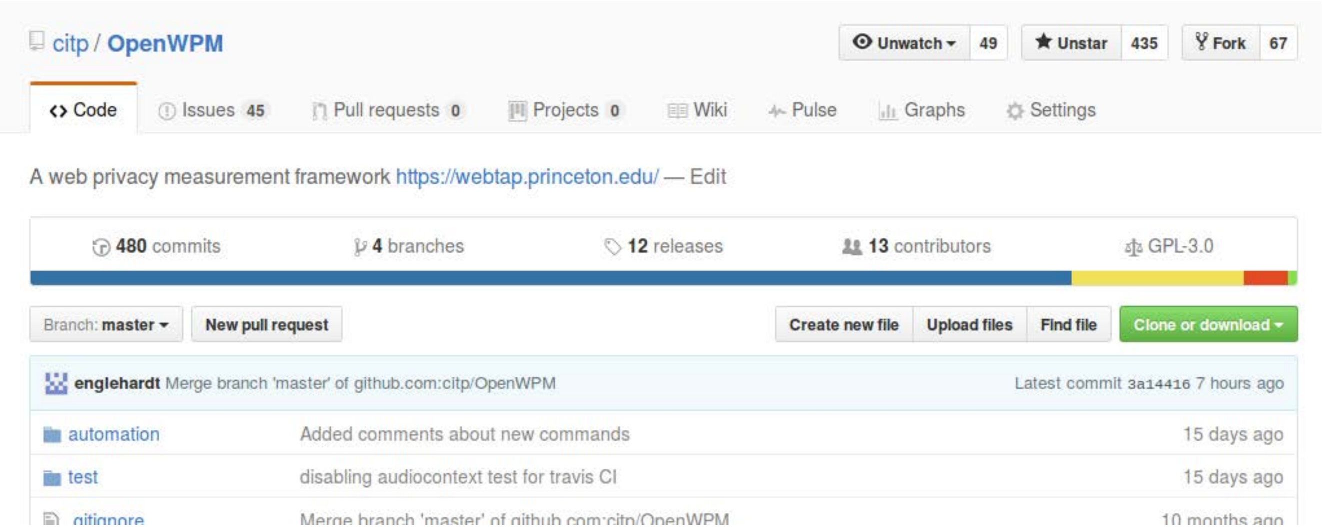


DATA GATHERED SINCE
AUG 15, 2016

YOU HAVE VISITED
2 SITES

YOU HAVE CONNECTED WITH
84 THIRD PARTY SITES

Open Web Privacy Measurement (OpenWPM)



A web privacy measurement framework https://webtap.princeton.edu/ — Edit

480 commits | 4 branches | 12 releases | 13 contributors | GPL-3.0

englehardt Merge branch 'master' of github.com:citp/OpenWPM Latest commit 3a14416 7 hours ago

automation	Added comments about new commands	15 days ago
test	disabling audiocontext test for travis CI	15 days ago
.gitignore	Merge branch 'master' of github.com:citp/OpenWPM	10 months ago

https://github.com/citp/OpenWPM

The Princeton Web Census

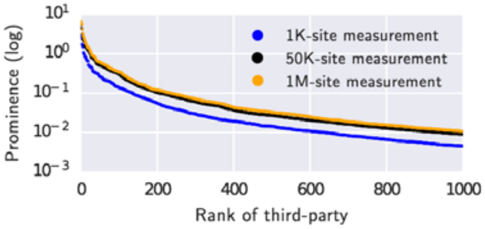
Monthly

1 Million Site Crawl

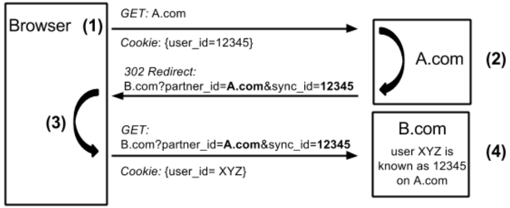
Collecting:

- Javascript Calls
- All javascript files
- HTTP Requests and Responses
- Storage (cookies, Flash, etc)

Results of the Princeton Web Census



New metric to rank third parties



Browser (1)

GET: A.com

Cookie: {user_id=12345}

A.com (2)

302 Redirect: B.com?partner_id=**A.com**&sync_id=**12345**

(3)

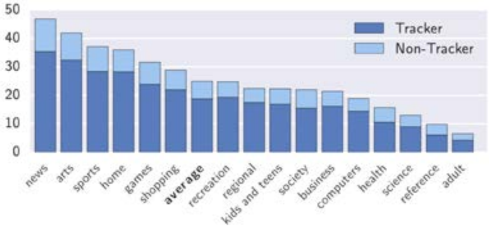
GET: B.com?partner_id=**A.com**&sync_id=**12345**

Cookie: {user_id= XYZ}

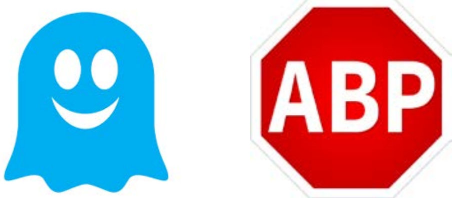
B.com

user XYZ is known as 12345 on A.com (4)

Nearly all top 3rd parties cookie sync



News site have the most trackers

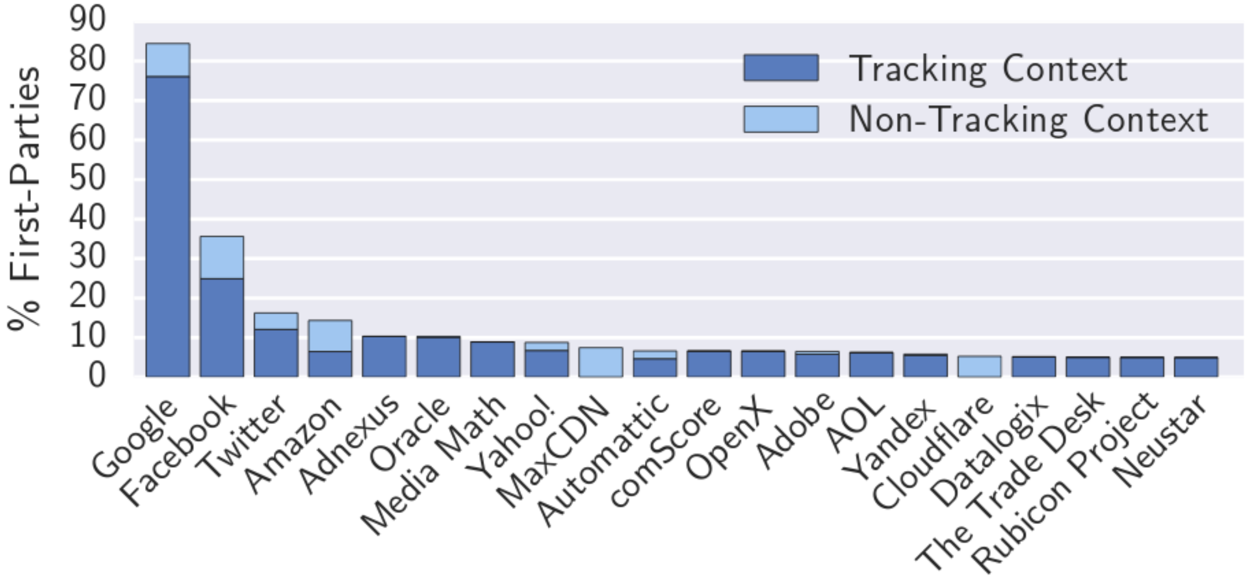


Tracking protection misses less popular 3rd parties and techniques

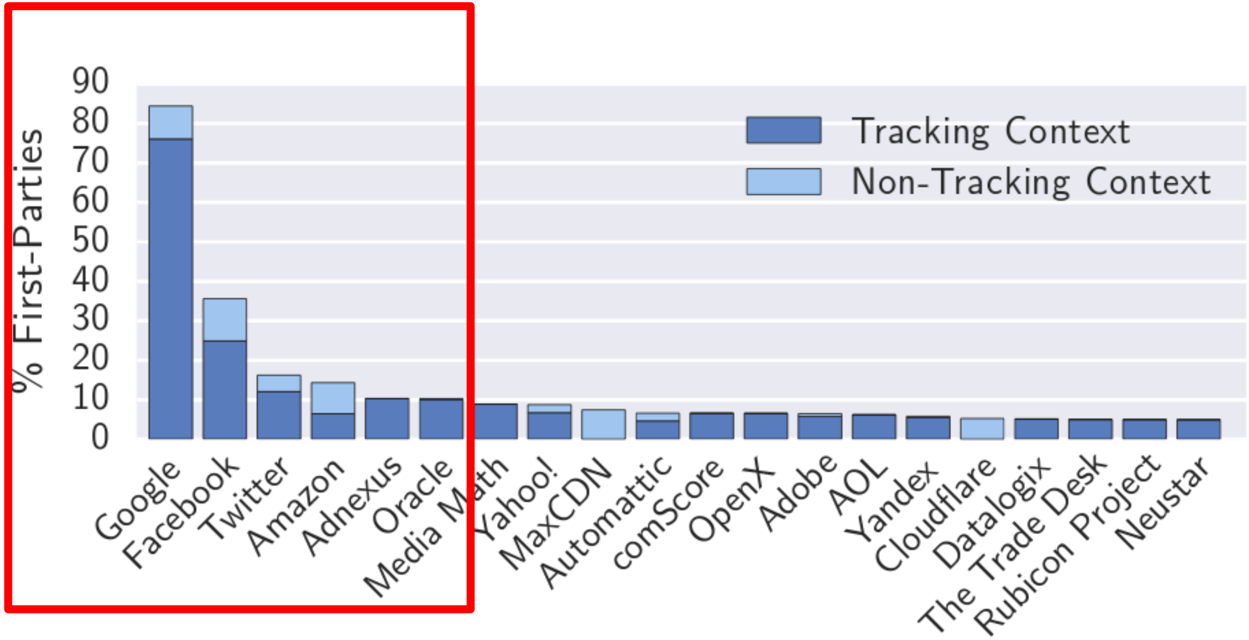
https://webtransparency.cs.princeton.edu/webcensus/

Insights from the Princeton Web Census

Consolidation of top trackers



Only 6 organizations are present on >10% of sites



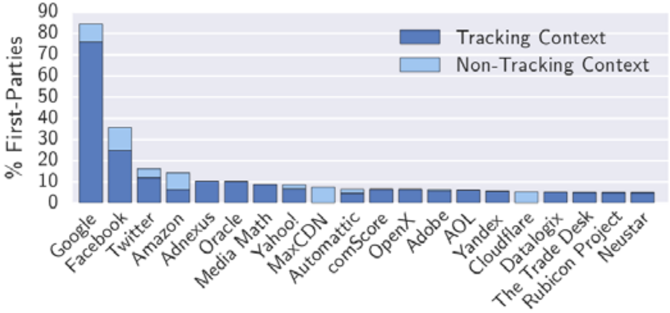
PRIVACYCON

Takeaways of consolidation

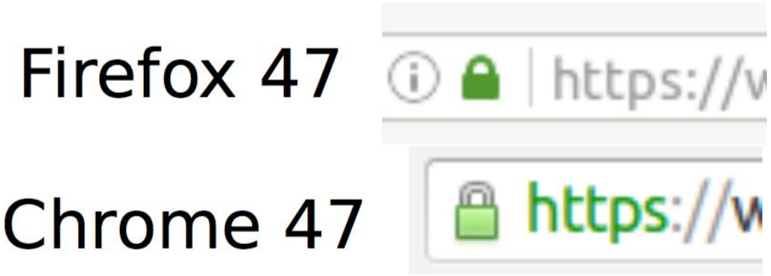
(1) Enforcement efforts can target large players, proactively set tracking norms.

(2) Large trackers can quickly deploy technique to a massive number of sites.

(3) Acquisitions can quickly shift tracking capability



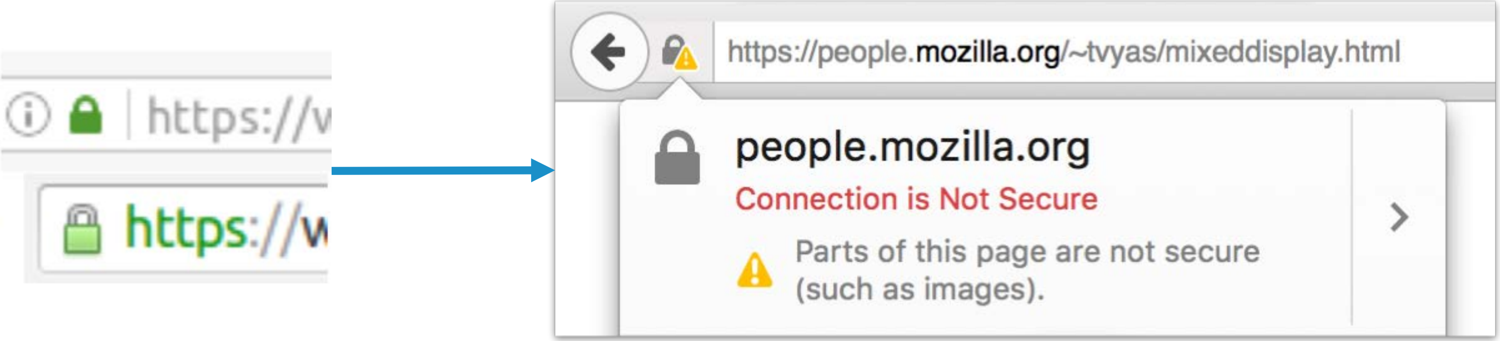
Trackers Impede HTTPS Adoption



Trackers Impede HTTPS Adoption

Firefox 47

Chrome 47



https://people.mozilla.org/~tvyas/mixeddisplay.html

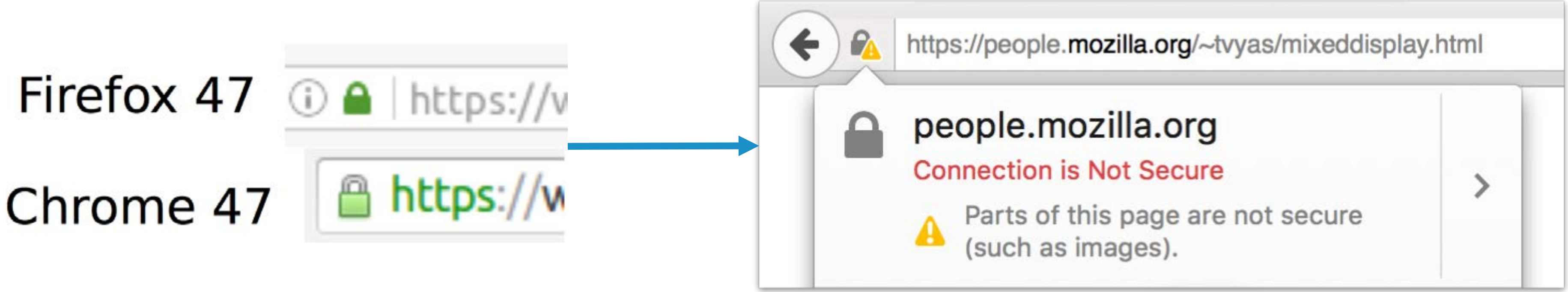
people.mozilla.org

Connection is Not Secure

Parts of this page are not secure (such as images).

Mixed content downgrades security indicator!

Trackers Impede HTTPS Adoption



https://people.mozilla.org/~tvyas/mixeddisplay.html

people.mozilla.org

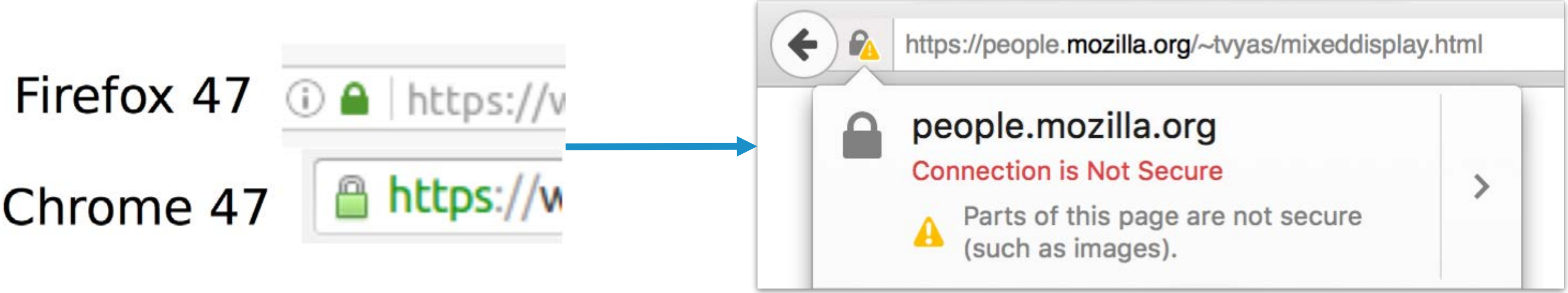
Connection is Not Secure

Parts of this page are not secure (such as images).

Of sites with mixed content:

half is caused solely by third parties (10% by trackers)

Trackers Impede HTTPS Adoption



https://people.mozilla.org/~tvyas/mixeddisplay.html

people.mozilla.org

Connection is Not Secure

Parts of this page are not secure (such as images).

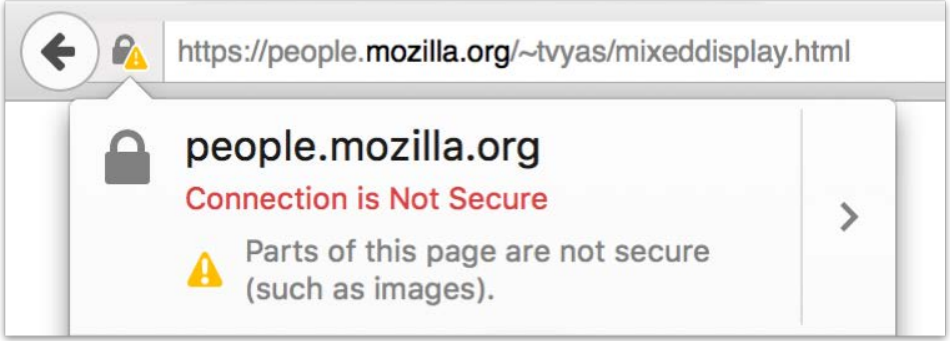
Of sites with mixed content:

half is caused solely by third parties (10% by trackers)

Half of all third-parties are HTTP-only

Takeaway: Tracking may have second-order privacy impacts

- (1) Slow the adoption of encryption
- (2) Identifier leakage in requests to
- (3) Can aid network surveillance efforts


https://people.mozilla.org/~tvyas/mixeddisplay.html

people.mozilla.org

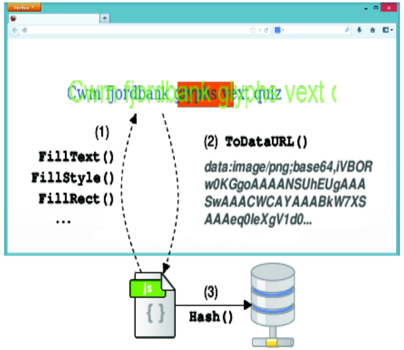
Connection is Not Secure

Parts of this page are not secure (such as images).

PRIVACYCON

New Browser Features Used for Fingerprinting

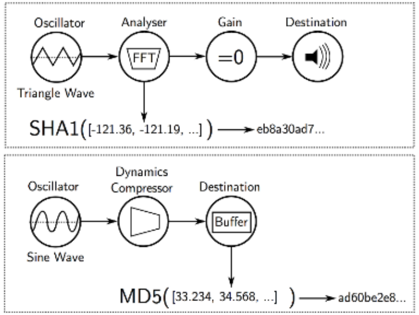
Canvas



```
data:image/png;base64,iVBOR
w0KGgoAAAANSUhEUgAAA
SwAAACWCAYAAABkW7XS
AAAeq0leXgV1d0...
```

WebRTC

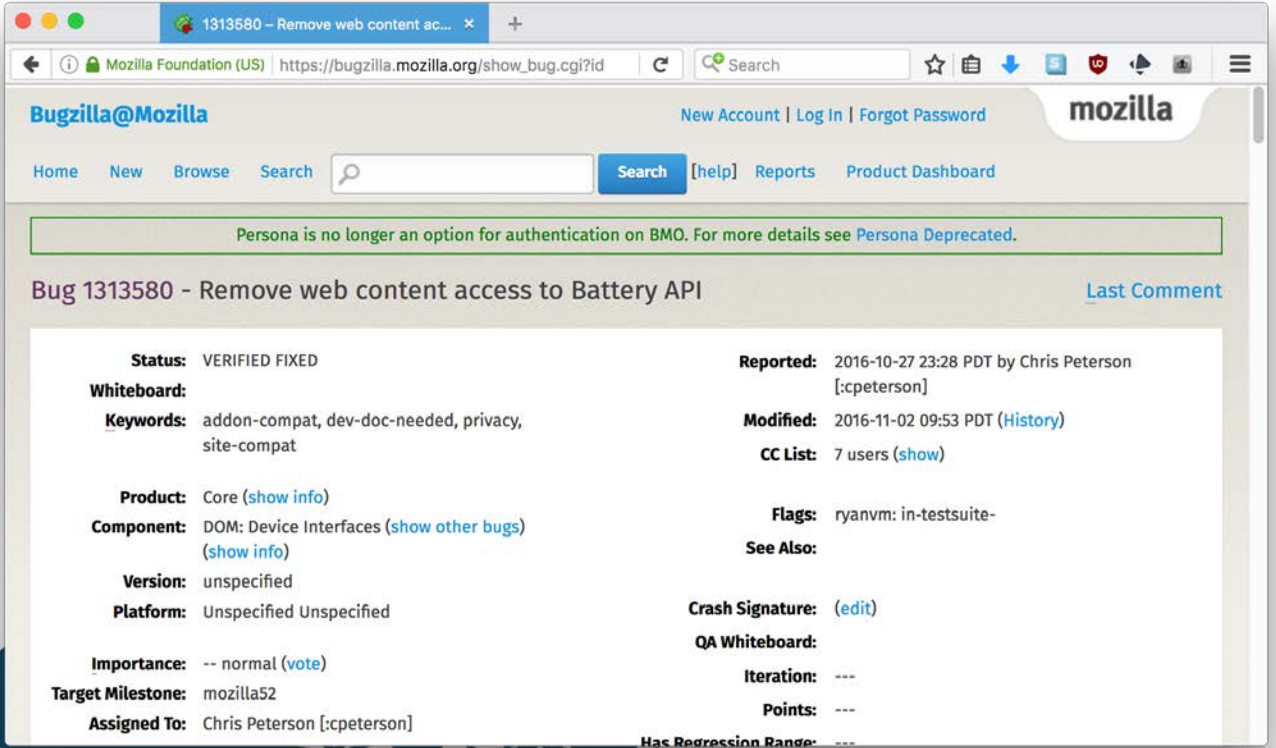
Audio



Battery

https://webtransparency.cs.princeton.edu/webcensus/

Browsers remove BatteryStatus API citing privacy


1313580 – Remove web content ac...
Mozilla Foundation (US)
https://bugzilla.mozilla.org/show_bug.cgi?id
Search
Bugzilla@Mozilla
New Account | Log In | Forgot Password
mozilla
Home
New
Browse
Search
Search
[help]
Reports
Product Dashboard

Persona is no longer an option for authentication on BMO. For more details see Persona Deprecated.

Bug 1313580 - Remove web content access to Battery API

Last Comment

Status: VERIFIED FIXED

Whiteboard:

Keywords: addon-compat, dev-doc-needed, privacy, site-compat

Product: Core (show info)

Component: DOM: Device Interfaces (show other bugs) (show info)

Version: unspecified

Platform: Unspecified Unspecified

Importance: -- normal (vote)

Target Milestone: mozilla52

Assigned To: Chris Peterson [:cpeterson]

Reported: 2016-10-27 23:28 PDT by Chris Peterson [:cpeterson]

Modified: 2016-11-02 09:53 PDT (History)

CC List: 7 users (show)

Flags: ryanvm: in-testsuite-

See Also:

Crash Signature: (edit)

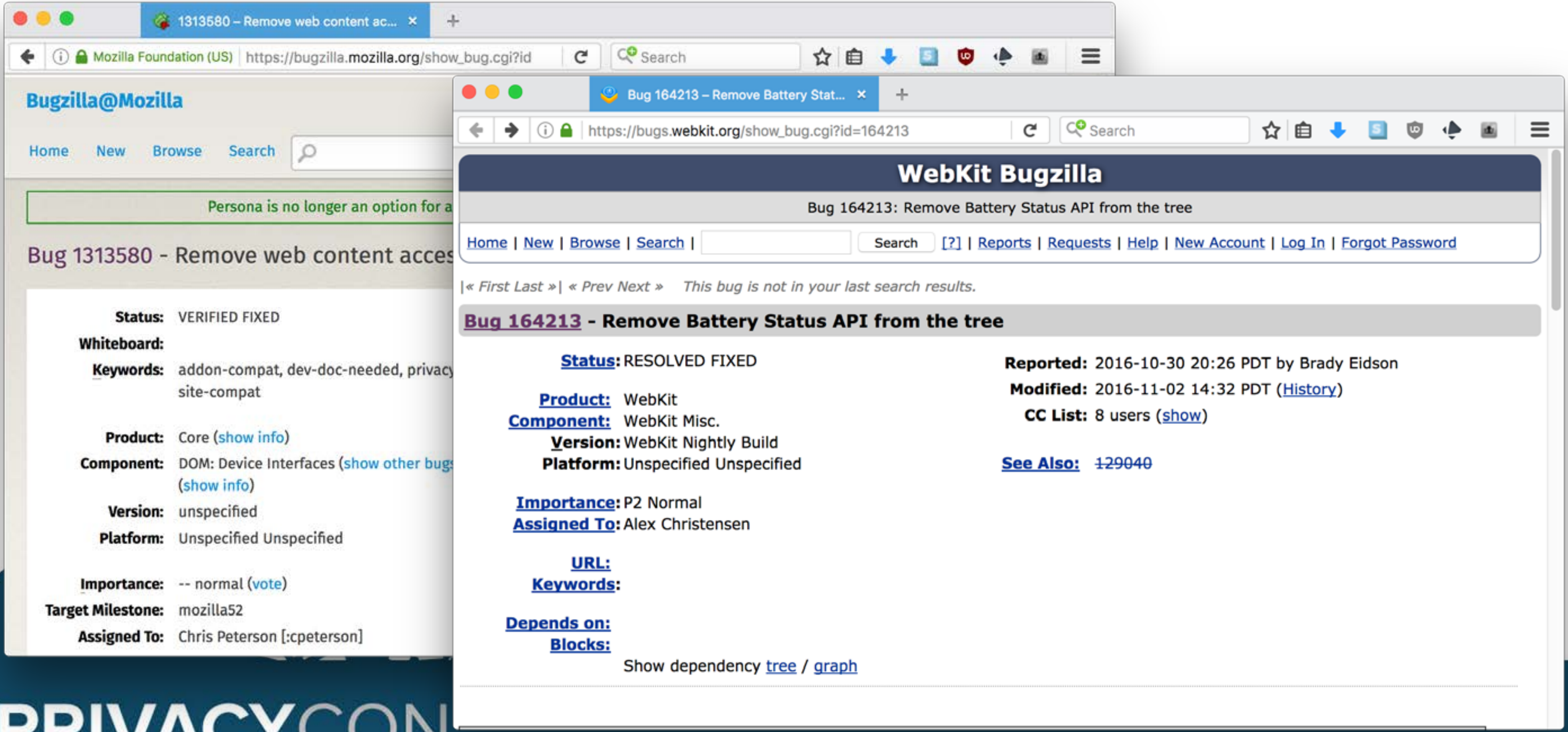
QA Whiteboard:

Iteration: ---

Points: ---

Has Regression Range: ---

Browsers remove BatteryStatus API citing privacy



1313580 – Remove web content ac...
Mozilla Foundation (US) https://bugzilla.mozilla.org/show_bug.cgi?id
Search

Bugzilla@Mozilla

Home New Browse Search

Persona is no longer an option for a

Bug 1313580 - Remove web content acces

Status:	VERIFIED FIXED
Whiteboard:	
Keywords:	addon-compat, dev-doc-needed, privacy site-compat
Product:	Core (show info)
Component:	DOM: Device Interfaces (show other bugs (show info)
Version:	unspecified
Platform:	Unspecified Unspecified
Importance:	-- normal (vote)
Target Milestone:	mozilla52
Assigned To:	Chris Peterson [:cpeterson]

Bug 164213 – Remove Battery Stat...
https://bugs.webkit.org/show_bug.cgi?id=164213
Search

WebKit Bugzilla

Bug 164213: Remove Battery Status API from the tree

Home | New | Browse | Search | Search [?] | Reports | Requests | Help | New Account | Log In | Forgot Password

|« First Last »| « Prev Next » *This bug is not in your last search results.*

Bug 164213 - Remove Battery Status API from the tree

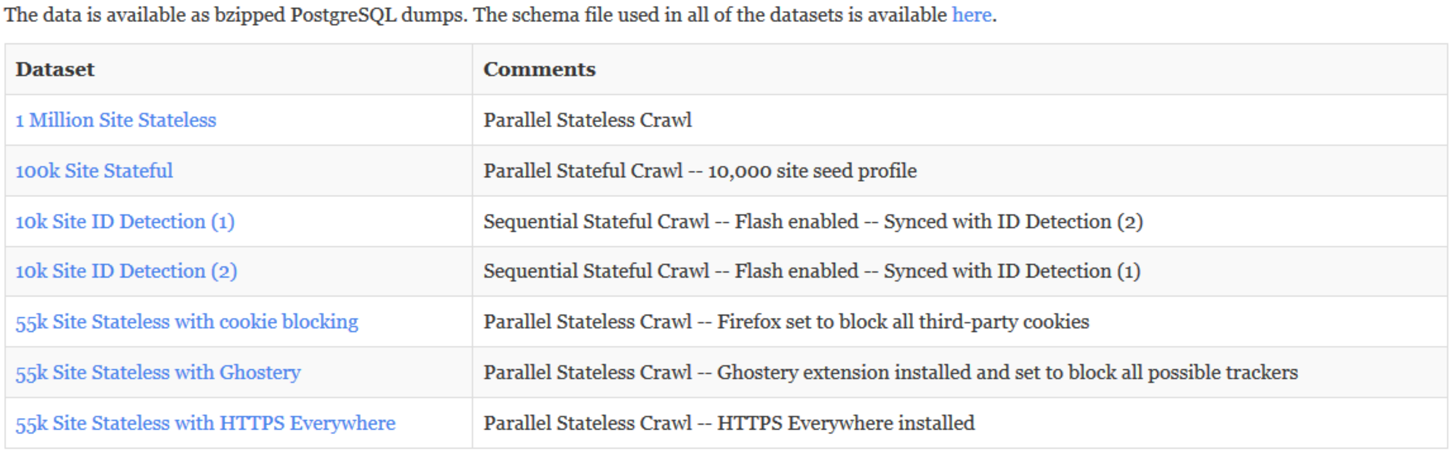
Status:	RESOLVED FIXED
Product:	WebKit
Component:	WebKit Misc.
Version:	WebKit Nightly Build
Platform:	Unspecified Unspecified
Importance:	P2 Normal
Assigned To:	Alex Christensen
URL:	
Keywords:	
Depends on:	
Blocks:	
	Show dependency tree / graph

Reported:	2016-10-30 20:26 PDT by Brady Eidson
Modified:	2016-11-02 14:32 PDT (History)
CC List:	8 users (show)
See Also:	~~129040~~

Takeaway: Expect any new API to be analyzed for its fingerprintability

1. Early detection of abuse can stem adoption
2. Browsers **view fingerprinting as abuse**
    a. Mitigate fingerprinting during standardization
    b. Remove APIs due to fingerprinting use

Our data is available!

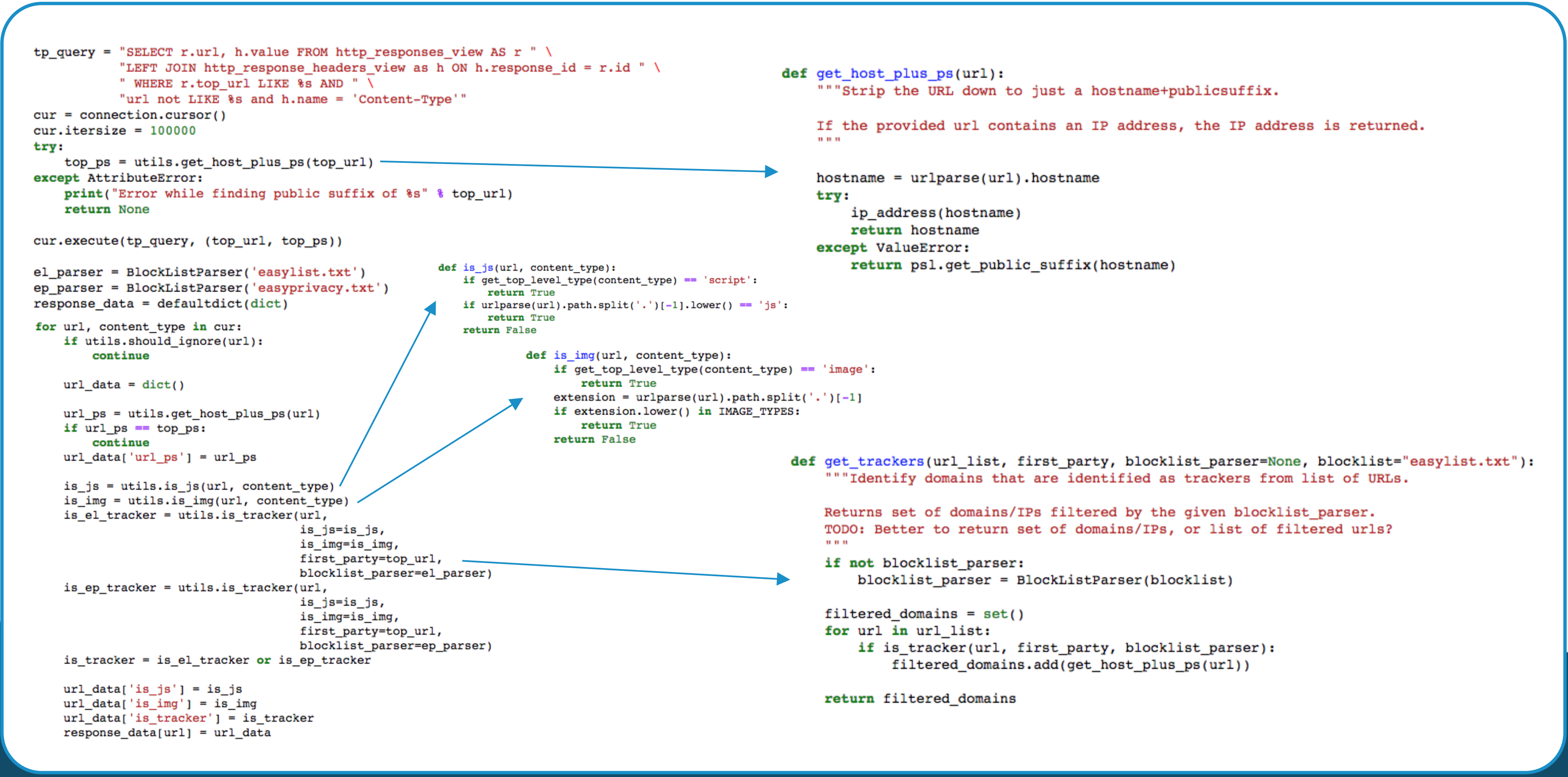


The data is available as bzipped PostgreSQL dumps. The schema file used in all of the datasets is available here.

Dataset	Comments
1 Million Site Stateless	Parallel Stateless Crawl
100k Site Stateful	Parallel Stateful Crawl -- 10,000 site seed profile
10k Site ID Detection (1)	Sequential Stateful Crawl -- Flash enabled -- Synced with ID Detection (2)
10k Site ID Detection (2)	Sequential Stateful Crawl -- Flash enabled -- Synced with ID Detection (1)
55k Site Stateless with cookie blocking	Parallel Stateless Crawl -- Firefox set to block all third-party cookies
55k Site Stateless with Ghostery	Parallel Stateless Crawl -- Ghostery extension installed and set to block all possible trackers
55k Site Stateless with HTTPS Everywhere	Parallel Stateless Crawl -- HTTPS Everywhere installed

https://webtransparency.cs.princeton.edu/webcensus/index.html#data

Getting third-party responses from our data



```python
tp_query = "SELECT r.url, h.value FROM http_responses_view AS r " \
           "LEFT JOIN http_response_headers_view as h ON h.response_id = r.id " \
           " WHERE r.top_url LIKE %s AND " \
           "url not LIKE %s and h.name = 'Content-Type'"
cur = connection.cursor()
cur.itersize = 100000
try:
    top_ps = utils.get_host_plus_ps(top_url)
except AttributeError:
    print("Error while finding public suffix of %s" % top_url)
    return None

cur.execute(tp_query, (top_url, top_ps))

el_parser = BlockListParser('easylist.txt')
ep_parser = BlockListParser('easyprivacy.txt')
response_data = defaultdict(dict)

for url, content_type in cur:
    if utils.should_ignore(url):
        continue

    url_data = dict()

    url_ps = utils.get_host_plus_ps(url)
    if url_ps == top_ps:
        continue
    url_data['url_ps'] = url_ps

    is_js = utils.is_js(url, content_type)
    is_img = utils.is_img(url, content_type)
    is_el_tracker = utils.is_tracker(url,
                                     is_js=is_js,
                                     is_img=is_img,
                                     first_party=top_url,
                                     blocklist_parser=el_parser)
    is_ep_tracker = utils.is_tracker(url,
                                     is_js=is_js,
                                     is_img=is_img,
                                     first_party=top_url,
                                     blocklist_parser=ep_parser)
    is_tracker = is_el_tracker or is_ep_tracker

    url_data['is_js'] = is_js
    url_data['is_img'] = is_img
    url_data['is_tracker'] = is_tracker
    response_data[url] = url_data
```

```python
def get_host_plus_ps(url):
    """Strip the URL down to just a hostname+publicsuffix.

    If the provided url contains an IP address, the IP address is returned.
    """

    hostname = urlparse(url).hostname
    try:
        ip_address(hostname)
        return hostname
    except ValueError:
        return psl.get_public_suffix(hostname)
```

```python
def is_js(url, content_type):
    if get_top_level_type(content_type) == 'script':
        return True
    if urlparse(url).path.split('.')[-1].lower() == 'js':
        return True
    return False
```

```python
def is_img(url, content_type):
    if get_top_level_type(content_type) == 'image':
        return True
    extension = urlparse(url).path.split('.')[-1]
    if extension.lower() in IMAGE_TYPES:
        return True
    return False
```

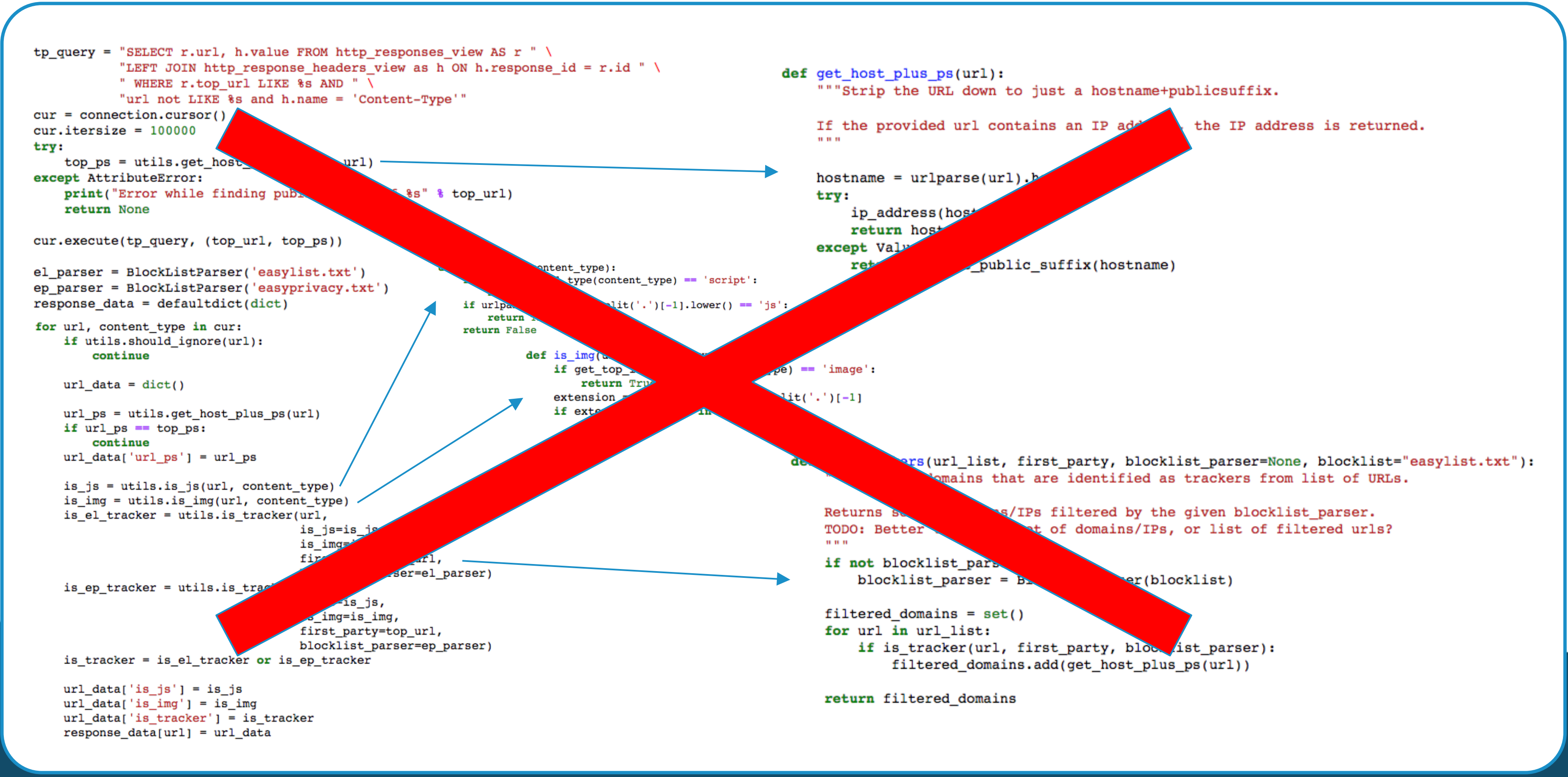
```python
def get_trackers(url_list, first_party, blocklist_parser=None, blocklist="easylist.txt"):
    """Identify domains that are identified as trackers from list of URLs.

    Returns set of domains/IPs filtered by the given blocklist_parser.
    TODO: Better to return set of domains/IPs, or list of filtered urls?
    """
    if not blocklist_parser:
        blocklist_parser = BlockListParser(blocklist)

    filtered_domains = set()
    for url in url_list:
        if is_tracker(url, first_party, blocklist_parser):
            filtered_domains.add(get_host_plus_ps(url))

    return filtered_domains
```

Getting third-party responses from our data



```
tp_query = "SELECT r.url, h.value FROM http_responses_view AS r " \
           "LEFT JOIN http_response_headers_view as h ON h.response_id = r.id " \
           " WHERE r.top_url LIKE %s AND " \
           "url not LIKE %s and h.name = 'Content-Type'"
cur = connection.cursor()
cur.itersize = 100000
try:
    top_ps = utils.get_host_          url)
except AttributeError:
    print("Error while finding pub          %s" % top_url)
    return None

cur.execute(tp_query, (top_url, top_ps))

el_parser = BlockListParser('easylist.txt')
ep_parser = BlockListParser('easyprivacy.txt')
response_data = defaultdict(dict)

for url, content_type in cur:
    if utils.should_ignore(url):
        continue

    url_data = dict()

    url_ps = utils.get_host_plus_ps(url)
    if url_ps == top_ps:
        continue
    url_data['url_ps'] = url_ps

    is_js = utils.is_js(url, content_type)
    is_img = utils.is_img(url, content_type)
    is_el_tracker = utils.is_tracker(url,
                                     is_js=is_js
                                     is_img=
                                     fir          url,
                                              ser=el_parser)
    is_ep_tracker = utils.is_trac
                                           =is_js,
                                     s_img=is_img,
                                     first_party=top_url,
                                     blocklist_parser=ep_parser)
    is_tracker = is_el_tracker or is_ep_tracker

    url_data['is_js'] = is_js
    url_data['is_img'] = is_img
    url_data['is_tracker'] = is_tracker
    response_data[url] = url_data
```

```
def get_host_plus_ps(url):
    """Strip the URL down to just a hostname+publicsuffix.

    If the provided url contains an IP add          the IP address is returned.
    """

    hostname = urlparse(url).h
    try:
        ip_address(hos
        return hos
    except Valu
        ret                 _public_suffix(hostname)
```

```
                 ontent_type):
                 _type(content_type) == 'script':

    if urlpa                lit('.')[-1].lower() == 'js':
        return
    return False

        def is_img(
            if get_top_               pe) == 'image':
                return Tru
            extension =                lit('.')[-1]
            if exte
```

```
de                  rs(url_list, first_party, blocklist_parser=None, blocklist="easylist.txt"):
                    omains that are identified as trackers from list of URLs.

    Returns s                s/IPs filtered by the given blocklist_parser.
    TODO: Better                t of domains/IPs, or list of filtered urls?
    """
    if not blocklist_pars
        blocklist_parser = B                 er(blocklist)

    filtered_domains = set()
    for url in url_list:
        if is_tracker(url, first_party, bloc      ist_parser):
            filtered_domains.add(get_host_plus_ps(url))

    return filtered_domains
```

Getting third-party responses with Census.py

```
census.get_third_party_responses_by_domain(
   database_connection,
   “http://nytimes.com”
)
```

Getting third-party responses with Census.py

- get_third_party_responses_by_domain
- get_third_party_responses_by_domain
- get_cookie_syncs_on_domain
- is_tracker
- get_trackers

Getting third-party responses with Census.py

- `get_third_party_responses_by_domain`
- `get_third_party_responses_by_domain`
- `get_cookie_syncs_on_domain`
- `is_tracker`
- `get_trackers`

Contact us for access to “alpha” analysis server and library!

Thanks for listening!

Full Paper:
senglehardt.com/papers/ccs16_online_tracking.pdf

Data and Analysis:
webtransparency.cs.princeton.edu/webcensus/

Collaborate:
webtap.princeton.edu/research/

Contact Me

Email: ste@cs.princeton.edu

Twitter: @s_englehardt

Web: senglehardt.com

Image Assets from the Noun Project:
Browser Network and Browser Battery by Aybige

Detection and Circumvention of Anti Ad-Blockers: A New Arms Race on the Web

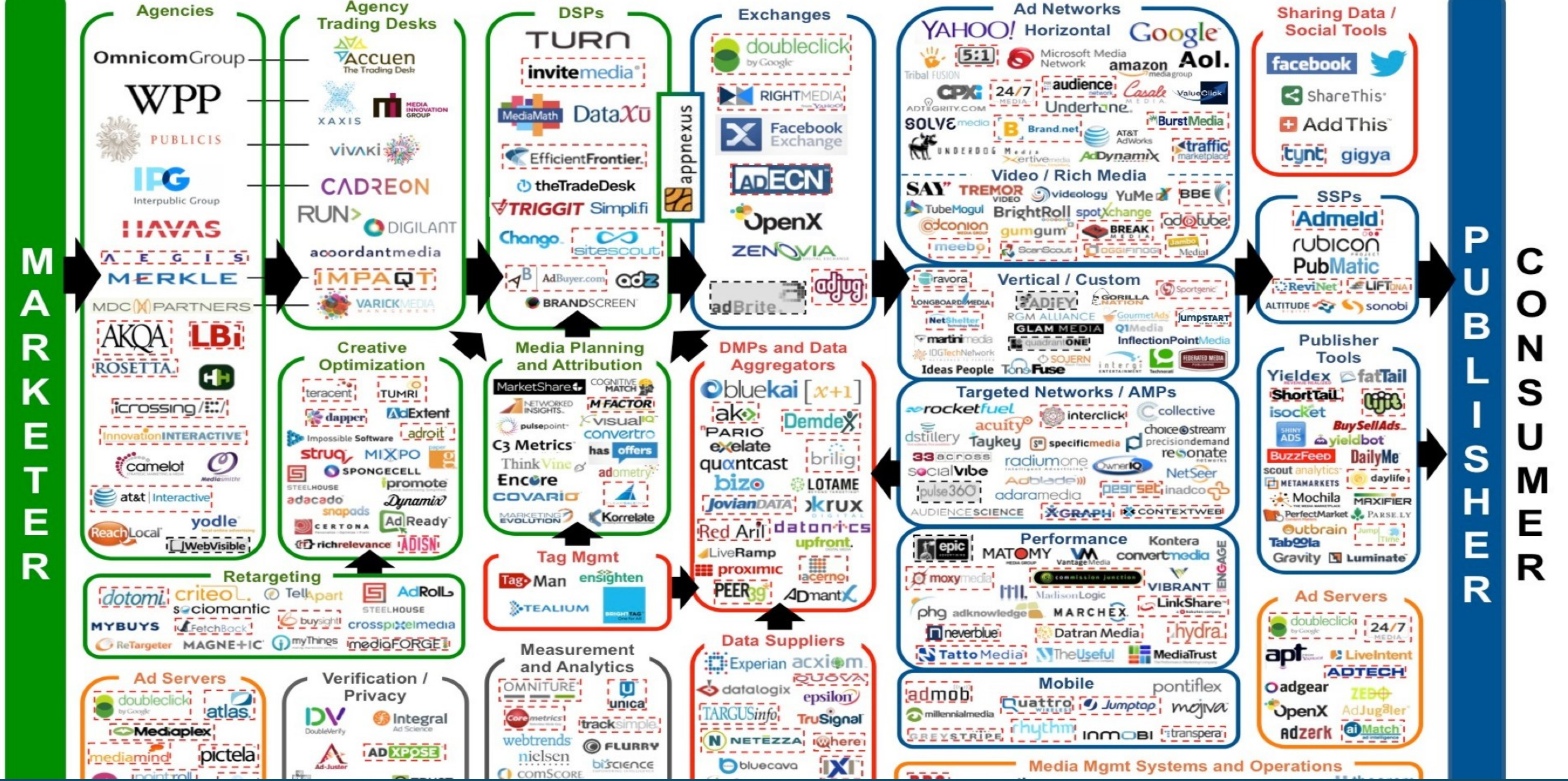
Zubair Shafiq (University of Iowa)
Zhiyun Qian (UC-Riverside)

This research was funded by the Data Transparency Lab (DTL)

PRIVACYCON

NEW!
Tell Something About Your Company HERE!
PUT YOUR WONDERFUL CAMPAIGN TEXT HERE
to increase your sale
50% OFF
Plus a chance to win bonus prizes on the grand draw!
BIG EVENT WILL COME THIS NOVEMBER AND IT'S AWESOME
50% OFF on selected items
the best offer of this year will come
AND GET A CHANCE TO WIN
$25,000
ON GRAND DRAW!
learn more
Corporate Banner Ad Campaign
Learn More
YOUR COMPANY NAME
YOUR IMPORTANT CAMPAIGN TEXT HERE
TO CATCH THEIR ATTENTION
LEARN MORE
YOUR COMPANY
some catchy words to attract the potential clients
FREE SIGN UP
HOT!
ONE BIG SUMMER SALE!
LEARN MORE
NEW!
TELL SOMETHING ABOUT YOUR BUSINESS
yourcompany
LEARN MORE
NEED A VACATION?
experience it now!
swim, surf, dive, and more!
BOOK NOW!
NEW!
Best Mobile Phone Deal
get it here first!
limited units only.
GET PRICES
YOUR LOGO
FREE SHIPPING
AFFORDABLE [YOUR ITEM]
shop now and get a chance to win gadgets!
YOUR COMPANY NAME
www.yourcompany.com
SOME MARKETING WORDS
we will increase your sales 10 times
Learn More
YOURLOGO
START YOUR CAREER in HEALTHCARE
LEARN MORE
Tell Something About Your Business
High Conversion Banner Ad for Your Campaign
learn more
are you the right one that we're looking for?
WE WANT YOU!
yourLOGO
APPLY NOW
www.yourcompany.com
INCREASE YOUR WEBSITE TRAFFIC!
LEARN MORE

PRIVACYCON


MARKETER
Agencies
OmnicomGroup
WPP
PUBLICIS
IPG Interpublic Group
HAVAS
AEGIS
MERKLE
MDC PARTNERS
AKQA
LBi
ROSETTA
icrossing
InnovationINTERACTIVE
camelot
Mediasmith
at&t Interactive
ReachLocal
yodle
WebVisible
Agency Trading Desks
Accuen The Trading Desk
XAXIS
MEDIA INNOVATION GROUP
VIVAKI
CADREON
RUN
DIGILANT
accordantmedia
IMPAQT
VARICKMEDIA MANAGEMENT
DSPs
TURN
invitemedia
MediaMath
DataXu
appnexus
EfficientFrontier
theTradeDesk
TRIGGIT
Simplifi
Chango
sitescout
AdBuyer.com
adz
BRANDSCREEN
Exchanges
doubleclick by Google
RIGHTMEDIA
Facebook Exchange
ADECN
OpenX
ZEDOVIA
adBrite
adjug
Ad Networks
Horizontal
YAHOO!
Google
5:1
Tribal FUSION
Microsoft Media Network
amazon media group
Aol.
CPX
24/7 MEDIA
audience network
Casale MEDIA
ValueClick
ADVERTISING.COM
Undertone
SOLVE media
Brand.net
AT&T AdWorks
BurstMedia
UNDERDOG Media
Xertive media
AdDynamix
traffic marketplace
Video / Rich Media
SAY
TREMOR VIDEO
videology
YuMe
BBE
TubeMogul
BrightRoll
spotXchange
adotube
adconion
gumgum
BREAK MEDIA
meebo
ScanScout
Jambo Media
Vertical / Custom
Gravora
LONGBOARD MEDIA
NetShelter
martini media
IDGTechNetwork
Ideas People
RGM ALLIANCE
GLAM MEDIA
quadrantONE
SOJERN
GORILLA NATION
GourmetAds
Q1Media
InflectionPointMedia
Sportgenic
jumpstart
FEDERATED MEDIA
Targeted Networks / AMPs
rocketfuel
interclick
collective
dstillery
acuity
Taykey
specificmedia
choicestream
precisiondemand
resonate networks
33across
radiumone
OwnerIQ
socialvibe
Adblade
NetSeer
pulse360
adaramedia
peer39
inadco
AUDIENCESCIENCE
XGRAPH
CONTEXTWEB
Performance
epic
MATOMY
Vantage Media
Kontera
convertmedia
ENGAGE
commission junction
VIBRANT
moxy media
MadisonLogic
LinkShare
phg
adknowledge
MARCHEX
neverblue
Datran Media
hydra
Tatto Media
TheUseful
MediaTrust
Mobile
admob
millennialmedia
GREYSTRIPE
Quattro wireless
Jumptap
rhythm
pontiflex
mojiva
inmobi
transpera
Sharing Data / Social Tools
facebook
ShareThis
AddThis
tynt
gigya
SSPs
Admeld
rubicon project
PubMatic
ReviNet
LIFT DNA
ALTITUDE digital
sonobi
Publisher Tools
Yieldex
fatTail
ShortTail
isocket
BuySellAds
yieldbot
BuzzFeed
DailyMe
scout analytics
daylife
METAMARKETS
Mochila
MAXIFIER
PerfectMarket
PARSE.LY
Outbrain
Jump Time
Taboola
Gravity
Luminate
Ad Servers
doubleclick by Google
24/7 MEDIA
apt
LiveIntent
ADTECH
adgear
ZEDO
OpenX
AdJuggler
adzerk
aiMatch
Creative Optimization
teracent
TUMRI
dapper
AdExtent
adroit
Impossible Software
struq
MIXPO
paper g
SPONGECELL
STEELHOUSE
ipromote
adacado
Dynamix
snapads
AdReady
CERTONA
richrelevance
ADISN
Media Planning and Attribution
MarketShare
COGNITIVE MATCH
NETWORKED INSIGHTS
M FACTOR
pulsepoint
visualIQ
convertro
C3 Metrics
has offers
ThinkVine
adometry
Encore
COVARIO
MARKETING EVOLUTION
Korrelate
Tag Mgmt
TagMan
ensighten
TEALIUM
BRIGHTTAG
DMPs and Data Aggregators
bluekai
[x+1]
aka
Demdex
PARIO
exelate
quantcast
brilig
bizo
LOTAME
JovianDATA
krux
Red Aril
datonics
upfront
LiveRamp
proximic
acerno
PEER39
ADmantX
Data Suppliers
Experian
acxiom
datalogix
QUOVA
epsilon
TARGUSinfo
TruSignal
NETEZZA
where
bluecava
IXI
Retargeting
dotomi
criteo
TellApart
AdRoll
sociomantic
buysight
STEELHOUSE
MYBUYS
FetchBack
crosspixelmedia
ReTargeter
MAGNETIC
myThings
mediaFORGE
Ad Servers
doubleclick by Google
atlas
Mediaplex
mediamind
pictela
pointroll
Verification / Privacy
DV DoubleVerify
Integral Ad Science
Ad-Juster
ADXPOSE
Measurement and Analytics
OMNITURE
unica
Coremetrics
tracksimple
webtrends
FLURRY
nielsen
bizscience
comScore
Media Mgmt Systems and Operations
PUBLISHER
CONSUMER

PRIVACYCON

(Self) Regulation

FTC

AdChoices

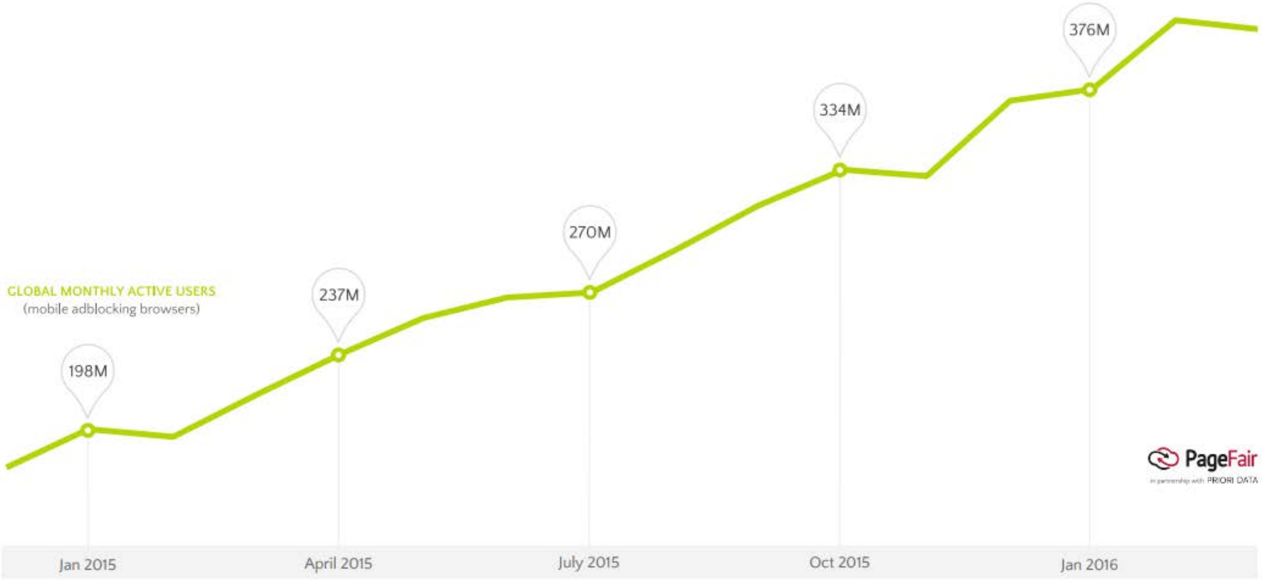
Do Not Track

FCC

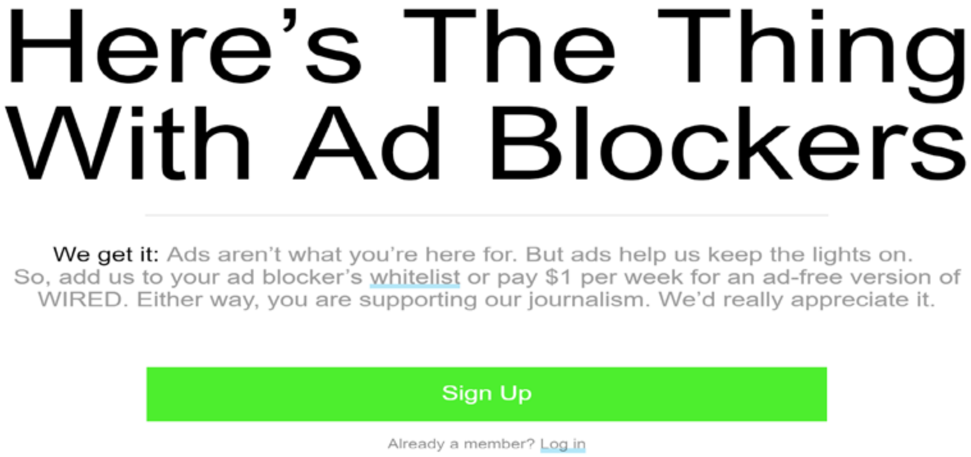
Privacy Preserving Tools

- EFF’s Privacy Badger
  - Target publishers that do not respect DNT
- Ghostery
  - Proprietary
- Ad-block (open-source, public filter lists)
  - Adblock Plus
  - uBlock Origin

Popularity of Ad-blockers



But…



Here's The Thing With Ad Blockers

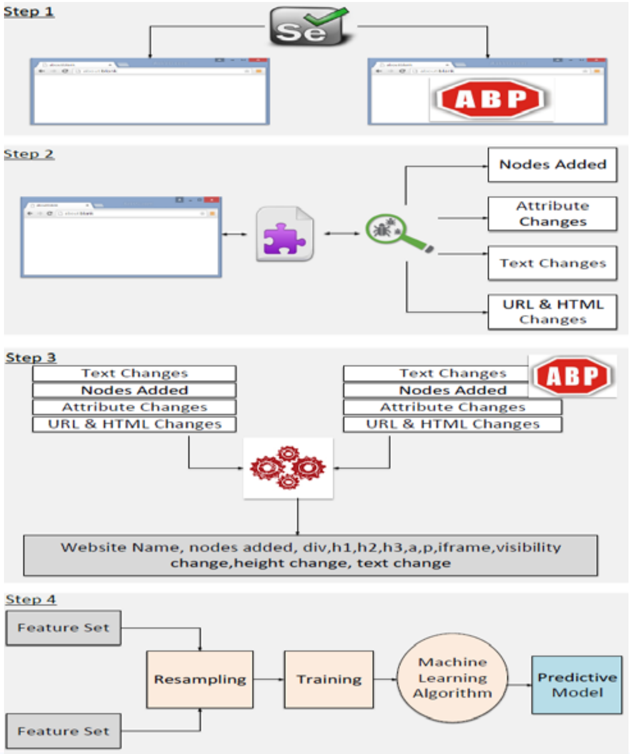
We get it: Ads aren't what you're here for. But ads help us keep the lights on. So, add us to your ad blocker's whitelist or pay $1 per week for an ad-free version of WIRED. Either way, you are supporting our journalism. We'd really appreciate it.

Project Goals

- Measure Anti Ad-blocking in the Wild
  - Number of anti ad-blocking publishers
  - Third-party anti ad-blocking services
  - Ad-block detection techniques
- Develop a Stealthy Ad-blocker
  - Automatically block anti ad-block scripts

Measuring Anti Ad-blockers

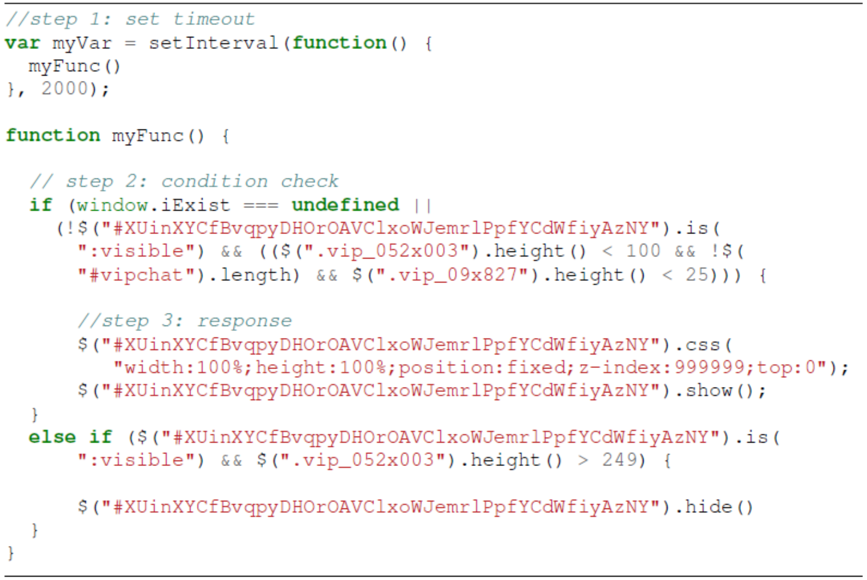
- Crawl Alexa top 100K websites
- A/B testing
  - with and without ad-block
- Feature extraction
  - nodes, attributes, text
- Machine learning models
  - Random forest, SVM, Bayesian
- Results
  - 95% precision, 93% recall
  - 1100 websites use anti ad-blockers



How Anti Ad-blockers Work?

Websites employ anti ad-block scripts

- Identify leaked extension information
- Verify ads (active or passive)



```
//step 1: set timeout
var myVar = setInterval(function() {
  myFunc()
}, 2000);

function myFunc() {

  // step 2: condition check
  if (window.iExist === undefined ||
    (!$("#XUinXYCfBvqpyDHOrOAVClxoWJemrlPpfYCdWfiyAzNY").is(
     ":visible") && (($(".vip_052x003").height() < 100 && !$(
     "#vipchat").length) && $(".vip_09x827").height() < 25))) {

     //step 3: response
     $("#XUinXYCfBvqpyDHOrOAVClxoWJemrlPpfYCdWfiyAzNY").css(
        "width:100%;height:100%;position:fixed;z-index:999999;top:0");
     $("#XUinXYCfBvqpyDHOrOAVClxoWJemrlPpfYCdWfiyAzNY").show();
  }
  else if ($("#XUinXYCfBvqpyDHOrOAVClxoWJemrlPpfYCdWfiyAzNY").is(
     ":visible") && $(".vip_052x003").height() > 249) {

     $("#XUinXYCfBvqpyDHOrOAVClxoWJemrlPpfYCdWfiyAzNY").hide()
  }
}
```

Towards a Stealthy Ad-blocker

- Remove anti ad-block scripts through filter lists
- Crowd-sourced, manually populated

Takeaway

- Users, Society, Economics

- “Ad blocking is like garlic.
  You hang it on your door to keep Dracula away from sucking your blood.
  Ad blocking is not the enemy.
  Just stop being Dracula.”

  - Doc Searls

Discussion of Session 4

Presenters:

- **James C. Cooper**, Antonin Scalia Law School, George Mason University
- **Steven Englehardt**, Princeton University
- **Zubair Shafiq**, University of Iowa

Moderator:

- **Kristin Cohen,** Federal Trade Commission

BREAK