

The Web Never Forgets: Persistent Tracking Mechanisms in the Wild

Steven Englehardt

Joint work with:

Güneş Acar, Marc Juarez, Christian Eubank, Arvind Narayanan, Claudia Diaz

Expanded version of CCS 2014 Talk



Background

Advertising

- \$42.8 billion revenues in 2013 (IAB)
- Relies on tracking



Source: IEEE Spectrum, Photo: Dan Saelinger; Prop Stylist: Dominique Baynes

The New York Times - Breaking News, World News & Multimedia

The New York Times - Breaking... http://www.nytimes.com/

GUCCI

The New York Times

Tuesday, February 1, 2011 Last Update: 10:16 PM ET

Advertising

see what's inside

Advertising

Search

Follow Us

Subscribe to Home Delivery

Switch to Global Edition

JOBS
REAL ESTATE
AUTOS
ALL CLASSIFIEDS

WORLD
U.S.
POLITICS
NEW YORK
BUSINESS
DEALBOOK
TECHNOLOGY
SPORTS
SCIENCE
HEALTH
OPINION
ARTS
Books
Movies
Music
Television
Theater
STYLE
Dining & Wine
Fashion & Style
Home & Garden
Weddings/
Celebrations
TRAVEL

Mubarak Advertising

Opposition Demands That He Leave Sooner

Video

MARKETS

MARKETS	As of 10:14 PM ET
JAPAN	CHINA
Nikkei	HangSeng
15,470.50	23,826.39
+190.00	+343.44
+1.91%	+1.48%

1.30% NEW-YORK

Advertising

OPINION

EGYPT

Kristof: The White House should support the protesters in Egypt.

Blog: Mubarak's Speech

Follow on Twitter

Brooks: In the democratic

ave, a quest for dignity.

Koehn: From a culture of

ictimhood to one of

self-empowerment.

Senator John Kerry: A

uture without Mubarak.

Social

Log In With Facebook

Log in to see what your friends are sharing on nytimes.com.

Privacy Policy | What's This?

WHAT'S POPULAR NOW

- The Paradox of Corporate Taxes in America
- Why Not Regulate

MORE IN OPINION

Editorial: Debt Limit

Bloggingheads: Abortions

GET QUOTES

My Portfolios

Stock, ETFs, Funds

Go

Path to Change in Power Still Unclear

By ANTHONY SHADID 56 minutes ago

President Hosni Mubarak's vow to step down in the fall was not enough for the hundreds of thousands who poured into Tahrir Square.

Post a Comment | Read (332)

Path to Change in Power Still Unclear

By DAVID D. KIRKPATRICK and MARK LANDLER 1 minute ago

The democracy movement is unfolding so rapidly in Egypt that Washington came close to being left behind.

More Photographs

Interactive Map of Protests

King of Jordan Dismisses His Cabinet

Antiquities Chief Says Sites Are Largely Secure

New Service Lets Voices From Egypt Be Heard

The Lede: Updates From Day 8

Interactive Feature: A Timeline of Mubarak's Presidency

A chronology of President Hosni Mubarak's 30-year rule in Egypt.

CHARLES LEBOY

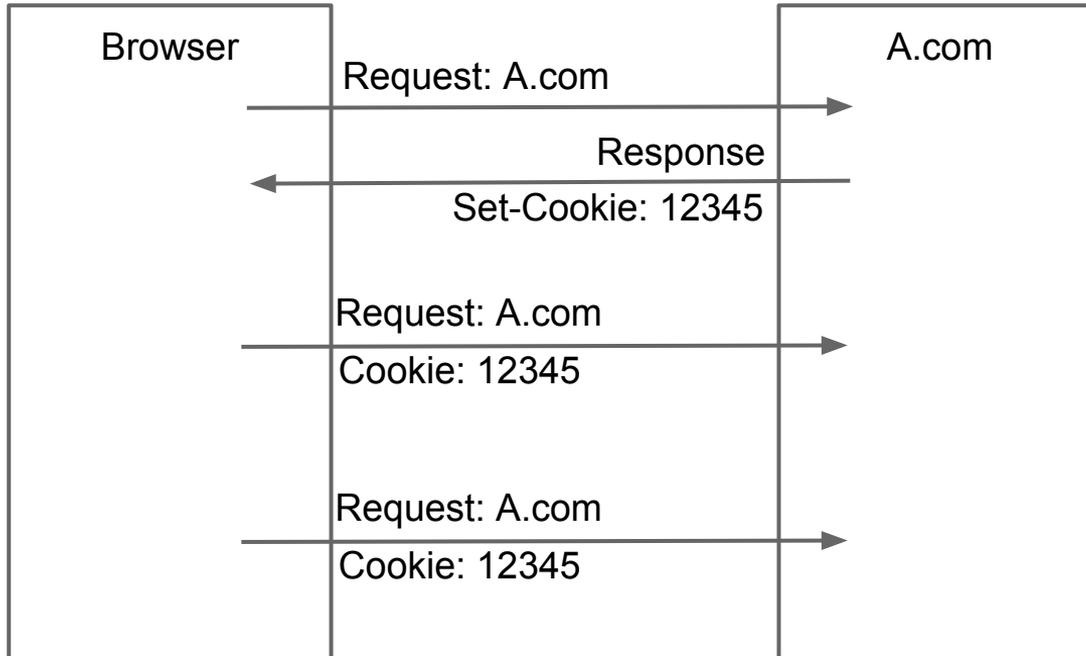
30-DAY FREE TRIAL

iMeet built a meeting room just for you.

Source: Mayer & Mitchell; Third-Party Web Tracking: Policy and Technology (and many COS432 lectures)

Tracking with Cookies

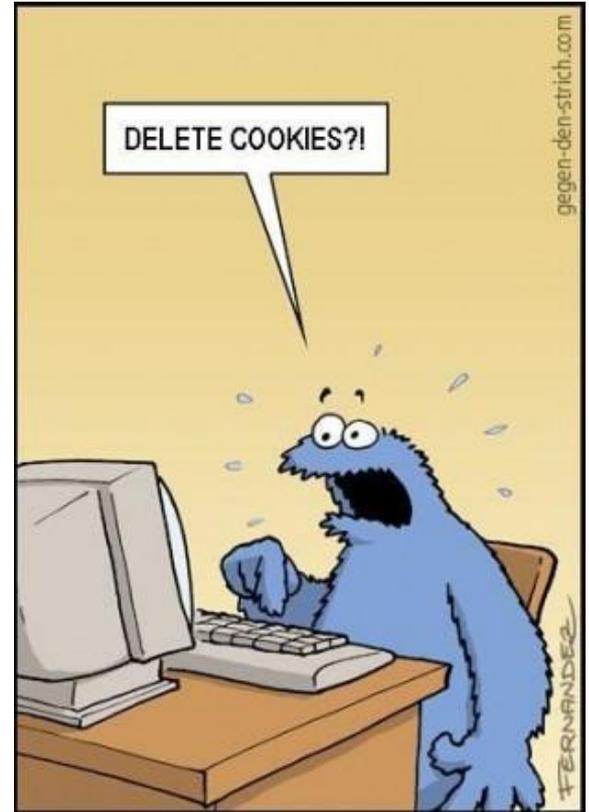
Primary tracking mechanism - built into requests



Tracking with Cookies

Not good enough?

- can be blocked
- don't work well on mobile
- 20-40% of the users delete cookies
(ComScore, April 2010)



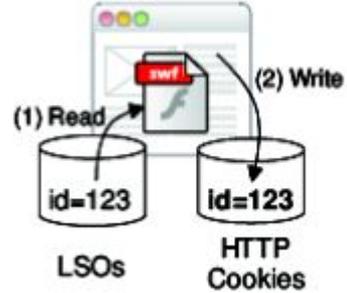
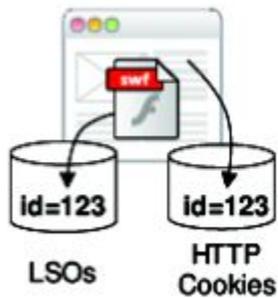
Advanced Tracking Mechanisms

Cookie Respawning

Fingerprinting

Cookie Syncing

Respawning (Evercookies)



Respawn cookies using obscure storage mechanisms

- e.g. Flash cookies, HTML5 localStorage, ETags

Respawning - Data collection & Analysis

Procedure:

1. Crawl top 10,000 sites with OpenWPM
2. Clear HTTP cookies, keep other location (e.g. LSOs)
3. Recrawl top 10,000 sites (independent visits)
4. Check for regenerated cookies
 - a. Check that IDs also in other location to rule out IDs regenerated by other means

Respawning

Findings - Top 10 Using Flash Cookies

Global rank	Site	CC	Respawning (Flash) domain	1st/3rd Party
16	sina.com.cn	CN	simg.sinajs.cn	3rd*
17	yandex.ru	RU	kiks.yandex.ru	1st
27	weibo.com	CN	simg.sinajs.cn	3rd*
41	hao123.com	CN	ar.hao123.com	1st
52	sohu.com	CN	tv.sohu.com	1st
64	ifeng.com	HK	y3.ifengimg.com	3rd*
69	youku.com	CN	irs01.net	3rd
178	56.com	CN	irs01.net	3rd
196	letv.com	CN	irs01.net	3rd
197	tudou.com	CN	irs01.net	3rd

Advanced Tracking Mechanisms

Cookie Respawning

Fingerprinting

Cookie Syncing



A research project of the **Electronic Frontier Foundation**

Panopti**cl**ick

How Unique **I** – and Trackable – Is Your Browser?

Is your browser configuration rare or unique? If so, web sites may be able to track you, *even if you limit or disable cookies.*

Panopti**cl**ick tests your browser to see how unique it is based on the **information** it will share with sites it visits. Click below and you will be given a uniqueness score, letting you see how easily identifiable you might be as you surf the web.

Only **anonymous data** will be collected by this site.

TEST
ME

Your browser fingerprint **appears to be unique among the 4,702,890 tested so far.**

Browser Characteristic	bits of identifying information	one in x browsers have this value	value
User Agent	9.3	631.77	Mozilla/5.0 (Windows NT 6.1; WOW64; rv:33.0) Gecko/20100101 Firefox/33.0
HTTP_ACCEPT Headers	3.96	15.52	text/html, */* gzip, deflate en-US,en;q=0.5
			Plugin 0: Adobe Acrobat; Adobe PDF Plug-In For Firefox and Netscape 10.1.12; nppdf32.dll; (Acrobat Portable Document Format;

■
■
■

System Fonts	22.17+	4702890	GulimChe, Gungsuh, GungsuhChe, Harlow Solid Italic, Harrington, High Tower Text, Impact, Intormal Roman, InsUPC, Iskoola Pota, JasmineUPC, Jokerman, Juice ITC, KaiTi, Kalinga, Kartika, Khmer UI, KodchiangUPC, Kokila, Kristen ITC, Kunstler Script, Lao UI, Latha, Leelawadee, Levenim MT, LilyUPC, Lucida Bright, Lucida Calligraphy, Lucida Console, Lucida Fax, Lucida Handwriting, Lucida Sans Unicode, Magneto, Malgun Gothic, Mangal, Marlett, Matura MT Script Capitals, Meiryo, Meiryo UI, Microsoft Himalaya, Microsoft JhengHei, Microsoft New Tai Lue, Microsoft PhagsPa, Microsoft Sans Serif, Microsoft Tai Le, Microsoft Uighur, Microsoft YaHei, Microsoft Yi Baiti,
--------------	--------	---------	--

■
■
■

Browser fingerprinting

FPDetective (Acar et al., CCS'13)

- detecting **font based fingerprinting**
- 16 previously unknown providers
- Flash based, 1.5% 10K sites
- JavaScript based, 0.04% of top 1M

The image shows two lines of text for comparison. The top line is 'Arial mmmmmmmmmml' with a red bracket above it labeled '44 PX'. The bottom line is 'Times New Roman mmmmmmmmmml' with a red bracket below it labeled '33 PX'. The 'm' characters in both lines are of different heights, with the Arial 'm' being taller than the Times New Roman 'm'.

Source: lalit.org

Canvas Fingerprinting

- Canvas and WebGL API
 - Allows sites to draw and render images from javascript
- Depends on:
 - OS
 - browser
 - fonts
 - font library
 - **graphics card & driver**
 - **font rendering (ClearType)**

Different than other fingerprinting vectors

Windows:

How quickly daft jumping zebras vex. (Also, pur
How quickly daft jumping zebras vex. (Also, pu

OS X:

How quickly daft jumping zebras vex. (Also, pu
How quickly daft jumping zebras vex. (Also, pu
How quickly daft jumping zebras vex. (Also, pu
How quickly daft jumping zebras vex. (Also, pu

Linux:

How quickly daft jumping zebras vex. (Also, pu
How quickly daft jumping zebras vex. (Also, pur
How quickly daft jumping zebras vex. (Also, p

Arial; 20px rendering

Chrome, Window(XP, Vista, 7)

How quickly daft jumping zebras vex. (

Linux:

How quickly daft jumping zebras vex. (f
How quickly daft jumping zebras vex. (f

OSX:

How quickly daft jumping zebras vex. (f

Windows (XP, Vista, 7):

ow e. e<y ea". " "e "g /e e's vex (f
(-low quickly daft jumping zebrae vex. (f
(-low quickly daft jumping zebrae vex. (f

Windows 8:

ow e. e<y ea". " "e "g /e e's vex (f

Diffs between renderings

Cwm fjordbank glyphs vext quiz vext c

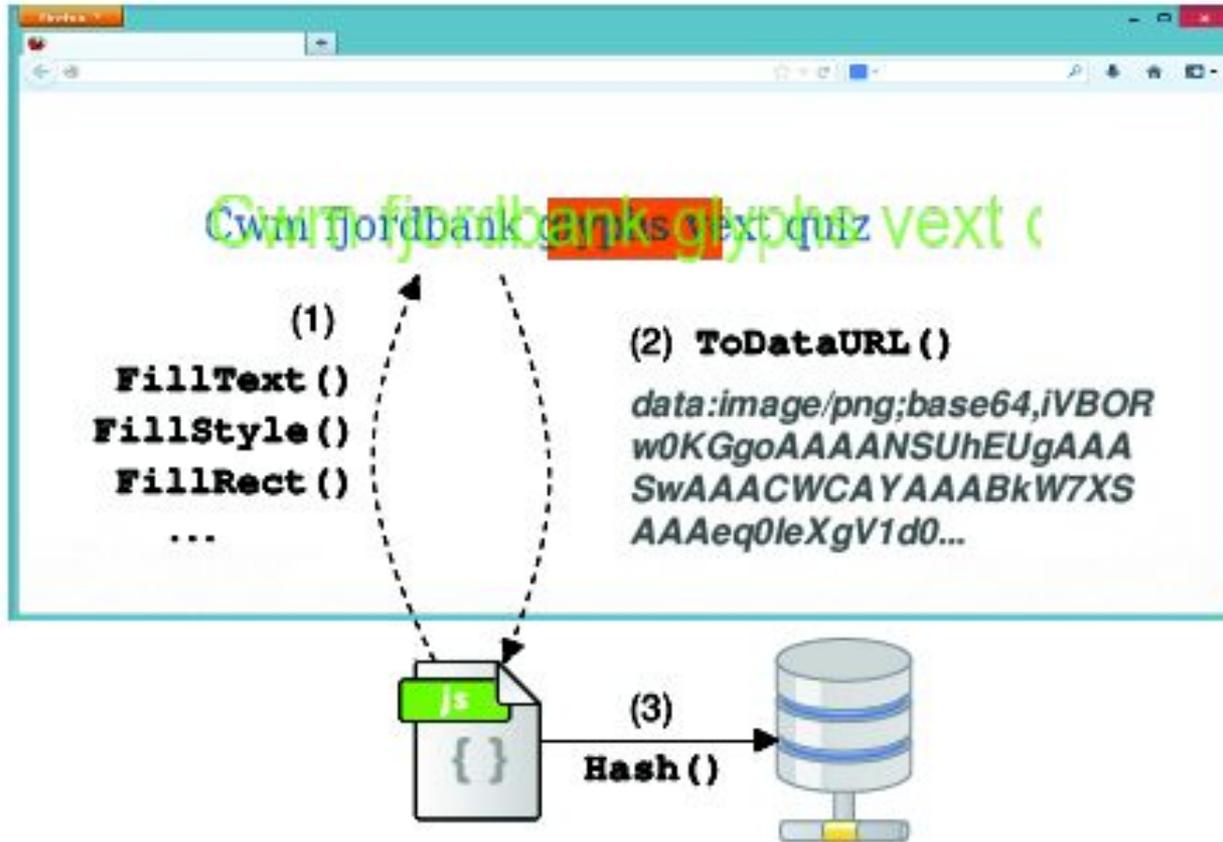
<http://valve.github.io>

<http://admicro.vn/>

<http://www.plentyoffish.com>

ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789+/-

Canvas Fingerprinting



Canvas Fingerprinting - Experimental Setup

- Crawler based on Selenium, mitmproxy and modified Firefox
 - Alexa top 100K (May 1-5, 2014)
 - Ran in parallel
 - Log canvas access functions (R/W)
 - Insert into a SQLite database
- Analysis & false positive removal

Canvas Fingerprinting Findings - Alexa Top 100K

Fingerprinting script	Number of including sites	Text drawn into the canvas
ct1.addthis.com/static/r07/core130.js	5282	Cwm fjordbank glyphs vext quiz, ☺
i.ligatus.com/script/fingerprint.min.js	115	http://valve.github.io
src.kitcode.net/fp2.js	68	http://valve.github.io
admicro1.vcmedia.vn/fingerprint/figp.js	31	http://admicro.vn/
amazonaws.com/af-bdaz/bquery.js	26	Centillion
*.shorte.st/js/packed/smeadvert-intermediate-ad.js	14	http://valve.github.io
stat.ringier.cz/js/fingerprint.min.js	4	http://valve.github.io
cya2.net/js/STAT/89946.js	3	ABCDEFGHIJKLMNOPQRSTUVWXYZ abcdefghijklmnopqrstuvwxyz0123456789+/ /
images.revtrax.com/RevTrax/js/fp/fp.min.jsp	3	http://valve.github.io
pof.com	2	http://www.plentyoffish.com
*.rackcdn.com/mongoose.fp.js	2	http://api.gonorthleads.com
9 others*	9	(Various)
TOTAL	5559 (5542 unique ¹)	-

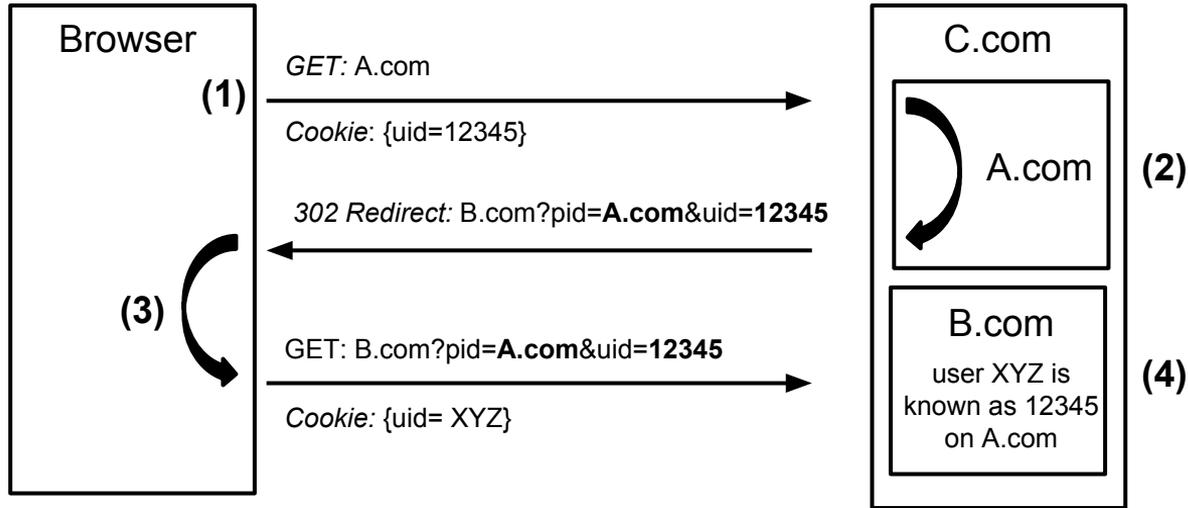
Advanced Tracking Mechanisms

Cookie Respawning

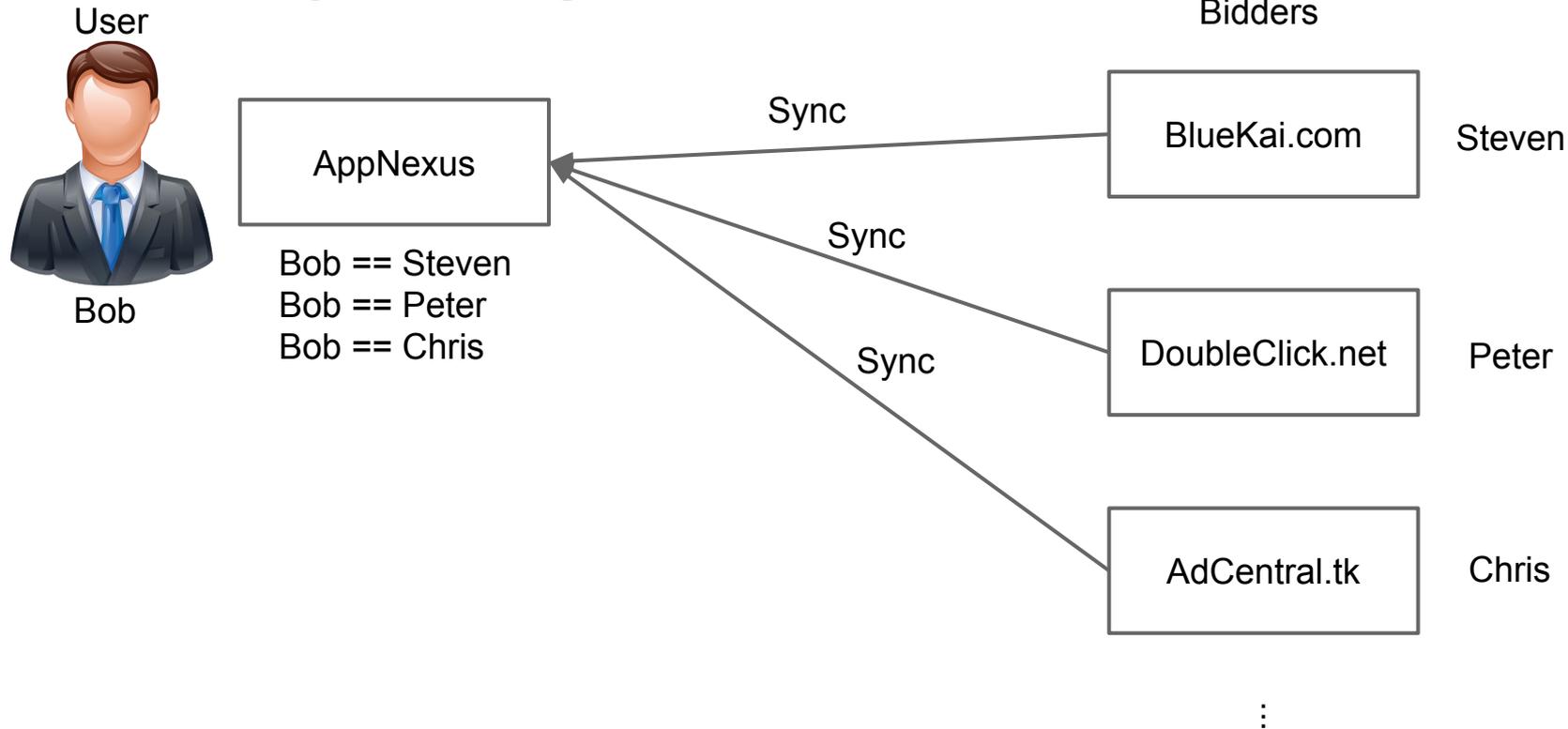
Fingerprinting

Cookie Syncing

Cookie Syncing



Cookie Syncing



Cookie Syncing - Experiments

- Top 3000 sites crawled while accumulating the state using OpenWPM
- Procedure:
 1. Detect identifying cookies and extract identifiers
 2. Search for identifiers in request / redirect URLs
 3. Mark all pairs of third-parties which share IDs

Cookie Syncing - Results

- Olejnik et al. finds **125 companies** syncing cookies (NDSS'14)
- We find **40%** of all tracking IDs are synced
- On average, **3.4 domains** see each ID
 - Turn.com's ID cookie is seen by **43 domains**

Why do we care?

This can make it much more difficult to start with a fresh profile.

I want to start a new profile

Clear cookies (from the beginning of time)?

→ Oops, forgot to check “Cache”

→ Forced cached PNGs respawns cookies

I want to start a new profile

Clear all local browser state?

→ Your browser fingerprint didn't change.

→ You visit site again

→ Site links your old history to new cookie using fingerprint

I want to start a new profile

Okay, so a few parties fingerprinted me or respawned cookies -- no big deal, most of my tracking history is gone?

→ respawned ID is cookie-synced to a large ad exchange

→ fingerprint linkage may still be shared through business relationships

I want to start a new profile

Well maybe I'll clear all state and simultaneously change some fonts and browser plugins?

→ You *might* be okay.

→ We observed one instance of ID respawning through (what we assume is) passive, server-side fingerprinting.

Defenses

Opt-out

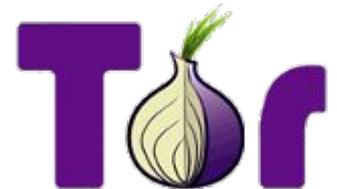
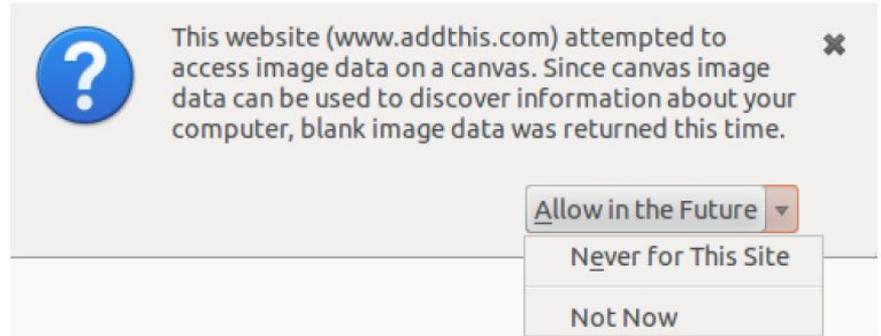
- opted-out on...
 - Network Advertising Initiative (NAI)
 - European Interactive Digital Advertising Alliance (EDAA)
 - AddThis' own Data Collection Opt-Out website
- no change for canvas fingerprinting and evercookies
- # of IDs involved in cookie sync: reduced by 30%
- # of parties involved in cookie sync: reduced by 5%

Cookie Blocking and DNT

- DNT reduces # of parties and # of IDs in cookie syncing by 3%
- Blocking all third-party cookies reduces by a factor of two
- Ad-blockers?
 - Ghostery cuts HTTP request traffic in half

Defenses - Canvas Fingerprinting

- no option to disable canvas
- Tor Browser returns an empty image
- no significant steps from browser manufacturers
 - pessimistic, “lost cause”





GeoIP® Databases and Services

Industry leading IP intelligence

Use GeoIP intelligence for content personalization, ad targeting, traffic analysis, digital rights management, and more.

[LEARN MORE](#)

m
Too

Preve
manu

CanvasFingerprintBlock

Blocked **1** potential HTML canvas fingerprinting attempt on this page

Prevented a script on <https://www.maxmind.com> from capturing the following 300px × 150px canvas:



[Donate](#)

[LEARN MORE](#)

Try our GeoIP2 Precision service demo:

Enter an IP address

[GO](#)

Conclusions

Move along, nothing to see here

- Significant media & public attention
- Top two providers (AddThis, Ligatus) stopped using canvas fingerprinting
 - calling it an “experiment”
- community made anti-canvas fingerprinting extensions



Takeaways

- third-party tracking isn't getting any better
- defense is hard, opt-out is not effective
- fingerprinting is becoming more common
- can automation help?

Thanks for listening :)



More info, code and the data:

- <https://securehomes.esat.kuleuven.be/~gacar/persistent/>
- <https://webtap.princeton.edu/>



the web never forgets

Images Videos

The Web never forgets: Persistent tracking mechanisms in the wild

The Web never forgets: Persistent tracking mechanisms in the wild is the first large-scale study of three advanced **web** tracking mechanisms - canvas fingerprinting, evercookies and use of "cookie syncing" in conjunction with evercookies.

> securehomes.esat.kuleuven.be