

CS 6431

# Web Tracking and Fingerprinting

Vitaly Shmatikov



New Yorker Collection 1993 Peter Steiner  
m cartoonbank.com. All rights reserved.

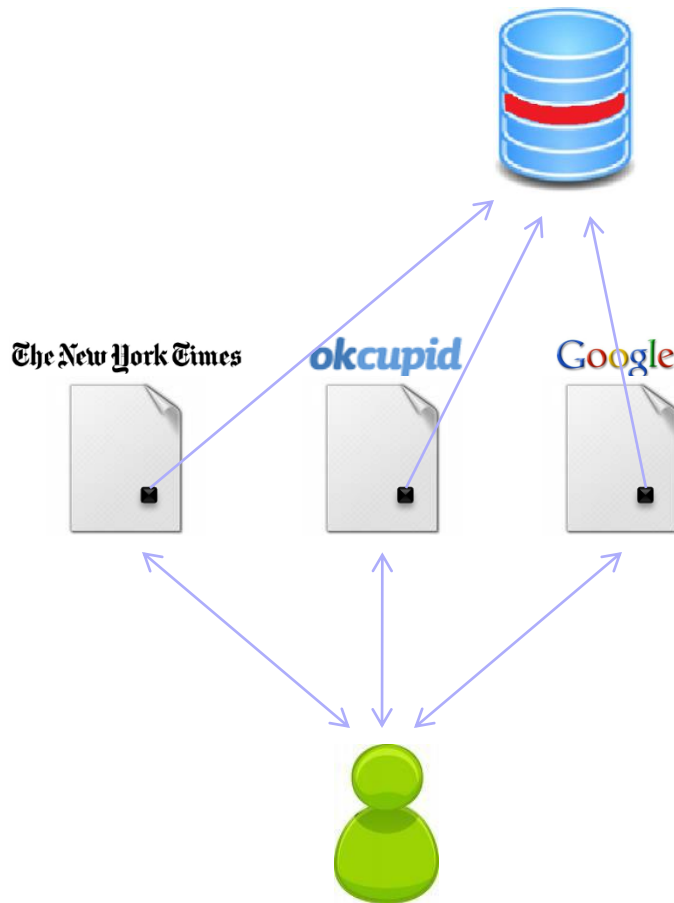
*It's the Internet! Of course they know you're a dog.  
They also know your favorite brand of pet food and  
the name of the cute poodle at the park that you  
have a crush on!*

# Tracking via Cookies

---

- ◆ **Cookie**: value set by Web server, automatically sent by the browser on subsequent requests to same(ish) origin
- ◆ Link two sessions at same site
- ◆ Link sessions between different sites (third-party cookies)
- ◆ Can be combined with user-identifying information

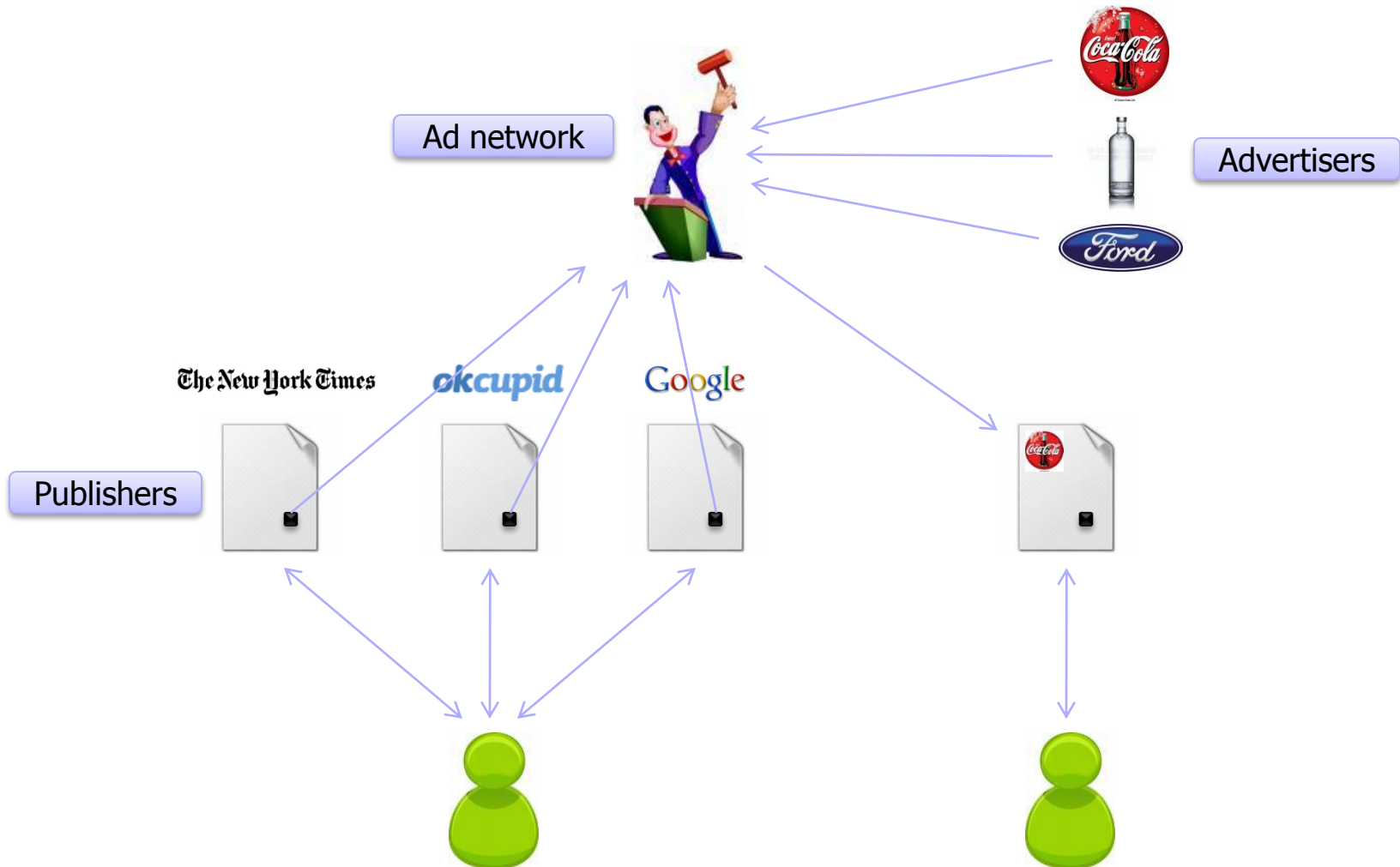
# Third-Party Tracking



## Third-party cookies:

- Disabled by default (Safari)
  - Can be disabled by user (many browsers)
  - Cannot be disabled (Android)
- ... but there are many other tracking technologies

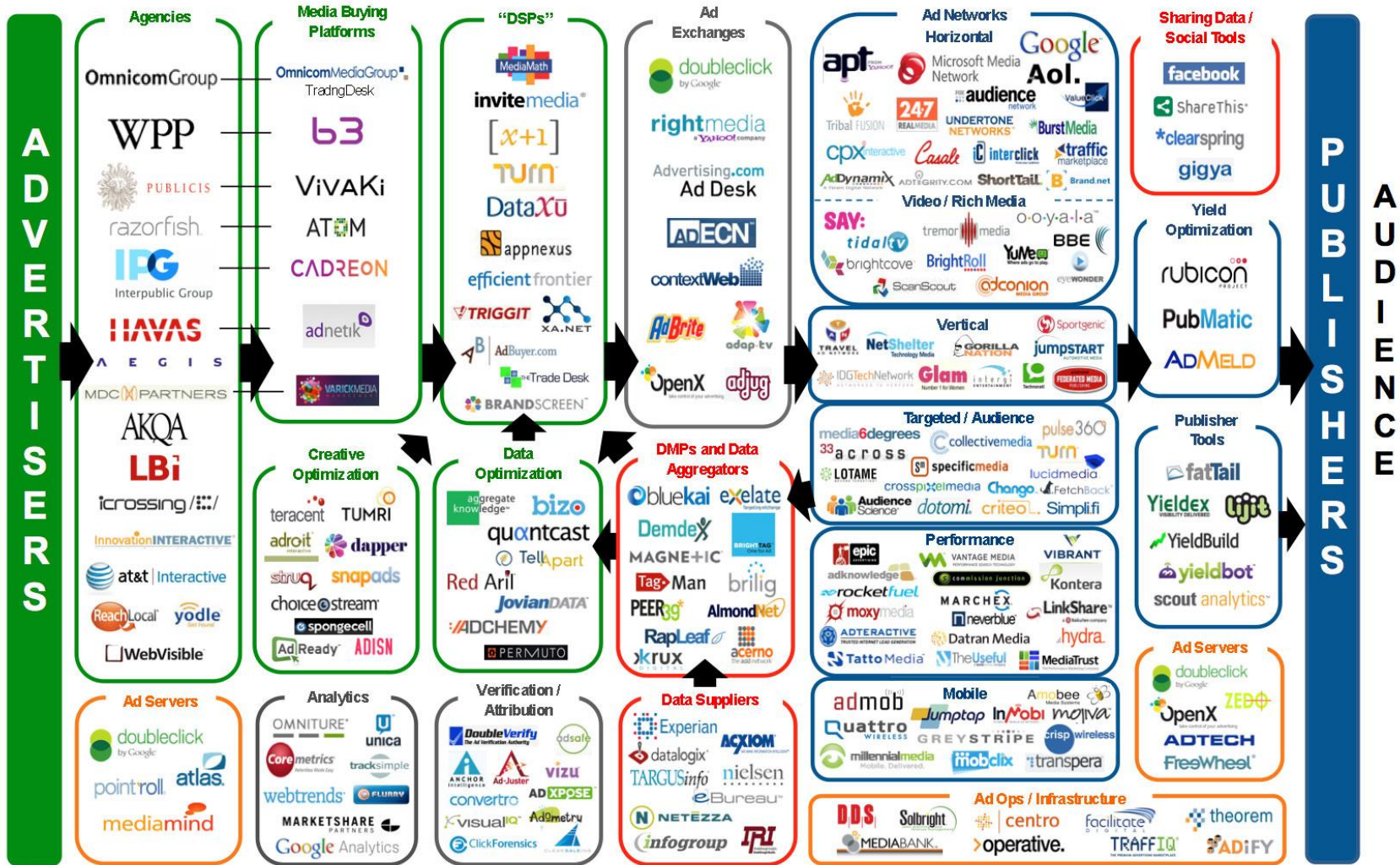
# Behavioral Targeting



# Partial List of Ad Networks

<a href="#">24/7 Real Media</a>	<a href="#">33Across</a>	<a href="#">Acerno</a>	<a href="#">Acxiom Relevance-X</a>	<a href="#">AdAdvisor</a>	<a href="#">AdBrite</a>
<a href="#">Adify</a>	<a href="#">AdInterax (Yahoo!)</a>	<a href="#">AdJuggler</a>	<a href="#">AdShuffle</a>	<a href="#">ADTECH (AOL)</a>	<a href="#">Advertising.com (AOL)</a>
<a href="#">Aggregate Knowledge</a>	<a href="#">Akamai</a>	<a href="#">AlmondNet</a>	<a href="#">Atlas (Microsoft)</a>	<a href="#">AudienceScience</a>	<a href="#">Bizo</a>
<a href="#">Blue Kai</a>	<a href="#">BlueLithium (Yahoo!)</a>	<a href="#">Bluestreak</a>	<a href="#">BrightRoll</a>	<a href="#">BTBuckets</a>	<a href="#">Burst Media</a>
<a href="#">Casale Media</a>	<a href="#">Chitika</a>	<a href="#">ChoiceStream</a>	<a href="#">ClickTale</a>	<a href="#">Collective Media</a>	<a href="#">comScore VoiceFive</a>
<a href="#">Coremetrics</a>	<a href="#">Cossette</a>	<a href="#">Criteo</a>	<a href="#">Effective Measure</a>	<a href="#">Eloqua</a>	<a href="#">Eyeblander</a>
<a href="#">eXelate</a>	<a href="#">EyeWonder</a>	<a href="#">e-planning</a>	<a href="#">Facilitate Digital</a>	<a href="#">FetchBack</a>	<a href="#">Flashtalking</a>
<a href="#">Fox Audience Network</a>	<a href="#">FreeWheel</a>	<a href="#">Google</a>	<a href="#">Hurra</a>	<a href="#">interCLICK</a>	<a href="#">Lotame</a>
<a href="#">Navegg</a>	<a href="#">NextAction</a>	<a href="#">NexTag</a>	<a href="#">Mediaplex (ValueClick Media)</a>	<a href="#">Media 6 Degrees</a>	<a href="#">Media Math</a>
<a href="#">Microsoft</a>	<a href="#">MindSet Media</a>	<a href="#">Nielsen Online</a>	<a href="#">nugg.ad</a>	<a href="#">Omniture</a>	<a href="#">OpenX</a>
<a href="#">Outbrain</a>	<a href="#">PointRoll</a>	<a href="#">PrecisionClick</a>	<a href="#">Pulse 360</a>	<a href="#">Quantcast</a>	<a href="#">Quigo (AOL)</a>
<a href="#">richrelevance</a>	<a href="#">Right Media (Yahoo!)</a>	<a href="#">Rocket Fuel</a>	<a href="#">Safecount *</a>	<a href="#">ScanScout</a>	<a href="#">Smart Adserver</a>
<a href="#">Snoobi</a>	<a href="#">Specific Media</a>	<a href="#">TACODA (AOL)</a>	<a href="#">Tatto Media</a>	<a href="#">Tealium</a>	<a href="#">TradeDoubler</a>
<a href="#">Traffic Marketplace</a>	<a href="#">Tribal Fusion / Exponential</a>	<a href="#">TruEffect</a>	<a href="#">Tumri</a>	<a href="#">Turn</a>	<a href="#">Undertone Networks / Zedo</a>
<a href="#">ValueClick Media</a>	<a href="#">Vizu</a>	<a href="#">Weborama</a>	<a href="#">WebTrends</a>	<a href="#">Yahoo!</a>	<a href="#">[x+1]</a>

# Display Advertising Technology Landscape





# 2012 DISPLAY ADVERTISING ECOSYSTEM EUROPE

PUBLISHERS

ADVERTISERS





# Tracking Is Pervasive

---

64

independent tracking mechanisms in an  
average top-50 website

# Sticky Tracking

---

Subverting same origin policy  
(publisher also runs an ad network)

[ad.hi5.com](http://ad.hi5.com) = [ad.yieldmanager.com](http://ad.yieldmanager.com)

Flash cookies

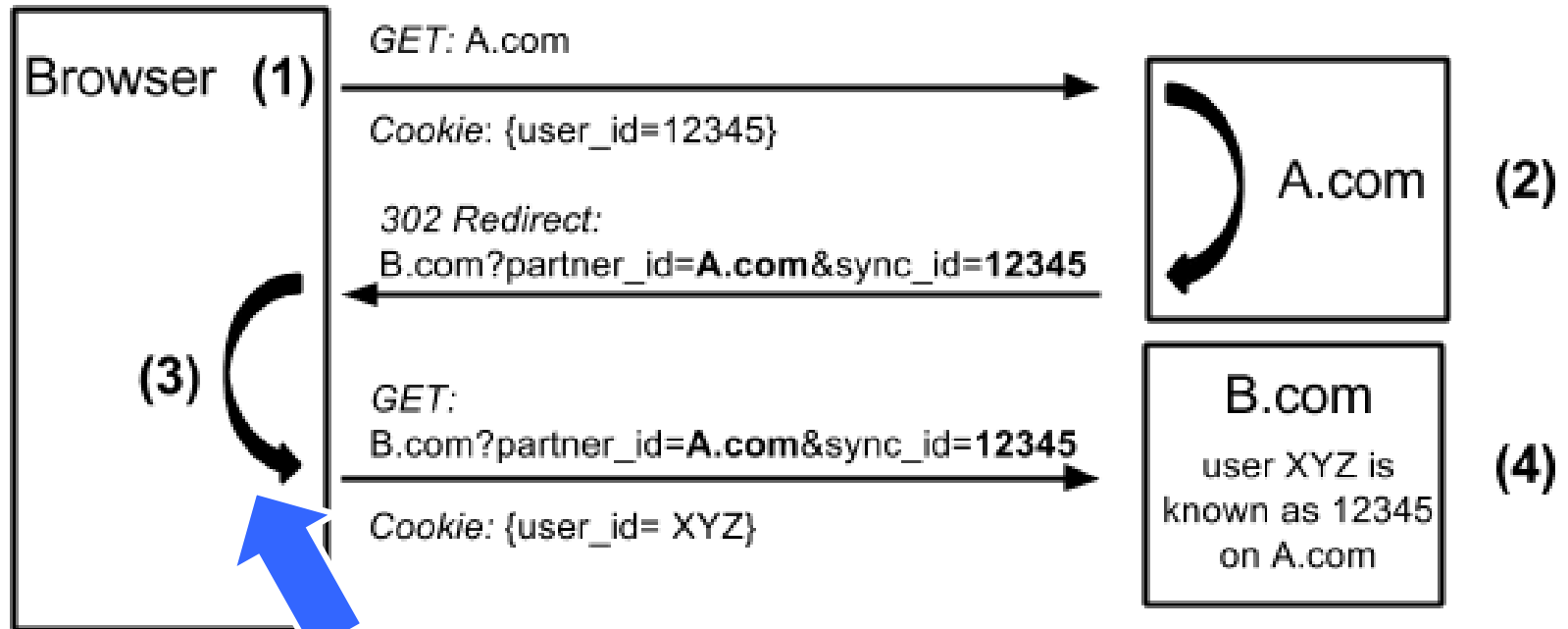


Browser fingerprinting



History sniffing

# Cookie Syncing



Site A informing site B about user's identity (via user's browser)

Allows aggregation across multiple trackers

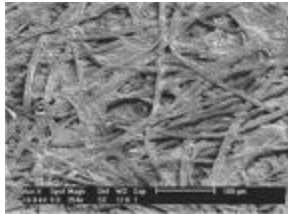
# Tracking Technologies

---

- ◆ HTTP Cookies
- ◆ HTTP Auth
- ◆ HTTP Etags
- ◆ Content cache
- ◆ IE userData
- ◆ HTML5 protocol and content handlers
- ◆ HTML5 storage
- Flash cookies
- Silverlight storage
- TLS session ID & resume
- Browsing history
- window.name
- HTTP STS
- DNS cache

# Everything Has a Fingerprint

---





# Fingerprinting Web Browsers

---

- ◆ User agent
- ◆ HTTP ACCEPT headers
- ◆ Browser plug-ins
- ◆ MIME support
- ◆ Clock skew
- Installed fonts
- Cookies enabled?
- Browser add-ons
- Screen resolution



A research project of the [Electronic Frontier Foundation](#)

# Panopti**cl**ick

How Unique – and Trackable – Is Your Browser?

Is your browser configuration rare or unique? If so, web sites

Your browser fingerprint **appears to be unique** among the 3,435,834 tested so far

you see how easily identifiable you might be as you surf the web.

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



A paper reporting the statistical results of this experiment is now available: [How Unique Is Your Browser?](#), Proceedings of the Privacy Enhancing Technologies Symposium (PETS 2010), Springer Lecture Notes in Computer Science.

[Learn about Panopti\*\*cl\*\*ick and web tracking.](#)

[The Panopti\*\*cl\*\*ick Privacy Policy.](#)

[Learn about the \[Electronic Frontier Foundation\]\(#\).](#)

# Panoptlick Example

Plugin 0: Adobe Acrobat; Adobe Acrobat Plug-In Version 7.00 for Netscape; nppdf32.dll; (Acrobat Portable Document Format; application/pdf; pdf) (Acrobat Forms Data Format; application/vnd.fdf; fdf) (XML Version of Acrobat Forms Data Format; application/vnd.adobe.xfdf; xfdf) ( Acrobat XML Data Package; application/vnd.adobe.xdp+xml; xdp) (Adobe FormFlow99 Data File; application/vnd.adobe.xfd+xml; xfd). Plugin 1: Adobe Acrobat; Adobe PDF Plug-In For Firefox and Netscape; nppdf32.dll; (Acrobat Portable Document Format; application/pdf; pdf) (Acrobat Forms Data Format; application/vnd.fdf; fdf) (XML Version of Acrobat Forms Data Format; application/vnd.adobe.xfdf; xfdf) ( Acrobat XML Data Package; application/vnd.adobe.xdp+xml; xdp) (Adobe FormFlow99 Data File; application/vnd.adobe.xfd+xml; xfd). Plugin 2: Google Gears; Google Gears; nppdf32.dll; (Acrobat Portable Document Format; application/pdf; pdf) (Acrobat Forms Data Format; application/vnd.fdf; fdf) (XML Version of Acrobat Forms Data Format; application/vnd.adobe.xfdf; xfdf) ( Acrobat XML Data Package; application/vnd.adobe.xdp+xml; xdp) (Adobe FormFlow99 Data File; application/vnd.adobe.xfd+xml; xfd). Plugin 3: Microsoft® Windows Media Player Firefox Plugin; np-mswmp; np-mswmp.dll; (np-mswmp; application/x-ms-wmp; \*) (; application/asx; \*) (; video/x-ms-asf-plugin; \*) (; application/x-mplayer2; \*) (; video/x-ms-asf; asf,asx,\*) (; video/x-ms-wm; wm,\*) (; audio/x-ms-wma; wma,\*) (; audio/x-ms-wax; wax,\*) (; video/x-ms-wmv; wmv,\*) (; video/x-ms-wvx; wvx,\*). Plugin 4: Move Media Player; npmnqmp 07103010; npmnqmp07103010.dll; (npmnqmp; application/x-vnd.moveplayer.qm; qmx,qpl) (npmnqmp; application/x-vnd.moveplay2.qm; ) (npmnqmp; application/x-vnd.movenetworks.qm; ). Plugin 5: Mozilla Default Plug-in; Default Plug-in; npnul32.dll; (Mozilla Default Plug-in; \*; \*). Plugin 6: Shockwave Flash; Shockwave Flash 10.0 r32; NPSWF32.dll; (Adobe Flash movie; application/x-shockwave-flash; swf) (FutureSplash movie; application/futuresplash; spl). Plugin 7: Windows Genuine Advantage; 1.7.0059.0; npLegitCheckPlugin.dll; (npLegitCheckPlugin; application/WGA-plugin; \*).

**84% of browser fingerprints are unique**

**With Flash or Java, 94% are unique**

# <CANVAS>

---

- ◆ Programmatic drawing in the browser
  - Draw shapes, add text, 3D (via WebGL)
- ◆ Access to drawn pixels
  - Array of RGBA values
  - PNG-encoded data URL

# Text Rendering ...

---

```
<script type="text/javascript">
  var canvas =
    document.getElementById("drawing");
  var context = canvas.getContext("2d");
  context.font = "18pt Arial";
  context.textBaseline = "top";
  context.fillText("Some letters", 2, 2);
</script>
```



# ... Text Inspection

---

```
<script type="text/javascript">
  var canvas =
    document.getElementById("drawing");
  var context = canvas.getContext("2d");
  context.font = "18pt Arial";
  context.textBaseline = "top";
  context.fillText("Some letters", 2, 2);

  var pixels =
    canvas.toDataURL("image/png");
</script>
```

# WebFonts

---

- ◆ Problem: Clients ship with ugly fonts
- ◆ Solution: Browsers should download fonts from the Internet on demand!

```
@font-face { font-family: 'Sirin Stencil';  
font-style: normal; font-weight: 400; src:  
url(http://themes.googleusercontent.com/stat  
ic/fonts/sirinstencil/v1/[...].woff)  
format('woff'); }
```

# 45 Ways To Sirin Stencil

```
context.font = "12pt 'Sirin Stencil'";
```

## Windows

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

## OS X

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

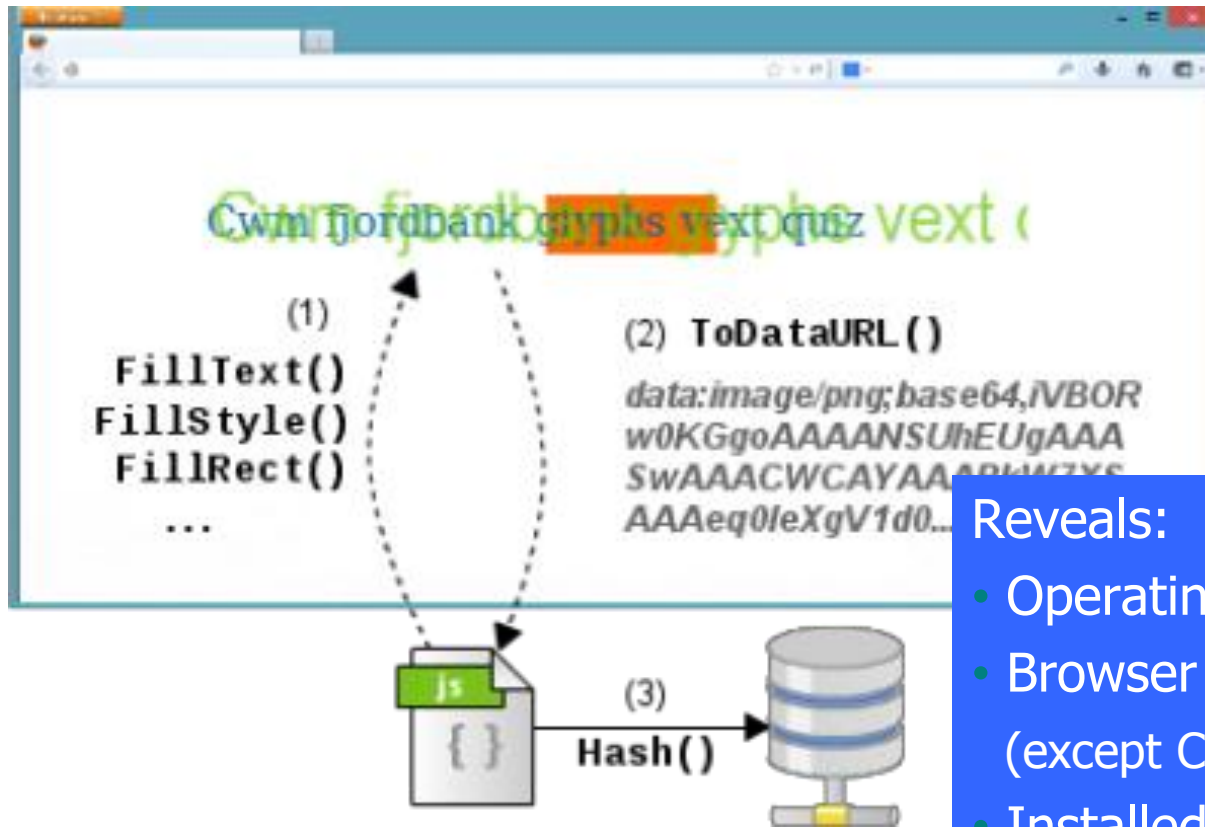
## Linux

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

How quickly daft jumping zebras vex. (Also, punctuation: &/c.)

# Canvas Fingerprinting

[Mowery and Shacham.  
"Pixel Perfect". W2SP 2012]



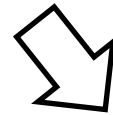
## Reveals:

- Operating system family
- Browser family  
(except Chrome, Safari on OS X)
- Installed fonts
- Font smoothing parameters

# How Pervasive?

[Acar et al. "The Web Never Forgets". CCS 2014]

- ◆ Present in 5.5% of top 100,000 websites
- ◆ Fingerprinting code comes from 20 different domains
  - addthis.com by far the most popular (95%)



Cwm fjordbank glyphs vext quiz  
<http://valve.github.io>  
<http://admicro.vn/>  
<http://www.plentyoffish.com>

Draws

Cwm fjordbank glyphs vext quiz  
into the canvas

Why this text?

ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789+/'



# “Don’t Worry, It’s All Anonymous”

---

- ◆ Is it?
- ◆ What’s the difference between
  - “anonymous”
  - “pseudonymous”
  - “identified”
- ◆ Which technology changed data collection from anonymous to pseudonymous?

# How Websites Get Your Identity

---

Third party is sometimes the site itself

## Leakage of identifiers

```
GET http://ad.doubleclick.net/adj/...
```

```
Referer: http://submit.SPORTS.com/...?email=jdoe@email.com
```

```
Cookie: id=35c192bcfe0000b1...
```

## Security bugs

XSUH: cross-site URL hijacking

Third party buys your identity

Syphilis - NHS Choices

http://www.nhs.uk/conditions/syphilis/pages/introduction.aspx

Home | About | Contact | Communities | Tools | Video | Choose and Book

Log in or create an account

**NHS choices** Your health, your choices

Enter a search term **Search**

Health A-Z | **Live Well** | Carers Direct | Health news | Find and choose services

# Syphilis

Share Save Easy print Like 5

Overview | **Map of Medicine** | Medicines info | Clinical trials

Syphilis | Symptoms | Causes | Diagnosis | Treatment | Complications | Prevention

## Introduction

Is your sex life putting your health at risk? Take the test and find out more.

How safe is your sex life?

QUIET PLEASE SAFE SEX TEST

Type your first name here

START

NHS choices

Syphilis is a bacterial infection that is usually passed on through having sex with someone who is infected. It can also be passed from an infected mother to her unborn child and, in rare cases, can be caught through injecting drugs.

It is extremely rare to catch syphilis through a blood transfusion in the UK as blood donors are carefully screened.

### Three stages of disease

**Stage 1 (primary syphilis).** Symptoms of syphilis begin with a painless but highly infectious sore on the genitals or sometimes around the mouth. If somebody else comes into close contact with the sore, typically during sexual contact, they can also become infected. The sore lasts two to six weeks before disappearing.

**Stage 2 (secondary syphilis).** Secondary symptoms, such as a skin rash and sore throat, then develop. These symptoms may disappear within a few weeks, after which you experience a latent (hidden) phase with no symptoms, which can last for years. After this, syphilis can progress to its third, most dangerous stage.

**Stage 3 (tertiary syphilis).** At this stage, it can cause serious damage to the body.

The primary and secondary stages are when you are most infectious to other people. In the latent phase (and usually around two years after becoming infected), syphilis cannot be passed onto others but can still cause symptoms. See Symptoms of syphilis for more information on the

### Useful links

**NHS Choices links**

- Video: gay healthcare
- Video: condom negotiation
- Live Well: condoms
- Live Well: drugs
- Health A-Z: HIV and AIDS
- Health A-Z: STIs
- Find sexual health services
- Infections you can catch through oral sex

**External links**

- British Association for Sexual Health and HIV
- Brook: for under-25s
- FPA: sexual health
- Health Protection Agency: syphilis
- Lab Tests Online: syphilis test
- Men's Health Forum

**Screening and testing for gays and lesbians**

Research shows that gay men and lesbians are less likely to have NHS screening and testing than heterosexuals. But it's important.

# History Sniffing

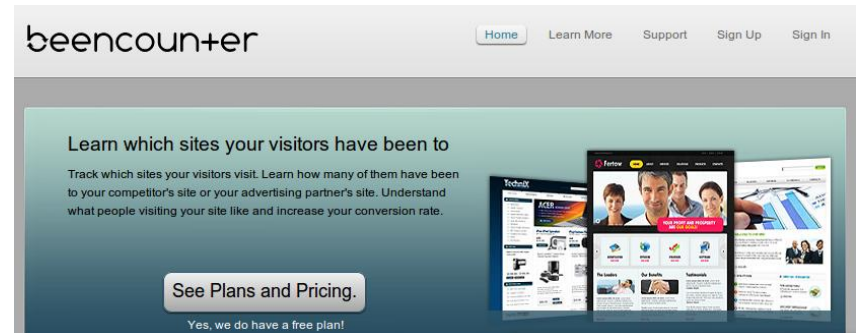
How can a webpage figure out which sites you visited previously?

## ◆ Color of links

- CSS :visited property
- getComputedStyle()

## ◆ Cached Web content timing

## ◆ DNS timing



# Identity Sniffing

[Wondracek et al. Oakland 2010]

- ◆ All social networking sites allow users to join groups
- ◆ Users typically join multiple groups
  - Some of these groups are public
- ◆ Group-specific URLs are predictable

```
http://www.facebook.com/group.php?gid=[groupID]&v=info&ref=nf+  
https://www.xing.com/net/[groupID]/forums+
```

- ◆ Intersection of group affiliations acts as a fingerprint
  - Can sometimes infer identity by computing the intersection of group membership lists



# One-Click Fraud

[Cristin et al. CCS 2010]

*Thank you for your patronage! You successfully registered for our premium online services, at an incredible price of 50,000 JPY. Please promptly send your payment by bank transfer to ABC Ltd at Ginko Bank, Account 1234567. Questions? Please contact us at 080-1234-1234.*

*Your IP address is 10.1.2.3, you run Firefox 3.5 over Windows XP, and you are connecting from Tokyo.*

*Failure to send your payment promptly will force us to mail you a postcard reminder to your home address. Customers refusing to pay will be prosecuted to the fullest extent of the law. Once again, thank you for your patronage!*



# One-Click Fraud

◆ Estimated costs to victims:  
USD 260 million / year

◆ What's going on here?

◆ Why only Japan?

- Cultural factors:  
susceptibility to authoritative language  
threat of public shaming



Credible because the website  
does have your real identity!



# Instant Personalization

Now in the UK!

Yelp is using Facebook to personalize your experience. Options Friends' Activity 0 Sign Up for Yelp Log In

**yelp**  
Real people. Real reviews.®

Search for (e.g. taco, cheap dinner, Max's)

Welcome About Me Write a Review Find Reviews Invite Friends

Are You Looking For **Yelp New York?**

Atlanta Austin Berkeley Boston Brooklyn Chicago Dallas Denver Detroit Houston Los Angeles London Los Angeles Orange County Palo Alto

**Yelp New York**

Yelp is the fun and easy way to find and talk about great things in your city. Sign Up Now

**Best of yelp** More "Best Of" »

**Restaurants**  
16360 reviewed

**Shopping**  
25112 reviewed

**Browse by Category**

- Professional Services 27589
- Shopping 25529
- Health and Medical 20221
- Restaurants 16712
- Home Services 15377
- Food 10222
- Local Services 7743
- Beauty and Spas 6035
- Automotive 4284
- Event Planning & Services
- Education 3777

**Hey, 4 of your friends have joined Yelp!**  
Sign up and never miss their reviews

**Review of the Day** Archive »

Voted by our members!

1. **Graham Avenue Meats...**  
2. **Tofu Guy**  
3. **G Esposito & Sons**

1. **Fuego 718**  
2. **Park Slope Eye**  
3. **10/10 Optics**

Alicia L. bookmarked Americas Travel. Lers Ros Thai, Vietnam House, Bodega Bistro, The Brick Yard Restaurant & Bar, Marina Nails, Pazzia Restaurant & Pizzeria, Red Door Cafe, Jackson Hewitt Tax Service, Beretta

amit n. reviewed Toyota of Long Beach

amit n. added a photo

Jeremy P. reviewed La Duni Latin Kitchen & Coffee Studio

All Friends' Activity

Member Search

Hide

**Creepy is the new normal**

# Do Not Track



## Basics

HTTP header

- [DNT: 1](#)

Standardization

Browser support in FF4, IE9

Beginning to see adoption  
(AP, NAI)... or not

## Privacy protections

No tracking across sites

- Who is the “third” party?

Can't be based on domain 

Example: amazonaws.com, ad.hi5.com ...

No intrusive tracking

Limits on regular log data

Exceptions for fraud  
prevention, etc.

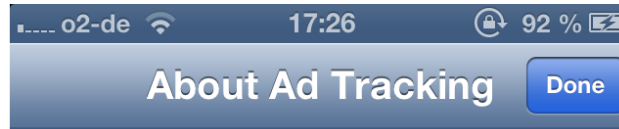
# DNT Adoption Issues

---

“But the NAI code also recognizes that companies sometimes need to continue to collect data for operational reasons that are separate from ad targeting based on a user’s online behavior. For example, online advertising companies may need to gather data to prove to advertisers that an ad has been delivered and should be paid for; to limit the number of times a user sees the same ad; or to prevent fraud.”

Translation: we’re going to keep tracking you, but we’ll simply call it “operational reasons.”

# Brave New World?



## Ad Tracking

iOS 6 introduces the Advertising Identifier, a non-permanent, non-personal, device identifier, that advertising networks will use to give you more control over advertisers' ability to use tracking methods. If you



**How are these identifiers different from third-party cookies?**



Advertising Identifier. However, advertising networks transition to the Advertising Identifier you may have targeted ads from other

Ads from the ground up to tackle today's marketing challenges, like reaching people across devices and bridging the gap between online impressions and offline purchases

