

Online Tracking

Vitaly Shmatikov

Slides courtesy of Arvind Narayanan

Reading Assignment

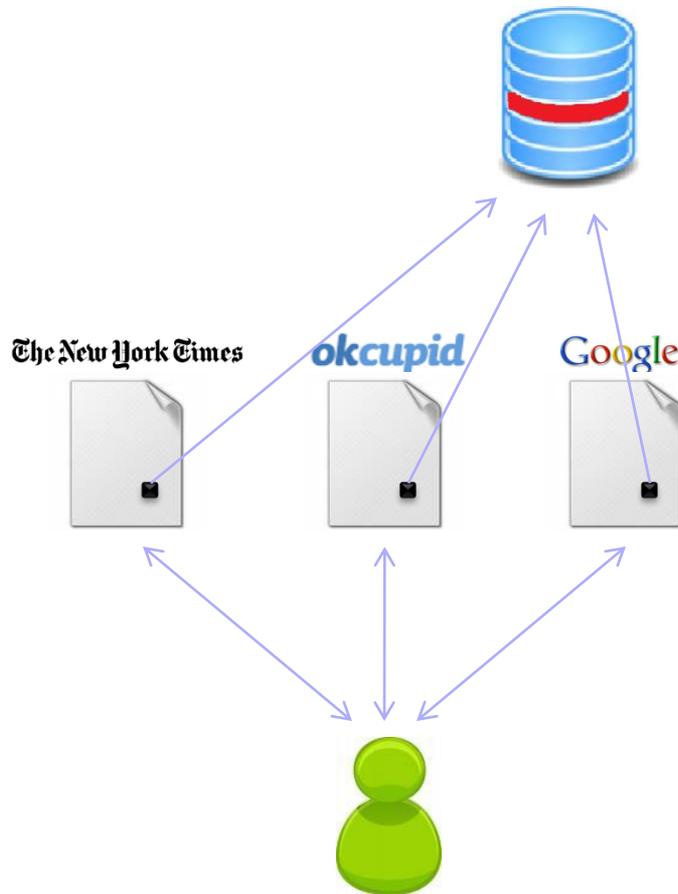
- ◆ “Third-Party Web Tracking: Policy and Technology”
- ◆ “Cookieless Monster: Exploring the Ecosystem of Web-based Device Fingerprinting”



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!*

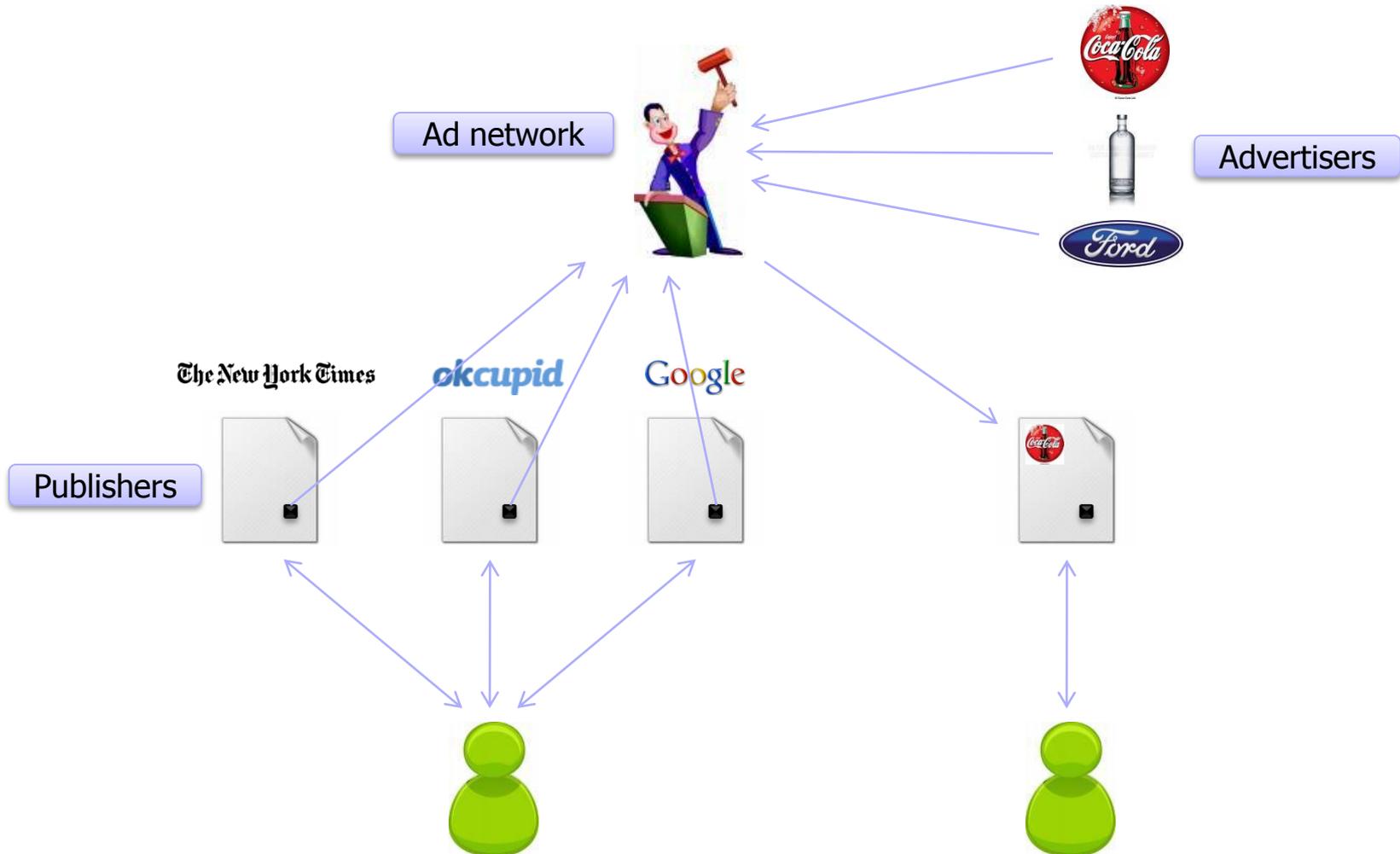
Third-Party Tracking



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

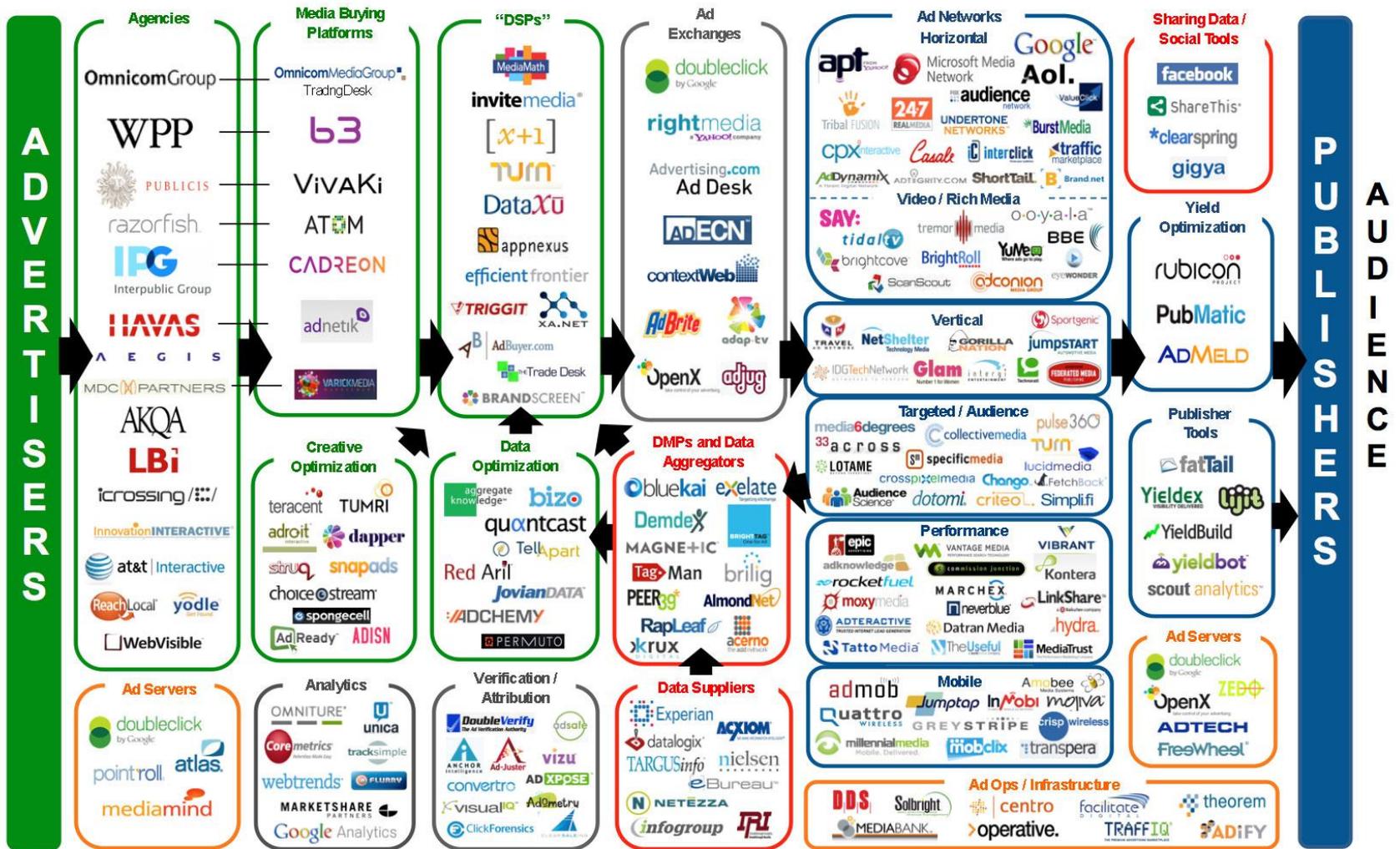
Behavioral Targeting



Partial List of Ad Networks

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

Display Advertising Technology Landscape



2012 DISPLAY ADVERTISING ECOSYSTEM EUROPE

PUBLISHERS

ADVERTISERS

Data Suppliers	CACI, DoubleClick, Epsilon, comScore, EQUIFAX, Experian, ACQUOM, nielsen, dunhumby, Almond		
Data Management Platforms	XRUx, LOTAME, Audience Science, TURN, bluekal, enreach, exelate, Demand		
Data Exchanges	Adatus, exelate, quantcast, dotvantage, weborama		
Sales Houses	Ad Networks	Demand Side Platforms	Agencies
InteractiveMedia, SanomaMedia, Yahoo!, FOX, AOL, etc.	ad pepper, Microsoft Advertising, etc.	Turn, etc.	WPP, Omnicom Media Group, Havas, dentsu, IPG, etc.
SSP & Private Ad Exchanges	Agency Trading Desks	Delivery Systems, Tools & Analytics	Trading Desks
Improve Digital, Admeld, Rubicon, etc.	Cadreon, Vivaki, etc.	DoubleClick, Appnexus, 24/7, etc.	Intelligence, Moxod, etc.
Delivery Systems, Tools & Analytics	Audience Targeting / Re-targeting	Verification & Privacy	Published by
Adtech, Appnexus, etc.	Effigee, etc.	DoubleVerify, Adsafe, etc.	IMPROVE DIGITAL
Verification & Privacy	Ad Exchanges		
DoubleVerify, Adsafe, Evidon, etc.	Microsoft Advertising, HiMedia, etc.		

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 = ad.yieldmanager.com

Flash cookies



Browser fingerprinting

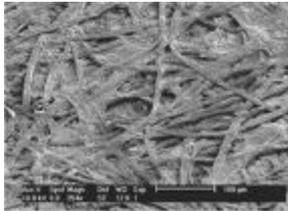


History sniffing

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](#)

Panopticlick

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 Panopticlick and web tracking.](#)

[The Panopticlick 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 Update; Google Update; npGoogleOneClick8.dll; (; application/x-vnd.google.oneclickctrl.8;). 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

“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

Remember XSS (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 PLEASE SAFE SEX TEST

Type your first name here

START



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

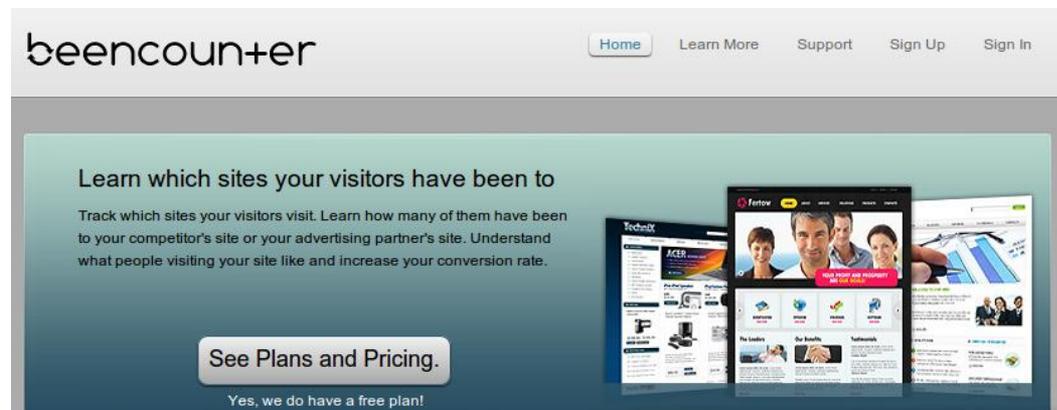
How can a webpage figure out which sites you visited previously?

◆ Color of links

- CSS :visited property
- getComputedStyle()

◆ Cached Web content timing

◆ DNS timing



The screenshot shows the homepage of beencounter. The header includes the logo 'beencounter' and navigation links: Home, Learn More, Support, Sign Up, and Sign In. The main content area features a large green banner with the text: 'Learn which sites your visitors have been to. Track which sites your visitors visit. Learn how many of them have been to your competitor's site or your advertising partner's site. Understand what people visiting your site like and increase your conversion rate.' Below this text is a button labeled 'See Plans and Pricing.' and a smaller line of text: 'Yes, we do have a free plan!'. To the right of the text are three overlapping images of various websites, including Techradar, Acer, and a site with a bar chart.

Identity Sniffing

[Wondracek et al.]

- ◆ 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



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 Co Palo Alto

Yelp New York

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

All Friends' Activity 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

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 ability to choose the networks you may not want you target advertising use the until advertising using the still receive networks



How are these identifiers different from third-party cookies?