Privacy and Tracking on the Internet

Dan S. Wallach
Department of Computer Science
Rice University
Seen in my postal mail...
“Customer code” identifies the consumer.
“First party” identifiers

Relationships between consumer and vendor

Supermarket “rewards” cards
Vendor-specific credit cards / gift cards
Unique IDs printed on receipts
Credit card numbers *
“First party” identifiers

Relationships between consumer and vendor
Supermarket “rewards” cards
Vendor-specific credit cards / gift cards
Unique IDs printed on receipts
Credit card numbers *

Map consumers to database records
Track your purchasing history
Provide you with targeted advertising and coupons

Big data allows for powerful inferences (e.g., pregnancy)

www.nytimes.com/2012/02/19/magazine/shopping-habits.html
Use the coupon, connect your purchases to you, even if you pay cash.
I’m refinancing my mortgage and I got this in the mail.
I’m refinancing my mortgage and I got this in the mail.

“You have received this letter as a result of our relationship with the national credit bureaus. We were notified of your recent mortgage inquiry…”

“To opt out of future correspondence you may write to Experian…”
“Third party” relationships
“Third party” relationships
“Third party” relationships
Web Ads: First Party
Simple display ads

Some web sites host their own ads


<img src="http://graphics8.nytimes.com/adx/images/ADS/32/05/ad.320528/NYT_DOT_LEFT.jpg" width="184" height="90" border="0"></a>
If the user clicks the ad

GET /adx/bin/adx_click.html? ...

User

nytimes.com
If the user clicks the ad

GET /adx/bin/adx_click.html? ...

HTTP/1.1 302 Found
Location: http://www.marcjacobs.com/?utm_source=nyt ...
If the user clicks the ad

Main web site tells the user’s browser to redirect the advertiser, give it credit for the click.
Targeted advertising

Option #1: Search queries.
If a user is searching for a product, they might buy it.
Targeted advertising

Option #2: User profiling
If an advertiser knows something specific about a user, they can customize advertising to them.

More Recommendations for You

- OXO Good Grips 3-Piece Angled...
  - Rating: 4 stars (78)
  - Price: $19.99
  - Why recommended?
- Kindle US Power Adapter
  - Rating: 4 stars (1,258)
  - Price: $19.99
  - Why recommended?
- Sentinel 3: Homeworld
  - Rating: 4 stars (194)
  - Price: $2.49
  - Why recommended?
- OXO Steel Cocktail Strainer
  - Rating: 4 stars (19)
  - Price: $9.99
  - Why recommended?
- 2-Year Protection Plan for Kindle
  - Service Plan: $29.99
  - Why recommended?
Gathering user data?

**Volunteered data**
Users tell Facebook all kinds of demographics
Users “like” vendors, products, bands, etc.

**Measured data**
Users’ history of purchases, visited web sites
IP geolocation (accurate to the city, if not better)

**Inferred data**
Users might be a lot like their friends / neighbors
Web Ads: Third Party
Typical 3rd party display ads

Similar operation to 1st party ads

<a target="_blank" href="http://ad.doubleclick.net/click;h=v8/3d41/4/a2/%2a/g;
255075316;0-0;0;78667076;3454-728/90;47210945/47227088/4;;~okv=;pc=aol
02555489;;~sscs=%3fhttp%3A//at.atwola.com/adlink
%2F5113%2F786827%2F0%2F225%2FAdId%3D2555489%3BBnId%3D1%3Bitime
%3D595358030%3Bdecode%3Dyes%3Bimpref%3D13545953581141452207%3Blink
78667076.&dcc=47210945.255075316">
<img src="http://s0.2mdn.net/viewad/1361550/All-New
+Altima_MSRP_728x90.jpg" alt="Advertisement" border="0">
</a>
If the user clicks the ad

User → GET /click;h=v8/3d41/4/a2/ ... → ad.doubleclick.net
If the user clicks the ad

User

GET /click;h=v8/3d41/4/a2/ ...

HTTP/1.1 302 Moved Temporarily
Location: http://at.atwola.com/adlink...

User

GET /adlink...

ad.doubleclick.net

at.atwola.com
If the user clicks the ad

DoubleClick redirects through Atwola then to destination.
Tracking Cookies & Mobile Geolocation
Back to our paper example
VALID FOR IN-STORE USE ONLY. PRESENT THIS COUPON
Valid for in-store use only. Copies not accepted. Limit one coupon, Savings Certificate, special offer or discount (including price match) per item. Coupon must be surrendered at time of purchase; any return of purchase will reduce your savings proportionately. The discount cannot be applied to gift cards, shipping, or sales tax. Offer excludes the following: Alessi, Arthur Court, Breville, Brita Collection, Brookstone®, DKNY, Kate Spade, Kenneth Cole Reaction Home, Kosta Boda, Le Creuset®, Lladro®, Monique Lhuillier, Nambe®, Nautica®, Noritake®, Riedel, Shun, Starbucks® Espresso, Swarovski, T-Fal, Vera Wang®, Victorinox Luggage, Viamex, Waterford®, Wusthof®, or Zwilling; Argintino®, babyybrezza™, Baby Jogger™, BEABA®, BOB, Bugaboo, Bumprider™, ERGOBaby™, Foundations®, iCandy®, Meci-Cosi®, Mountain Buggy, Oeuf, Orbit Baby™, Peg Pérego®, petunia pickle bottom®, Phil & Ted's, Plan Toys®, Quinny®, Svan®, Teutonia®, Under Armour®, UPPAbaby®, baby furniture, diapers, wipes, formula, baby food or portrait studio services.

MOBILE Firsttime subscribers get a 20% OFF one single item in-store or online mobile offer. Text OFFER to 239663.
Message and data rates may apply. Mobile internet access required. Up to 5 msg/month. Text STOP to 239663 to cancel. Send HELP to 239663 for help.

EMAIL Firsttime subscribers get a 20% OFF one single item e-mail offer for in-store use. Visit bedbathandbeyond.com/circulars.asp

CONNECT WITH US

FREE SHIPPING
FREE STANDARD SHIPPING WITH ANY ONLINE PURCHASE OF $99 OR MORE
Offer valid through 12.17.12 BEDBATHANDBEYOND.COM

We gladly accept competitors’ coupons for products also available at those competitors. Prices valid only in the contiguous U.S. If you find a lower price at any of our competitors, we will match that price. Exceptions may apply.

If you wish to opt-out of future mailings from us, go to our website or visit bedbathandbeyond.com/unssubscribe.asp

Dole

yOnanas

Ice Cream Treat Maker
• Turns frozen bananas and other fruit into a healthy treat that looks and tastes like soft serve ice cream
• Dishwasher safe parts
49.99

To learn more, scan the QR code.

Hollywood Popcorn Maker
• Makes up to 8 cups of popcorn
• Marquee header lights up and can be personalized, includes 150 letters
• Includes warming light and pull-out drawer
79.99

MERRY CHRISTMAS
BEBRATH & BEYOND

We wish you a Very Merry Christmas and Happy Holidays. Thank you for shopping with us.

MOVIE NIGHT
What happens if you scan this QR code with your phone?
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
  <title>QRCodeTracking</title>
  <meta http-equiv="Content-type" value="text/html; charset=utf-8"/>
  <script type="text/javascript" src="/tracking.js"></script>
  <script type="text/javascript" src="/js/mini.ajax.js"></script>
  <script type="text/javascript">
    function tagAction(result)
    {
      window.location="/936647AE1AE6C73/r";
      //show("content");
      document.body.innerHTML = res;
    }
  
    var res = null;
  </script>
</head>

<body>
  <div id="user_action">
    Waiting ...
  </div>
  <div id="content" style="display: none">
  </div>
  <script type="text/javascript">
    tracking('2957203_db7b71b43a3ad916d08ac54b0b2c86a8');
  </script>
</body>
</html>
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
<head>
<title>QRCodeTracking</title>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<meta http-equiv="Content-type" value="text/html; charset=utf-8" />
<script type="text/javascript" src="/tracking.js"></script>
<script type="text/javascript" src="/js/mini.ajax.js"></script>
<script type="text/javascript">
function tagAction(result)
{
    window.location="/936647AE1AE6C73/r";
    //show("content");
    document.body.innerHTML = res;
}

var res = null;
</script>
</head>
<body>
<div id="user_action">
Waiting ...
</div>
<div id="content" style="display: none">
</div>
<script type="text/javascript">
tracking('2957203_db7b71b43a3ad916d08ac54b0b2c86a8');
</script>
</body>
</html>
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8"/>
<title>QRCodeTracking</title>
<meta http-equiv="Content-type" value="text/html; charset=utf-8"/>
<script type="text/javascript" src="/tracking.js"></script>
<script type="text/javascript" src="/js/mini.ajax.js"></script>
<script type="text/javascript">
function tagAction(result)
{
    window.location="/936647AE1AE6C73/r";
    //show("content");
    document.body.innerHTML = res;
}

var res = null;
</script>
</head>
<body>
<div id="user_action">
Waiting ...
</div>
<div id="content" style="display: none">
<script type="text/javascript">
tracking('2957203_db7b71b43a3ad916d08ac54b0b2c86a8');
</script>
</div>
</body>
</html>
Cookies

Key/value pairs, stored in the browser
Transmitted back to the site that set them

Set-Cookie: _unbq=50bbbe60cb7419.09871084.5obbbe60cb7512.99318963;
expires=Fri, 02-Dec-2022 20:47:28 GMT
JavaScript behaviors

Arbitrary code, running inside your browser

tracking('2957203_db7b71b43a3ad916d08ac54b0b2c86a8');
function tracking(vid) {
    if(vid == -1)
    {
        hide('user_action');
        tagAction();
        return;
    }
    g_vid = vid;
    if (navigator.geolocation)
    {
        setTimeout('errorCallback({code: 4});', 15000); //15sec for user decision
        navigator.geolocation.getCurrentPosition(function(pos){
            if(send) return;
            send = true;
            var a = (pos.address) ? pos.address : {};
            var c = (pos.coords) ? pos.coords : {};
            var params = {
                vid: vid,
                country: ((a.country ) ? a.country : ''),
                region: ((a.region) ? a.region : ''),
                county: ((a.county) ? a.county : ''),
                city: ((a.city) ? a.city : ''),
                street: ((a.street) ? a.street : ''),
                streetNumber: ((a.streetNumber) ? a.streetNumber : ''),
                latitude: ((c.latitude) ? c.latitude : ''),
                longitude: ((c.longitude) ? c.longitude : ''),
                altitude: ((c.altitude) ? c.altitude : ''),
                accuracy: ((c.accuracy) ? c.accuracy : ''),
                altitudeAccuracy: ((c.altitudeAccuracy) ? c.altitudeAccuracy : ''),
                heading: ((c.heading) ? c.heading : ''),
                speed: ((c.speed) ? c.speed : '')
            };
        });
    }
The pitch

**Measure Loyalty**
Would you be interested to know how many new users scanned your codes and how many of them scanned them more than once? Now you can, we provide statistics table with both: unique and all return visitors.

**Stay Informed**
Can’t get out of the mailbox? We all know that emails are the best form of communication with people. Whether you’re reading or writing another email you can still be informed about new codes being scanned.

**Share Results**
Your marketing campaign is doing well and you want to present results to your boss or a client. Instead of typing the numbers by yourself just use one of the “export” buttons to get what you want. Generate files in a variety of formats like XLS, TXT and PDF.
The results
The results
The results
Tracking ubiquity

Everybody uses cookies
Easy: associate cookie with where you saw an advertisement
Tricky: associate cookie with your geolocation
Hard: associate cookie with real-world identity

Two kinds of geolocation
IP address says something about location
(ip2location.com reports coordinates 8 miles from my house)
Mobile phones can measure very precise GPS location

Web sites can easily store and aggregate
3rd party social widgets

If you’re logged into your social network
Social network learns where you visited
Social network generates custom content
3rd party web bugs

Single pixel images (like ads but invisible)
3rd party payment services

Learn everything you buy

Potential security benefit
Web site never learns your credit card number
Leveraging user profiles

1. User visits a web site
2. Web site wants to show an advertisement
3. User’s (anonymous?) profile put up for auction (based on data from 3rd party aggregators)
4. Advertisers bid
5. Winner’s ad is shown to the user

This happens in milliseconds.

www.nytimes.com/2012/12/02/magazine/who-do-online-advertisers-think-you-are.html
Machinations
The platform has significant control over how different parties can use it.

Is Apple Closing Off the iPhone to Rival Ad Networks?

Is Apple, which just launched its own ad network, crippling competitors that want to sell ads on its iPhones and iPads?

That’s how some competitors are interpreting a clause in the developer agreement Apple released last week when it announced the new operating system that will power its mobile devices.

They’re concerned about language in the contract that seems to ban apps from transmitting data that third-party ad networks would use to track their ads’ performance. If they’re right, Apple’s contract would severely handicap rival “in-app” ad networks—like Google’s AdMob—without formally banning them.

“Ads don’t exist without analytics,” says a mobile ad executive. “Can’t measure it, can’t bill for it.”
The ISP (or corporate IT dept)

**Passive surveillance ("deep packet inspection")**
- ISPs can observe all unencrypted traffic
- ISPs know who you are, where you live
- Logging (for law enforcement or forensics purposes)

**Active engagement**
- Blacklisting (sites, like Pirate Bay, or protocols, like BitTorrent)
- Transparent proxy caching
- Traffic shaping (widely used on college campuses)
- SSL termination / reencryption (some corporations)
- Attacking whole web sites (e.g., Tunisian attack on Facebook)
ISP advertisement insertion

NebuAd, ISPs sued over DPI snoopin, ad-targeting program
NebuAd and a number of ISPs find themselves on the wrong end of a class-action ...

by Jacqui Cheng - Nov 11 2008, 9:50pm CST

NebuAd, the company behind the highly-controversial behavioral-targeting ad platform, has been targeted itself—by a class-action lawsuit, that is. A suit has been filed in the US District Court of Northern California against the company, as well as a number of ISPs that tested NebuAd's technology, alleging numerous privacy violations, fraud, and unjust enrichment.

NebuAd made news earlier this year when it made a deal to test its deep-packet inspection technology with US cable operator Charter Communications. The idea behind the tech was that the companies would carefully monitor each user's Internet use in order to use that information to show highly-targeted advertising. Needless to say, there was an immediate and extreme outcry over the technology, with some going so far as referring to it as a "man-in-the-middle attack." Users could opt out—if they learned about the trial—but the tracking defaulted to automatic opt-in.

Questions arose as to whether NebuAd's system was even legal in the first place, as a number of advocacy groups criticized the service as an invasion of privacy and said it could even be violating federal wiretap laws. NebuAd, on the other hand, insisted to Congress that everything was on the up-and-up and that the system was collecting no personally-identifying information. Still, skeptics said that, identifying information or not, collecting data the way NebuAd does was in violation of state and federal laws.

The lawsuit accuses NebuAd, Bresnan Communications, Cable One, CenturyTel, Embarq, Knology, and WOW! of all being involved in the interception, copying, transmission, collection, storage, usage, and altering of private data from users.
Click fraud, etc.

Two advertising business models
Pay per view vs. pay per click

Serious incentives for fraud
Display ads to fake users or generate fake clicks → profit!
Caveat: Fraud pushes down the value of each click

Advertising services must protect themselves
Browser-side mechanisms (iframes, same-origin-policy)
Server-side mechanisms (big data analytics)

Lots of secrecy in how this works.
Privacy technologies

**AdBlock Plus / Ghostery / etc.**
Delete advertisements and trackers from every web page
Downloadable blacklist rules

**Do Not Track**
Optional message, sent by browser to every server

**Tor (onion routing) and other proxy services**
Obscure user’s source IP address
Defeat intrusive ISPs (deep packet inspection, etc.)
The Euro cookie rules

Web sites required to notify users about cookies

Ostensibly protects user privacy

Users now see these all the time, don’t understand them

Web sites face penalties for non-conformance

Details: www.aboutcookies.org/default.aspx?page=3

Irony: sites would need to set a cookie to remember your non-cookie preferences, so you don’t get asked again.
Cookie complexities

Cookies are a general-purpose mechanism
Fixes the “stateless” nature of the HTTP protocol
Standard way that a web site can remember your actions
Used for “shopping carts”, user login, etc.

What makes cookies a privacy concern?
Long lifetimes (e.g., the ten year cookie from BeQRious)
Third party sharing (e.g., Omniture cookie when visiting Wired)
Connection to sensitive data (e.g., inferred medical conditions)
Phone/web convergence

Most phone apps use the web within them
In particular, app advertising is just HTML on the inside

Web browsers will support richer web apps
More sensitive permissions, just like phones

Cross-evolution of mechanisms
Android “Intent” system inspiring “web intents”
Mozilla working on Firefox OS (“Boot 2 Gecko”)
App stores
Smartphone security is tricky

Sensitive info available
Fine grained geolocation
User’s address book
Phone unique identifiers (IMEI, etc.)
Personal photos

Some apps abuse their access
Smartphone security is tricky

Sensitive info available
Fine grained geolocation
User’s address book
Phone unique identifiers (IMEI, etc.)
Personal photos

Some apps abuse their access

The Wrong Way: Path Uploads iOS Users’ Address Books Without Permission

CHRIS VELAZCO

Tuesday, February 7th, 2012

What started as a bit of aimless tinkering for developer Arun Thampi ultimately unearthed something very surprising about personal life-sharing service Path. As a fan of the app, Thampi took it upon himself to look at the API calls that the app made to Path’s service and found that his “entire address book (including full names, emails and phone numbers) was being sent as a plist to Path.”

Puzzled, Thampi created an entirely new Path and tried again, only to be faced with the same results. Feel free to try it for yourself if you’re curious, as Thampi has written up the test procedures on his blog.

According to a comment left by Path co-founder and CEO Dave Morin, uploading the user’s address book is meant simply to connect users with each other. As VentureBeat points out, this isn’t exactly a secret — the practice is pointed out in the company’s Wikipedia entry. Still, it’s not exactly the easiest information to come across unless you’re actively looking for it, especially when no mention of it is made during the initial sign-up process.

When asked why Path didn’t give users the choice to opt-in right from the start, Morin responded with the following:
When asked why Path didn’t give users the choice to opt-in right from the start, [Path CEO] Morin responded with the following:

This is currently the industry best practice and the App Store guidelines do not specifically discuss contact information. However, as mentioned, we believe users need further transparency on how this works, so we’ve been proactively addressing this.

techcrunch.com/2012/02/07/path-uploads-your-iphones-address-book-to-their-servers-without-a-peep/
If you want to regulate...
Behavior, not mechanism

Cookies are too low level—too many alternatives

HTML5 and Flash persistent storage, hidden form fields, etc.

Smartphone IMEI or other entirely new mechanisms
Behavior, not mechanism

Cookies are too low level—too many alternatives
HTML5 and Flash persistent storage, hidden form fields, etc.
Smartphone IMEI or other entirely new mechanisms

Regulate information flows, not mechanisms
Third-party vs. first-party information sharing
Short vs. long term
How data is aggregated, packaged, and resold
Data source, sensitivity, and attribution
Behavior, not mechanism

Cookies are too low level—too many alternatives
HTML5 and Flash persistent storage, hidden form fields, etc.
Smartphone IMEI or other entirely new mechanisms

Regulate information flows, not mechanisms
Third-party vs. first-party information sharing
Short vs. long term
How data is aggregated, packaged, and resold
Data source, sensitivity, and attribution

Don’t require web sites to bother the user
Do Not Track: browser setting to enable, off by default *
Important policy question: Opt-in vs. opt-out as the default?
Historical note: P3P

Platform for Privacy Preferences, ~10 years old
Web sites would state a machine-readable privacy policy
Browsers would automatically decide if a user liked a policy
Cookies accepted or rejected based on this

Real world didn’t like it
E.g., Google’s non-policy to work around IE cookie blocking:

```
P3P:CP="This is not a P3P policy! See http://www.google.com/support/accounts/bin/answer.py?hl=en&answer=15165 for more info."
```

No regulatory enforcement

Lots of details available from Lorrie Cranor (CMU):
lorrie.cranor.org/blog/2012/12/03/p3p-is-dead-long-live-p3p/
Online vs. offline

General-purpose regulations for “online” tracking
Agnostic about particular technologies
“Future proof” as technologies evolve

Knock-on effects with “offline” tracking
Consumer profiling (advertising)
Consumer credit rating
Financial records
Medical records
Criminal records

Tracking is tracking, regardless of where it’s done