Could Googling Take Down A President, a Prime Minister, or an Average Citizen?

Gregory Conti
United States Military Academy
West Point, New York
gregory.conti@usma.edu
Outline

• Information Disclosure
  – Computing Platform
  – Network Eavesdropping
  – Destination Websites

• Vectors

• Cross-site Tracking
  – Advertising and Embedded Content

• User Motivation
  – The survey says

• Solutions
“Never talk when you can nod, and never nod when you can wink, and never write an e-mail because it's death. You're giving prosecutors all the evidence we need.”

- Eliot Spitzer
Two Years before his resignation

Eliot Spitzer
Former-Governor of New York
Maf54 (7:43:27 PM): well dont ruin my mental picture

Xxxxxxxxxx (7:43:32 PM): oh lol...sorry

Maf54 (7:43:54 PM): nice

Maf54 (7:43:54 PM): youll be way hot then

Xxxxxxxxxx (7:44:01 PM): haha...hopefully
Can anyone help me please! This stalking thing is not funny at all. When I type my name in keyword it gives a list of places that show where I have been on aol on the net. This is nobodys business. I have not done anything wrong at all and I have contacted aol about this matter and they keep saying they will do something about it but never do.

-Debbie

How do I get stuff removed from aol stalker? Can anyone tell me? Aol won't respond even though they claim willingness to remove data when requested. Someone, anyone, please help!

-Sally

IT: Details of Cyber Storm War Games Released

Posted by Soulskill on Thursday January 31, @07:29PM
from the defending-against-gravitic-mines dept.

I Don't Believe in Imaginary Property writes

"Apparently, the participants in the U.S. 'Cyber Storm' war games are familiar with the Kobayashi Maru, because some of them tried to cheat by hacking the games themselves. They also prepare for some very interesting scenarios. Among other things, the organizers are worried about having too many people on the 'No Fly' list show up at an airport, finding 'mystery liquids' in the subway, and having bloggers reveal the classified location of railcars with hazardous materials. The Department of Homeland Security has already analyzed the results of the games, and plans to hold 'Cyber Storm 2' in March."

IT: Microsoft Launches IT Superhero Comic
AOL Demo

- User #10291
- User #2708
AOL is embarking on a new direction for its business making its content and products freely available to all consumers. To support those goals, AOL is also embracing the vision of an open research community. To get started, we invite you to visit us at http://research.aol.com, where you will find:

- 20,000 hand labeled, classified queries
- 3.5 million web question/answer queries (who, what, where, when, etc.)
- Query streams for 500,000 users over 3 months (20 million queries)
- 2 million queries against US Government domains

Also, please feel free to provide feedback on the site, datasets you'd like to see in the future, and any other comments about our vision.
Definitions

• googling: The full spectrum of free online tools and services (such as search, mapping, email, Web-based word processing and calendaring etc.)

• web-based information disclosure: the information we disclose as we surf the web
Global Computing Statistics

- World Population: ~6.6 Billion
- Cell Phones: ~3.3 Billion
- Personal Computers: ~1.2 Billion
- MP3 Players: ~220 Million
- Digital Cameras: ~120 Million
- Webcams: ~100 Million
- PDAs: ~85 Million
- DVRs: ~44 Million
- Servers: ~27 Million

Data Collection

Number of times data is collected on each visitor in a Month (Average)

- Yahoo: 2500
- MySpace: 1100
- AOL: 550
- Google: 480
- Facebook: 350
- Microsoft: 200
- Ebay: 100
- Amazon: 50

http://www.nytimes.com/2008/03/10/technology/10privacy.html?pagewanted=1&_r=1&hp / Comscore
Unique Visitors
(millions / month)

Yahoo: 160 million
MySpace: 80 million
AOL: 160 million
Google: 150 million
Facebook: 30 million
Microsoft: 140 million
Ebay: 80 million
Amazon: 70 million

http://www.nytimes.com/2008/03/10/technology/10privacy.html?pagewanted=1&_r=1&hp & Comscore
• 1.319 Billion for Year End 2007
• 20% Overall Penetration

source: http://www.internetworldstats.com/stats.htm
Information Leakage on a PC

- Phone Line: Modern Fax, Audio over Telephone
- EM Emissions: Wireless Networks, Spurious EM Emissions from electronic components and cables, Light
- Sound: Keystrokes, Speaker Output, Spoken Voice
- Things: USB Key rings, CDs / DVDs, Printed Output
- Power Lines
- Wired Network
- People
- Peripherals
Information Leakage and Spurious Emanations on a Network
ISPs vs. Large Online Companies

Online Company
- Sees global traffic from many customers
  - domain specific
- Advertising and embedded content brings in additional information
- Limited knowledge of user identity
- Extensive datamining

ISP
- Sees all traffic from its set of customers
  - except encrypted traffic
  - traffic analysis
- Limited to no visibility on non-customers
- Knows identity and location of accounts
- Ability to manipulate network flows
  - DNS
  - blocking P2P
Vantage Points of ISPs vs Online Companies
DNS Based Vulnerabilities
Major Threat Vectors

- Email
- Search
- Mobile Phones
- FTP Replacement
- Web office suites
- Mapping
- ...

# The Many Flavors of Search

(Simply Google)

<table>
<thead>
<tr>
<th>Sites</th>
<th>Searches</th>
<th>Blogs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounts</td>
<td>Web</td>
<td>Search</td>
</tr>
<tr>
<td>AdSense</td>
<td>Images</td>
<td>Search</td>
</tr>
<tr>
<td>Adswords</td>
<td>Groups</td>
<td>Search</td>
</tr>
<tr>
<td>Alerts</td>
<td>News</td>
<td>Search</td>
</tr>
<tr>
<td>Analytics</td>
<td>Products</td>
<td>Search</td>
</tr>
<tr>
<td>Anatomy</td>
<td>Blogs</td>
<td>Search</td>
</tr>
<tr>
<td>Authors</td>
<td>Books</td>
<td>Search</td>
</tr>
<tr>
<td>Mobile Books</td>
<td>Maps</td>
<td>Search</td>
</tr>
<tr>
<td>Calendar</td>
<td>Scholar</td>
<td>Search</td>
</tr>
<tr>
<td>Checkout</td>
<td>Video</td>
<td>Search</td>
</tr>
<tr>
<td>Co-op</td>
<td>Code</td>
<td>Search</td>
</tr>
<tr>
<td>Code for Educators</td>
<td>Music</td>
<td>Search</td>
</tr>
<tr>
<td>Custom Search</td>
<td>Reader</td>
<td>Search</td>
</tr>
<tr>
<td>Current TV</td>
<td>Base</td>
<td>Search</td>
</tr>
<tr>
<td>Docs</td>
<td>Directory</td>
<td>Search</td>
</tr>
<tr>
<td>Earth Enterprise</td>
<td>Finance</td>
<td>Search</td>
</tr>
<tr>
<td>Earth Outreach</td>
<td>Catalogs</td>
<td>Search</td>
</tr>
<tr>
<td>Executives</td>
<td>Answers</td>
<td>Search</td>
</tr>
<tr>
<td>Foundation</td>
<td>History</td>
<td>Search</td>
</tr>
<tr>
<td>Gadget Ventures</td>
<td>History</td>
<td>Search</td>
</tr>
<tr>
<td>Gadgets</td>
<td>Bookmarks</td>
<td>Search</td>
</tr>
<tr>
<td>Grants</td>
<td>Linux</td>
<td>Search</td>
</tr>
<tr>
<td>Hot Trends</td>
<td>Apple</td>
<td>Search</td>
</tr>
<tr>
<td>iGoogle</td>
<td>Microsoft</td>
<td>Search</td>
</tr>
<tr>
<td>Investor Relations</td>
<td>Trends</td>
<td>Search</td>
</tr>
<tr>
<td>Help</td>
<td>Notebooks</td>
<td>Search</td>
</tr>
<tr>
<td>Jobs</td>
<td>Accessible</td>
<td>Search</td>
</tr>
<tr>
<td>JobsSpot</td>
<td>Code</td>
<td>Search</td>
</tr>
<tr>
<td>Mail</td>
<td>Maps</td>
<td>Search</td>
</tr>
<tr>
<td>Maps</td>
<td>Patents</td>
<td>Search</td>
</tr>
<tr>
<td>Mashup Editor</td>
<td>View Timeline</td>
<td>Search</td>
</tr>
<tr>
<td>Moon</td>
<td>View Map</td>
<td>Search</td>
</tr>
<tr>
<td>News Archive</td>
<td>View Map</td>
<td>Search</td>
</tr>
<tr>
<td>News Suggest</td>
<td>Move</td>
<td>Search</td>
</tr>
<tr>
<td>OpenSocial</td>
<td>Definitions</td>
<td>Search</td>
</tr>
<tr>
<td>Papers</td>
<td>Weather</td>
<td>Search</td>
</tr>
<tr>
<td>Press Centre</td>
<td>BBC</td>
<td>Search</td>
</tr>
<tr>
<td>Privacy</td>
<td>BBC</td>
<td>Search</td>
</tr>
<tr>
<td>The Answer</td>
<td>Feeling Lucky</td>
<td>Search</td>
</tr>
<tr>
<td>Research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SearchMash</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ads</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gadgets, Google, and SEO</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Malware Search
http://metasploit.com/research/misc/mwsearch/index.html

http://www.theregister.co.uk/2006/07/18/malware_search/
Motivation

• Cost benefit analysis
  – users
  – webmasters
  – bloggers

• Short-term gain vs. long term risks

• stats

• boils down to trust and awareness
Simply Google - November 2007

Want More Hits To Your Website?  
Get Guaranteed Traffic In 48 Hrs!  

www.AdvertisingWebService.com
If the content on the web it is fair game.
Case Studies
Craig’s List

personal

strictly platonic
women seek women
women seeking men
men seeking women
men seeking men
misc romance

casual encounters
missed connections
rants and raves
Feedback Profile

**salesman_jack** (66 ★)
Member since Dec 03/07 in Hong Kong

<table>
<thead>
<tr>
<th>Feedback Score</th>
<th>66</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Feedback</td>
<td>99.5%</td>
</tr>
<tr>
<td>Members who left a positive feedback</td>
<td>67</td>
</tr>
<tr>
<td>Members who left a negative feedback</td>
<td>1</td>
</tr>
<tr>
<td>All positive feedback</td>
<td>310</td>
</tr>
</tbody>
</table>

**Recent Feedback Ratings**

<table>
<thead>
<tr>
<th>Time Period</th>
<th>1 month</th>
<th>6 months</th>
<th>12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>310</td>
<td>310</td>
<td>310</td>
</tr>
<tr>
<td>Neutral</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Negative</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Detailed Seller Ratings** (since May 2007)

<table>
<thead>
<tr>
<th>Item as described</th>
<th>Average rating</th>
<th>Number of ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>★★★★★</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Communications</td>
<td>★★★★★</td>
<td></td>
</tr>
<tr>
<td>Shipping time</td>
<td>★★★★★</td>
<td>52</td>
</tr>
<tr>
<td>Shipping and handling charges</td>
<td>★★★★★</td>
<td>52</td>
</tr>
</tbody>
</table>

Feedback as a seller

- 311 Feedback received
- Ratings mutually withdrawn: 1

- **Feedback**
  - breastcancercharm ring keychain (pink) NIB (#20316669362)
  - Item as described
  - Received on 09/23/08
  - US $1.84
  - View Item

- **Thank you great transaction!!**
  - Red deg charms ring charm keychain NIB (#20315996326)
  - Item as described
  - Received on 09/11/08
  - US $1.84
  - View Item

- **Great transaction, Thanks**
  - Breast cancer awareness charm ring keychain (pink) NIB (#20315974062)
  - Item as described
  - Received on 09/29/08
  - US $1.84
  - View Item

- **Excellent price. Product exactly what was described. Fast shipping considering**
  - Pink Heart Heart Crystal Heart Charm Keychain New w/Box (#20316696233)
  - Item as described
  - Received on 09/22/08
  - US $2.00
  - View Item

- **Beautiful purchase. My daughter loves it. Perfect condition. Thank you**
  - Item as described
  - Received on 09/21/08
  - View Item
Customers with Similar Searches Purchased

- **Down and Out in the Magic Kingdom** Paperback by Cory Doctorow
- **Overclocked: Stories of the Future...** Paperback by Cory Doctorow
- **Little Brother** Hardcover by Cory Doctorow

> View or edit your browsing history

Recommended Based on Your Browsing History

- **Secure Programming with Static Analysis** Paperback by Brian Chess, Jacob West
- **Network Warrior** Paperback by Gary Donahue
- **Linux Firewalls: Attack Detection and...** Paperback by Michael Rash

> View or edit your browsing history
Any email that touches any of these servers should be considered compromised.
Linked In

Social networking sites know your contacts and your contacts’ contacts. Old friends will find you and let the site know of the relationship.
Mapping sites reveal locations of interest, allowing diverse groups of users to be linked.
Even the most innocent appearing services should be considered as collecting your data.
Cross-site Tracking

- Web Analytics
- Embedded content
  - Webbugs
  - YouTube Videos
- Advertising
  - Doubleclick
  - AdSense
- Referer data
- Javascript Libraries
- Clickthrough Tracking
A Visit to MSNBC
- a365.ms.akamai.net
- a509.cd.akamai.net
- ad.3ad.doubleclick.net
- amch.questionmarket.com
- c.live.com.nsatc.net
- c.msn.com.nsatc.net
- rad.msn.com.nsatc.net
- context3.kanoodle.com
- global.msads.net.c.footprint.net
- hm.sc.msn.com.c.footprint.net
- msnbcom.112.2o7.net
- prpx.service.mirror-image.net
- wrpx.service.mirror-image.net
- switch.atdmt.com
- view.atdmt.com
- www-google-analytics.l.google.com

- 16 third-party sites
- 10 separate companies
Third Party Content and Web Bugs

Browser

Request and Receive Web Page

Request and Receive Image 1

Request and Receive Image 2

Web Server

Log
1 log entry
Linking a User Across Many Platforms

- **Home PC**: Same IP address due to NAT, different cookies indicate different machines.
- **Internet Cafe**: Same registered user account.
- **Work PC**: Personal laptop used in cafeteria on corporate IP space.
- **Internet Cafe**: No connection, due to minimal activity and no use of registered user account.
Linking Users, Groups, and Organizations

- Forwarded a set of directions from Google Maps
- Came from the same IP address
- Searching for the same phone number
- Third-party searched for both their names in the same query
- Collaboratively edited the same word processing document
- Read the same blog
- Shared the same cookie, but came from different IP addresses
- Read the same types of news stories
- Shared the same corporate IP address block
- Shared the same rare Google Alerts
- Googlebot found both of their email addresses on the same page
- Visited the same set of Google AdSense advertisers
- Emailed user's Gmail account
- Used cell phone-based search from the same location
Advertising – The DoubleClick Model
Click Through Tracking

Google Search Result Link Clicked

Host: www.google.com
Cookie: PREF=ID=9943d78763e5e18e:TM=1199411995:LM=1199411995:S=e5iiORgriiHNRjTP

No Additional User Action

GET / HTTP/1.1
Host: creativecommons.org
Referer: http://www.google.com/search?hl=en&q=creativecommons&btnG=Search
Browser Supplied Information

<table>
<thead>
<tr>
<th>Variable</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>appName</td>
<td>Netscape</td>
</tr>
<tr>
<td>appCodeName</td>
<td>Mozilla</td>
</tr>
<tr>
<td>appVersion</td>
<td>5.0 (Windows; en-US)</td>
</tr>
<tr>
<td>appMinorVersion</td>
<td>Property not supported or blank!</td>
</tr>
<tr>
<td>IE real version</td>
<td>Not Internet Explorer or version is undetectable!</td>
</tr>
<tr>
<td>UserAgent</td>
<td>Mozilla/5.0 (Windows U; Windows NT 5.1; en-US; rv:1.8.1.10) Gecko/20071115 Firefox/2.0.0.10</td>
</tr>
<tr>
<td>Security Policy</td>
<td>Property not supported or blank!</td>
</tr>
<tr>
<td>Browser Name</td>
<td>Mozilla Firefox 2.0.0.10</td>
</tr>
<tr>
<td>Operating System</td>
<td>Windows NT 5.1 - Windows XP</td>
</tr>
<tr>
<td>Platform</td>
<td>Win32 - Windows 32-bit platform</td>
</tr>
<tr>
<td>CPU</td>
<td>Property not supported or blank!</td>
</tr>
<tr>
<td>Online</td>
<td>true - The system is in global online mode</td>
</tr>
<tr>
<td>Connection Type</td>
<td>Property not supported or blank!</td>
</tr>
<tr>
<td>Vendor</td>
<td>Property not supported or blank!</td>
</tr>
</tbody>
</table>
AJAX Libraries API

- jQuery
- prototype
- script.aculo.us
- MooTools
- dojo
- cached

<script src="http://www.google.com/jsapi"></script>
<script>
  google.load("jquery", "1"); // Load jQuery
  google.setOnLoadCallback(function() {
    $.getJSON("http://ajax.googleapis.com/ajax/services/search/web?q=google&v=1.0&callback=?", function (data) {
      if (data.responseDate.results && data.responseDate.results.length>0) {
        renderResults(data.responseDate.results);
      }
    });
  });
</script>

http://code.google.com/apis/ajaxlibs/
Ebay pulling ads from a Yahoo server
Profiling

- Google hackers
- Security researchers
- Political activists
- Company XXX employee
- Corporate leaders
- Law enforcement officer
- Government official

Data Mining

DANGERS

Indexed by the number of Google results for
"Died in a ___ accident"

<table>
<thead>
<tr>
<th>Type of Accident</th>
<th>Google Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skydiving</td>
<td>710</td>
</tr>
<tr>
<td>Elevator</td>
<td>575</td>
</tr>
<tr>
<td>Surfing</td>
<td>496</td>
</tr>
<tr>
<td>Skateboarding</td>
<td>473</td>
</tr>
<tr>
<td>Camping</td>
<td>166</td>
</tr>
<tr>
<td>Gardening</td>
<td>100</td>
</tr>
<tr>
<td>Ice Skating</td>
<td>94</td>
</tr>
<tr>
<td>Knitting</td>
<td>7</td>
</tr>
<tr>
<td>Blogging</td>
<td>2</td>
</tr>
</tbody>
</table>

http://xkcd.com/369/
Countermeasures

• Patching Users
  – Raised Awareness
  – Know What You are Disclosing
  – Usable Security
  – ...

• Technical Countermeasures
  – Cookie Managers
  – Content Filtering
  – Self-monitoring
  – Search Term Chaffing
  – Encryption
  – Anonymizing Proxies
  – Tor
  – NAT Firewalls
  – ...

• Policy Countermeasure
  – Petition Law and Policy Makers
  – Support EFF and other Privacy Organizations
  – ...

Your IP is identified to be a Tor-EXIT. So you are using Tor successfully to reach the web!

**CHECK RESULT**

- **is a Tor-EXIT IP**
- **(color shows reachability of [dgraphxper@xenobite.com])**
- **ENABLED**
- **(disabled)**
- **possibly disabled**
- **http://torcheck.xenobite.eu/?
  presenttime=1213715403333
  useragent=Mozilla/5.0 (Windows NT 5.1; en-US; rv:1.8.1.9) Gecko/20070725
  Firefox/1.0.0.9
  LAUNCHER: en-us,en-gb,fr
  BROWSER: gzip deflate
  REFERRER: iso-9653p-1,usr-z,q=0.7,"","p=0.7"
Tor Button
Chained Proxies
Threat Spectrum

Cross-site tracking
User profiling
Targeted advertising

Likely

DNS Redirection
Third-party sharing
Government collaboration

Service eliminated
Search result ranking manipulation
User fingerprinting
Data Spills

Redirect to malicious sites
Less Likely
Future Countermeasures

• Plug-in that monitors information disclosures

• Plug-in that displays third-party contacts – at least serve as a history function for the status bar
Conclusions

• what online companies posses vs. what they share are two entirely different things
  – blurring streetview
  – blurring google maps
  – internal search logs vs. public interface
Acknowledgements

- 3efd09cddc148ee790d17e35ae323852
- Kulsoom Abdullah
- Ed Sobiesk
- New Security Paradigms Workshop
- Symposium on Usable Privacy and Security
Questions?