



What do your searches reveal about you?

By Kelechukwu (Kc) Udonsi and ZiXian (Wannie) Chen



Focus questions:

- Just how much does Google know about each of its users? What information is collected and how, how long is it kept, and how is it protected? What about other search-engines, such as Yahoo! and Bing?
- How much can you learn about a person just from their search history? Can a seemingly anonymous person be identified from their search history?









Google

- Privacy policy states that by just signing in, Google can collect information about you.



What information does Google collect?¹

-  Search history
-  Clicks on search results
-  Browse history
-  Geo location
-  Application history²
-  Even keeping a history of the ads that you have watched

1. (Pingdom, 2010)

2. Google dashboard/activities page



How?

- Google searches.
- Information aggregated from all Google platforms such as YouTube, etc.



Outside of Google

- Pretty much all web applications are integrated with Google.
- Examples of other ways Google might track you:
 - Google analytics
 - Google+ button
 - Google maps/GPS





How much does Google know about you? ¹

- Every bit of the information collected is stored.
- All of this is viewable in [Google Dashboard](#) and [Google Activities](#).

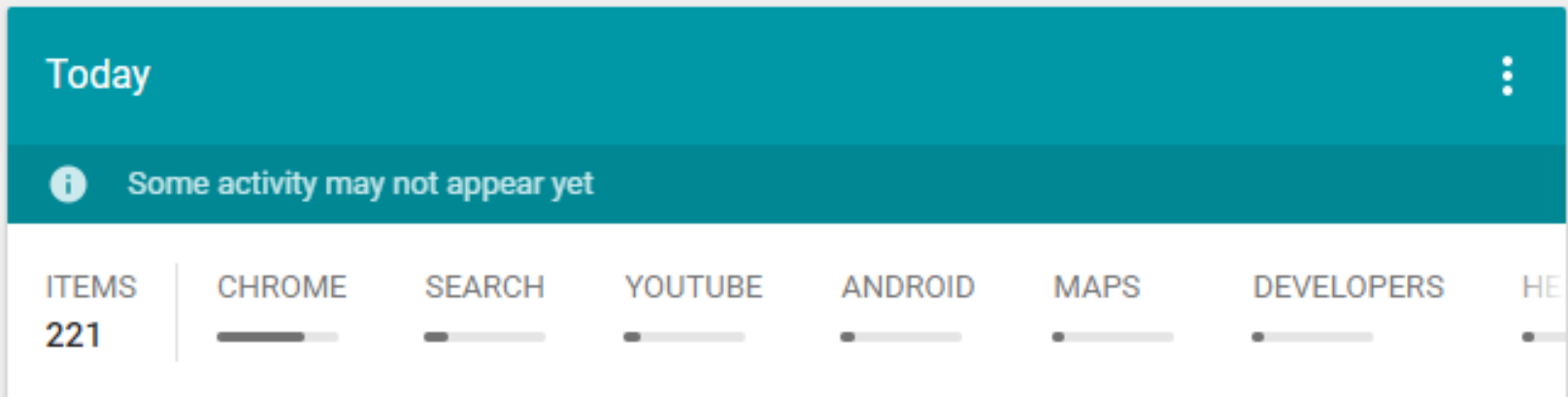


Google Dashboard

- Information from different google applications such as Gmail, Google Docs.

Google Activities

- A database of all your search & browse history.
- Information must be manually deleted.








Other search engines



Yahoo ¹

-  Personal information & cross-platform tracking.
-  Customized ads and personalized searches
-  “Provide anonymous reporting for internal and external clients.”

1. (Carpentier, 2013) □



Other search engines

- Bing ¹
 - Cookie tracking across Microsoft applications.
 - Search terms along with IP address, and geo-location
 - Device information



Other search engines

- Baidu
 - Chinese search engine
 - Collects GPS data, names of nearby wireless networks and device information. ¹
- Improvement of services and research ²

1. (VanderKlippe, 2017) □
2. Baidu research privacy policy



Other search engines

- Collect the same information
 - But much more limited
- 



How secure is this information?

- Study done in 2016, found multiple vulnerabilities in these search engine platforms. ¹
 - Using cookie hijacking.
- Important information or even functions compromised.

1. (Sivakorn, et al, 2016)

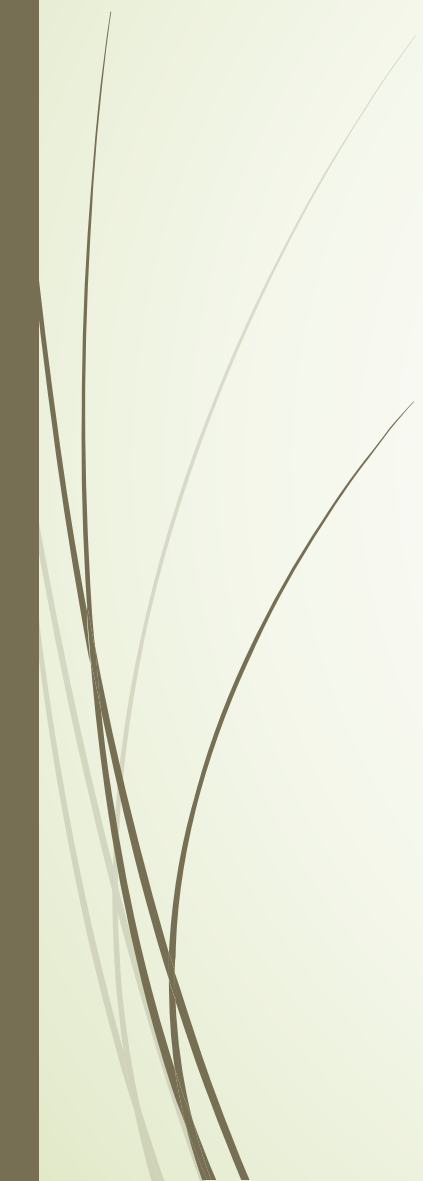


Yahoo, Bing and Baidu

- Yahoo:
 - Allowed access to your contact list and send emails.
- Bing & Baidu:
 - Allowed access to your entire search history.
- Little security is put into the privacy of your information.




Even Google failed to protect your privacy

- Same researched found a similar exploit through Google.
 - Allowed attackers access to your location.
- 



How much can you learn about a person just from their search history?




Just about anything. Their ethnicity, food habits, music tastes, health issues, names, names of friends or acquaintances




How much can you learn about a person just from their search history?

- Search providers don't think much can be learned from a person's search history. They store the search history with an index. Each user is assigned an Anonymity ID.



How much can you learn about a person just from their search history? Cont.

- Search engine companies argue they do not store names and addresses with search queries ¹



How much can you learn about a person just from their search history? Cont.

- Search queries reveal habit, life style, names, addresses, interests. All of which when gathered in sufficient amount can be analyzed to reveal identity



How much can you learn about a person just from their search history? Cont.

- ▶ AOL dumps search data for researchers. Outcry as AOL user is identified from released data ¹



Example analysis

- AOL query-log is about 2GB and available online
- 5MB will be used for an sample analysis
- This sample data is available at <http://www.cs.utoronto.ca/~gh/cscD03/AOL-data>.

Example Analysis Cont.

- Analyzing the user with anonymity id. 39509
 - Lives in Tahlequah, Oklahoma
 - Avid Myspace user
 - Youth pastor

39509	tahlequah public schools	2006-05-05 17:05:08	1	http://www.tahlequah.k12.ok.us
39509	oklahoma displined pastors	2006-03-04 19:17:36		
39509	oklahoma disciplined pastors	2006-03-04 19:17:44		
39509	oklahoma disciplined doctors	2006-03-04 19:18:18	3	http://www.medlaws.com
39509	games for church youth groups	2006-03-19 16:33:29	2	http://www.funandgames.org
39509	games for church youth groups	2006-03-19 16:33:29	10	http://www.thesource4ym.com
39509	games for church youth groups	2006-03-19 16:49:49		
39509	myspace layouts for guys	2006-03-18 22:16:38	1	http://myspace.nuclearcentury.com
39509	myspace layouts for guys	2006-03-18 22:16:38	3	http://www.freecodesource.com
39509	myspace layouts for guys	2006-03-18 22:16:38	4	http://www.pimpmycom.com

Example Analysis Cont.

- ▶ Analyzing the user with anonymity id. 39509
 - ▶ Loves cars, owns or wishes to own a Ford or Cummins

```
39509  ford engine head id      2006-03-09 19:26:56    1      http://raceabilene.com
39509  where are the casting numbers on ford cylinder heads 2006-03-09 19:44:00    1
http://www.classicmustang.com
39509  where are the casting numbers on ford cylinder heads 2006-03-09 19:44:00    1
http://www.classicmustang.com
39509  where are the casting numbers on ford cylinder heads 2006-03-09 19:44:00    1
http://www.classicmustang.com
39509  where are the casting numbers on ford cylinder heads 2006-03-09 19:44:00    3
http://www.mortec.com
39509  1961 ford fairlane        2006-03-09 20:05:04    3      http://www.musclegarclub.com
39509  1961 ford fairlane        2006-03-09 20:05:04    4      http://4wheeldrive.about.com
39509  1961 ford fairlane        2006-03-09 20:05:55    21     http://cgi.ebay.com
39509  engine torque specs for a 5.9 cummins 2006-03-09 12:53:31
39509  ask jeeves                2006-03-09 12:54:16    1      http://www.ask.com
39509  engine torque specs for a 5.9 cummins 2006-03-09 12:59:08
39509  engine torque specs for a 5.9 cummins 2006-03-09 12:59:58
39509  free chat                 2006-03-09 14:29:58    1      http://www.chat-avenue.com
```


Example Analysis Cont.


- Analyzing the user with anonymity id. 39509
 - Interesting sexual desires
 - Possibly pedophilic and drug user

39509	caught masturbating	2006-05-26 11:48:07	4	http://girlsmasturbating.ff4g.com
39509	caught masturbating	2006-05-26 11:48:07	6	http://hotgirlsmasturbating.ff4g.com
39509	caught masturbating	2006-05-26 11:48:07	9	http://masturbatinggirlfreevideo.ff4g.com
39509	dimethicone	2006-03-30 13:51:14	1	http://www.snowdriftfarm.com
39509	fentanyl	2006-03-30 13:52:24	1	http://www.streetdrugs.org
39509	side effects of fentanyl	2006-03-30 13:56:25	1	http://www.drugs.com
39509	misssing and expolited children	2006-03-05 21:55:35		
39509	pre teen porn pics	2006-03-26 14:16:54	3	http://xt32schoolgirl.proboards59.com
39509	pre teen porn pics	2006-03-26 14:16:54	7	http://www.mhhe.com
39509	48year old women sex pics	2006-03-26 14:22:10		
39509	48yearold women sex pics	2006-03-26 14:22:41	6	http://www.obeses.yi.org
39509	midget porn	2006-03-28 12:35:37		
39509	cruelart.pornoeye.com	2006-03-07 13:31:47		
39509	mandy moore	2006-05-14 14:23:04		
39509	mandy moore oops	2006-05-14 14:23:25		
39509	mandy moore naked	2006-05-14 14:24:02		

Example Analysis Cont.

- Analyzing the user with anonymity id. 39509
 - Likely ill or casually researching about Tourette and Parkinson's
 - Gym goer or fan

39509	parkinstons	2006-03-10 09:13:30			
39509	parkinson's	2006-03-10 09:14:02			
39509	adult tourettes	2006-05-03 20:17:09	4		http://www.vh.org
39509	adult tourettes	2006-05-03 20:17:09	3		http://www.tourettes-disorder.com
39509	early signs of tourettes	2006-05-03 20:34:28	1		http://www.ninds.nih.gov
39509	how does tourettes affect a childs speech	2006-05-03 20:36:20			
39509	tourettes in preschool children	2006-05-03 20:37:04	1		http://jaynagirl.cwd-cragin.com
39509	muscle man competition	2006-03-10 09:35:39	17		http://eastcoastmuscle.com
39509	muscle man competition	2006-03-10 09:35:39	18		http://www.highbeam.com
39509	muscle man competition	2006-03-10 09:42:25			



Can a seemingly anonymous person be identified from their search history?

- The answer to this question is neither a hard yes nor no and depends on a few factors.
- It's important to understand the meaning of 'seemingly anonymous'



Can a seemingly anonymous person be identified from their search history? Cont.

- If “seemingly anonymous” describes the regular internet user under the impression of anonymity then the answer is an obvious yes from reasons and examples earlier discussed.



Can a seemingly anonymous person be identified from their search history? Cont.

- ▶ If “seemingly anonymous” describes an internet privacy activist, then the answer will be No. The degree of difficulty identifying such a user from their search history would depend on steps taken to achieve anonymity

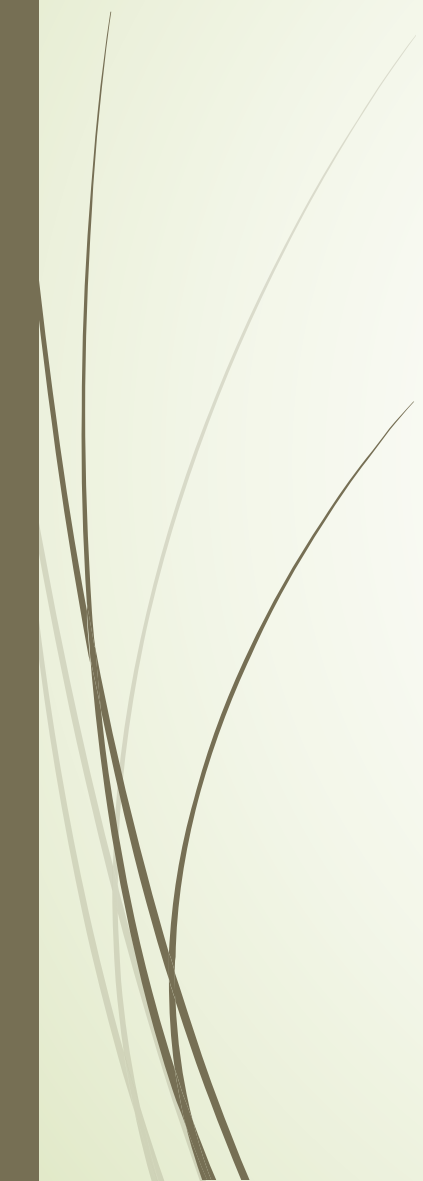


Can a seemingly anonymous person be identified from their search history? Cont.

- Some steps that can be taken to achieve anonymity include
 - Don't assume the cloak of anonymity
 - Use search engines wisely
 - Use TOR browser
 - Decline search engine cookies
 - Use VPN to change physical location



Discussion questions

- 
- Google and other search engines/platforms collect information in order to provide more personalized services such as ads, do you think they should be allowed to do this?
 - Could search query analysis return a false positive? That is, matching a wrong identity to a user.



Q

&

A



Thank you

References

- Hansell, S. (2006). *Marketers Trace Paths Users Leave on Internet*. [online] Nytimes.com. Available at: <http://www.nytimes.com/2006/08/15/technology/15search.html> [Accessed 21 Sep. 2017].
- Barbaro, M. and Zeller Jr., T. (2006). *A Face Is Exposed for AOL Searcher No. 4417749*. [online] Nytimes.com. Available at: <http://www.nytimes.com/2006/08/09/technology/09aol.html> [Accessed 21 Sep. 2017].
- Hafner, K. (2006). *Researchers Yearn to Use AOL Logs, but They Hesitate*. [online] Nytimes.com. Available at: <http://www.nytimes.com/2006/08/23/technology/23search.html> [Accessed 21 Sep. 2017].
- Sivakorn, S., Polakis, I., & Keromytis, A. D. (2016). The Cracked Cookie Jar: HTTP Cookie Hijacking and the Exposure of Private Information. *2016 IEEE Symposium on Security and Privacy (SP)*. doi:10.1109/sp.2016.49
- Pingdom. How Google collects data about you and the Internet. (2010, January 08). From <http://royal.pingdom.com/2010/01/08/how-google-collects-data-about-you-and-the-internet/>
- Carpentier, M. (2013, October 31). What Your Search History Says About You (And How to Shut It Up). From http://www.huffingtonpost.com/megan-carpentier/what-your-search-history-_b_4179728.html
- [VanderKlippe, N. \(2017, March 25\). China collecting sensitive personal data through Android apps: report. From https://beta.theglobeandmail.com/technology/tech-news/millions-of-android-apps-send-sensitive-data-to-china-u-of-t-report/article28865055/?ref=http%3A%2F%2Fwww.theglobeandmail.com&](https://beta.theglobeandmail.com/technology/tech-news/millions-of-android-apps-send-sensitive-data-to-china-u-of-t-report/article28865055/?ref=http%3A%2F%2Fwww.theglobeandmail.com&)
- G. Pass, A. Chowdhury, C. Torgeson, "A Picture of Search" The First International Conference on Scalable Information Systems, Hong Kong, June, 2006.
- Boatman, K. (n.d.). *Are Your Internet Searches Really Anonymous?*. [online] Us.norton.com. Available at: https://us.norton.com/yoursecurityresource/detail.jsp?aid=internet_searches [Accessed 24 Sep. 2017].
- Eckersley, P., Schoen, S., Bankston, K. and Slater, D. (2006). *Six Tips to Protect Your Search Privacy*. [online] Electronic Frontier Foundation. Available at: <https://www.eff.org/wp/six-tips-protect-your-search-privacy> [Accessed 24 Sep. 2017].