Privacy
CSE 120 Spring 2017

Instructor: Justin Hsia
Teaching Assistants:
Anupam Gupta, Braydon Hall, Eugene Oh, Savanna Yee
Administrivia

- Assignments:
  - Arrays and Elli due tonight (4/28)
  - Controlling Elli due Sunday (4/30)
  - Living Computers Museum Report (5/14)

- Midterm grades released on Gradescope by Sunday
  - Will have opportunity next week to submit regrade requests
  - You did great!
  - Why did we have a paper exam?
Outline

- Privacy
- Online Interactions
- Multimedia Retrieval
  - Geotagging
- Apps and Access
Being Online

- It’s the *World Wide Web*

- When we’re online, we typically only think about our *intended* audience
  - “Small” circle of friends and organizations

- Things to consider:
  - The Internet is a real, physical place
  - Digital information is “permanent” and easy to distribute
  - Who can access your stuff?
Privacy

- The right of people to *choose freely* what circumstances and to what extent they will reveal themselves, their attitude, and their behavior to others.

- Privacy is an *explicit* human right in many countries
  - Strong privacy protections mean that receivers of information must keep it private
  - Unfortunately, the US has almost no privacy protections
Privacy and Technology

- How did privacy work in the past?
  - Taking someone else’s stuff is illegal: theft, burglary, robbery, extortion
  - The 4th Amendment protects from search and seizure by the government

- Technology has changed the game!
  - It is now possible (and easy?) to violate people’s privacy without their knowledge
Privacy and Technology

- Former Supreme Court Justice Louis Brandeis:
  - “The narrower doctrine [of illegal search] may have satisfied the demands of society at a time when the abuse to be guarded against could rarely have arisen without violating a contract or a special confidence; but now that modern devices afford abundant opportunities for the perpetration of such wrongs without any participation of the injured party, the protection granted by the law must be placed upon a broader foundation. – “The Right to Privacy” (1890)

- An interesting conundrum at the intersection of technology and policy
  - e.g. Who is liable in today’s sharing economy? https://backchannel.com/the-most-important-law-in-tech-has-a-problem-64f5464128b6
Outline

- Privacy
- **Online Interactions**
  - Multimedia Retrieval
    - Geotagging
- Apps and Access
An Experiment

How's your tom muscle?
Audience Responses

- In what ways do you produce information (“digital footprints”) online?
Your Information

- Profile
  - Name, birthdate/age, job, photo, contact info, etc.

- Transactions
  - Timestamp, store, product, price, payment info

- Preferences
  - Clicks, advertisements, “likes”, browsing history

- Location
  - Geo-tags, location tags, GPS data

- Can these be connected to you?
  - Payment info, customer IDs, login accounts
Your Information

- **Profile**
  - Name, birthdate/age, job, photo, contact info, etc.

- **Transactions**
  - Timestamp, store, product, price, payment info

- **Preferences**
  - Clicks, advertisements, “likes”, browsing history

- **Location**
  - Geo-tags, location tags, GPS data

- **What happens to all of this information?**
  - Targeted advertising, sold for $, Big Data analysis
What Should Be Private?

- Birth Certificate demo
  - [See Panopto]

- Is this a violation of my privacy?
  - Actually no: birth certificates are public record

- Did this need to be online?
  - No... but it is and there’s nothing I can do about it

- Back to transactions:
  - Socks might not be a sensitive product to buy
  - Birth control might be a sensitive product to buy
Security Aside

“Security questions” usually ask for personal information that is specific to you:
- Mother’s maiden name
- Place of birth
- Street that you grew up on
- Model of first car
- Name of first pet
- Favorite sports team

How close would someone need to be in your social circles in order to know the answer to these questions?
Social Cause

- Increasingly, people use the Internet for regular activities
  - *e.g.* communication, shopping, information gathering
  - People *like* a highly-personalized web experience

- Industry is working to improve search and retrieval techniques
  - Improve the user experience (and their bottom lines)

- Governments improve search and retrieval to do forensics and intelligence gathering
Observations

- Internet sites and mobile apps encourage sharing of data too easily
  - Users often blindly follow along

- User *and* engineers are often unaware of search and retrieval possibilities of shared data

- Local privacy protections are often ineffective against inference across websites
Consequences: Cybercasing

- **Cybercasing**: Using online data and services to enable “real-world” crimes (that would not otherwise be possible)

Outline

- Privacy
- Online Interactions
- Multimedia Retrieval
  - Geotagging
- Apps and Access

Note: The work shown in this section is entirely that of Dr. Gerald Friedland, director of the Audio and Multimedia Group at the International Computer Science Institute (ICSI) in Berkeley, CA: https://www.icsi.berkeley.edu/icsi/projects/multimedia
Question

- On average, how often do you post images and/or videos on the Internet?
  - e.g. Facebook, Instagram, Craigslist, Youtube, Twitter

A. Never
B. About once a month or less
C. About once a week
D. About once a day
E. More than once a day
Internet Multimedia is Growing

- Youtube gets 300 hours of video uploaded every minute [2016]
- Instagram gets more than 80 million photos uploaded every day [2016]
- Twitter gets 6000 tweets per second (~ 500 million per day) [2017]
Multimedia Retrieval

- Can you search for an image, video, or audio file by anything other than file name?

- Multimedia retrieval is improving, but still really difficult
  - *e.g.* SoundHound/Shazam, [http://images.google.com](http://images.google.com)

- Yet multimedia retrieval is already good enough to cause major privacy issues that are not easy to solve
  - We will focus on *geotagging* here
Workaround: Manual Tagging

- User provides information that can be used for search and retrieval later
Workaround: Geotagging

- **Geotagging** is the process of adding geographical identification metadata to media.
Workaround: Geotagging

- Geotagging is the process of adding geographical identification metadata to media
  - Allows for easier clustering of photo and video series
Workaround: Geotagging

- **Geotagging** is the process of adding geographical identification metadata to media
  - Allows for easier clustering of photo and video series

- Many social media portals provide Application Programming Interfaces (APIs) to access geotag data – along with other user data!
  - This includes Youtube, Twitter, Instagram, Flickr, ...
Unintended Consequences

  - Map out geo-tagged data from unsecured Twitter or Instagram accounts
Case Study: Craigslist

- Many ads on Craigslist are anonymized... except for geotags on photos
  - More geotagged photos = higher coordinate accuracy
  - Sometimes ad is for high-valued goods (e.g. cars, diamonds)
  - Sometimes ad specifies availability (e.g. “call Sunday after 6pm”)

![Image of geotagged photo and map location]
Question

What do you think has to be done?


A. Nothing can be done – privacy is dead
B. We need to educate people about this and try to save privacy [fight]
C. I agree with (A) and will really think before I post
D. I agree with (B) and will really think before I post
E. I won’t post anything anymore [flee]
Outline

- Privacy
- Online Interactions
- Multimedia Retrieval
  - Geo-tagging
- Apps and Access
How Secure is Your Phone?

- Nowadays, your smart phone contains or has access to all sorts of personal information about you
  - How do you keep it safe?
  - Do you cover your screen when you unlock?
  - How often do you change your lock?

- Even if our phones are physically secure, we often give away our privacy!
Permissions: Android

- Contacts
  - “Use your device’s contacts, which may include the ability to read and modify your contacts”

- Calendar
  - “Read, add and modify calendar events as well as send email to guests without owners’ knowledge”
Permissions: Facebook

- “Most Used Words” went viral in late 2015
Permissions: Facebook

- In general, be very wary of “quizzes”
  - [http://www.itscovarr.com/2016/05/the-trouble-with-online-quizzes/](http://www.itscovarr.com/2016/05/the-trouble-with-online-quizzes/)
Privacy Settings

- *Always* pay attention to privacy settings
  - Usually dictates visibility, ownership, and distribution of data

- Actually read the Terms of Service (ToS)
  - tos;dr [https://tosdr.org/](https://tosdr.org/)
Website Tracking

- Recall the connectivity model of the Internet
  - Info sent between machines, meaning not everything is local
  - This is why “private browsing” is not truly anonymous

- Individual websites sometimes store information about you in “cookies”

- Browsers often collect data on you as well
Be Deliberate About What You Reveal

- Remember that you’re not really anonymous and that the WWW is accessible to everyone
  - Even on “closed” sites like Facebook

- Remember that digital information on the Internet rarely ever goes away, even if you try to delete it
  - Archives, or copies on other machines

- Even if you choose to do the revealing (no privacy violation), might not be wise to reveal “all”
  - What image do you want to present?
Privacy in Our New Reality

- Everything is “public” and persistent
  - People have lost jobs based on Facebook and Twitter posts
  - Resurfacing of digital information (old posts, emails, photos) is almost an inevitability

- People are aware of availability of information
  - Human Resources (HR) departments sometimes check your social media profiles when you apply for jobs
  - People often “cyberstalk” others before or after meeting
Want to Learn More?

- [http://teachingprivacy.org/](http://teachingprivacy.org/)