Privacy of Social Media

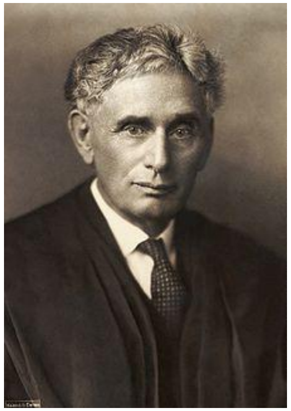
Irwin King

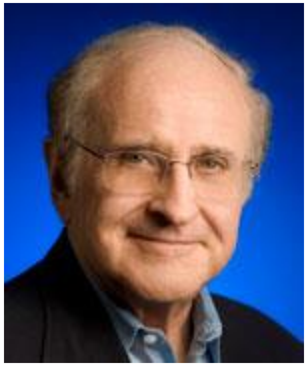
Department of Computer Science and Engineering
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Slides are modified from Privacy and Networks CPS 96 and Social Networking Security and Privacy

What do we mean by privacy?

- Louis Brandeis (1890)
  - "right to be left alone"
  - protection from institutional threat: government, press

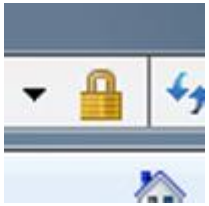




- Alan Westin (1967)
  - "right to control, edit, manage, and delete information about themselves and decide when, how, and to what extent information is communicated to others"

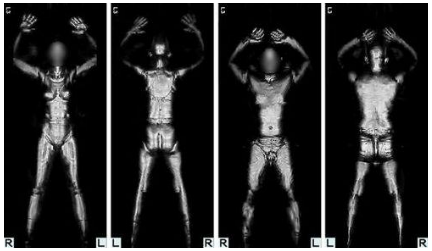
Privacy vs. security

Privacy: what information goes where?



Security: protection against unauthorized access

- *Security helps* enforce privacy policies
- Can be *at odds with each other*
  - e.g., invasive screening to make us more “secure” against terrorism



Types of social media services

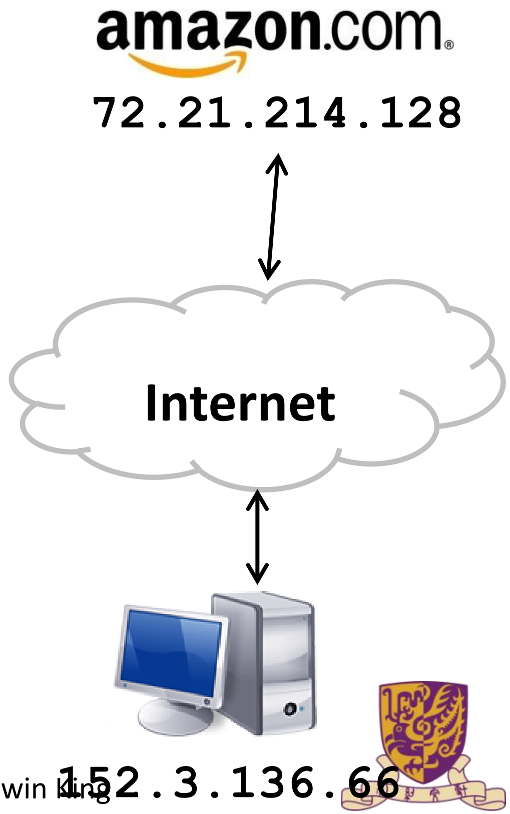
- Networking
  - Facebook, Google+, Linkedin, Twitter
- Content Sharing
  - Pinterest, Facebook, Dropbox, Google Drive
- Location-based Services
  - foursquare, Google Latitude, Facebook, Gowalla

Privacy-sensitive information

- Identity
  - name, address, SSN
- Location
- Activity
  - web history, contact history, online purchases
- Health records
- …and more

Tracking on the web

- IP address
  - Number identifying your computer on the Internet
  - Visible to site you are visiting
  - Not always permanent
- Cookies
  - Text stored on your computer by site
  - Sent back to site by your browser
  - Used to save prefs, shopping cart, etc.
  - Can track you even if IP changes



Types of Protection

- Security
  - Prevention of malicious action to systems, info
- Safety
  - Prevention from physical or mental harm
- Privacy
  - Prevention of exposing sensitive or private info

Default Privacy Modes

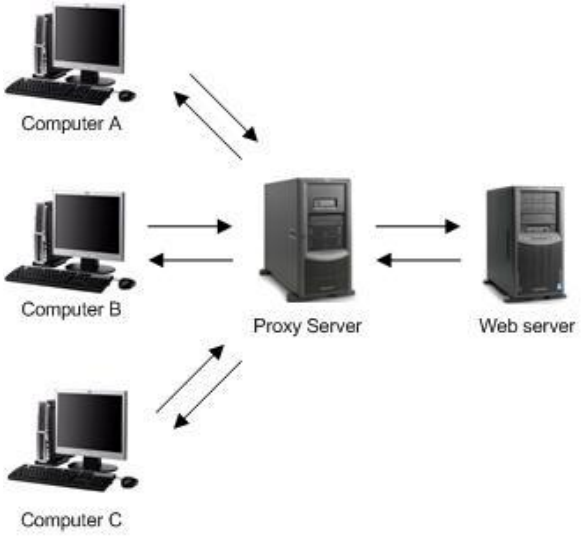
- "Mostly open"
  - The default sharing mode is **public**
  - You must choose to keep content private
- "Mostly closed"
  - The default sharing mode is **private**
  - You must choose to share content

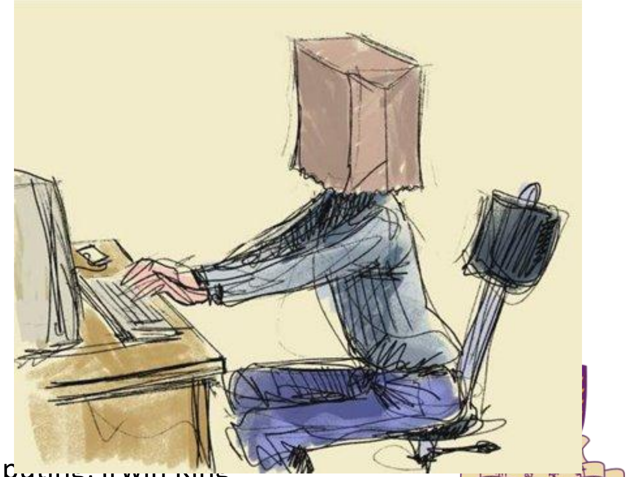
Alternatives?

- Anonymization
  - Do not use real names
- Encryption
  - NOYB, flyByNight
- Decentralization
  - Tighter control over data

Anonymization

- Hide identity, remove identifying info
- Proxy server: connect through a third party to hide IP
- Health data released for research purposes: remove name, address, etc





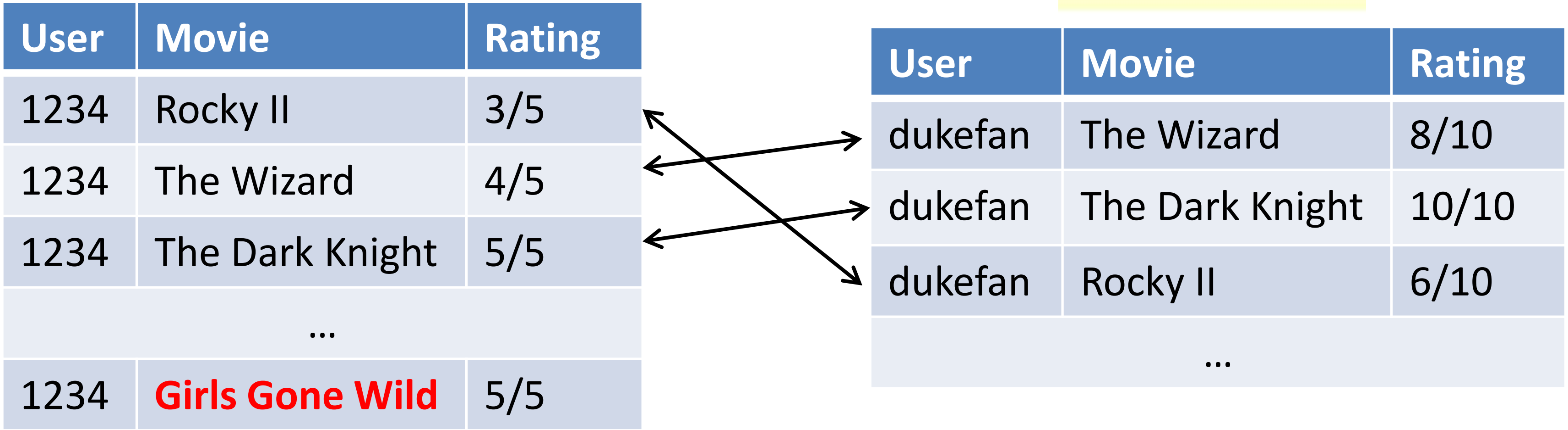
Threat: deanonymization

- Netflix Prize dataset, released 2006
- 100,000,000 (private) ratings from 500,000 users
- Competition to improve recommendations
  - i.e., if user X likes movies A,B,C, will also like D
- Anonymized: user name replaced by a number

Threat: deanonymization

- Problem: can combine “private” ratings from Netflix with public reviews from IMDB to identify users in dataset
- May expose embarrassing info about members…

Threat: deanonymization



User	Movie	Rating
1234	Rocky II	3/5
1234	The Wizard	4/5
1234	The Dark Knight	5/5
…		
1234	Girls Gone Wild	5/5

User	Movie	Rating
dukefan	The Wizard	8/10
dukefan	The Dark Knight	10/10
dukefan	Rocky II	6/10
…		

User 1234 is dukefan!

Threat: deanonymization

- Lesson: cannot always anonymize data simply by removing identifiers
- Vulnerable to aggregating data from multiple sources/networks
- Humans are predictable
  - E.g., try Rock-paper-scissors vs AI

Location privacy

- Mobile phones:
  - – Always in your pocket
  - – Always connected
  - – Always knows where it is: GPS
- Location-based services
- Location-based ads
- What are we giving up?

Why, when and what to disclose?

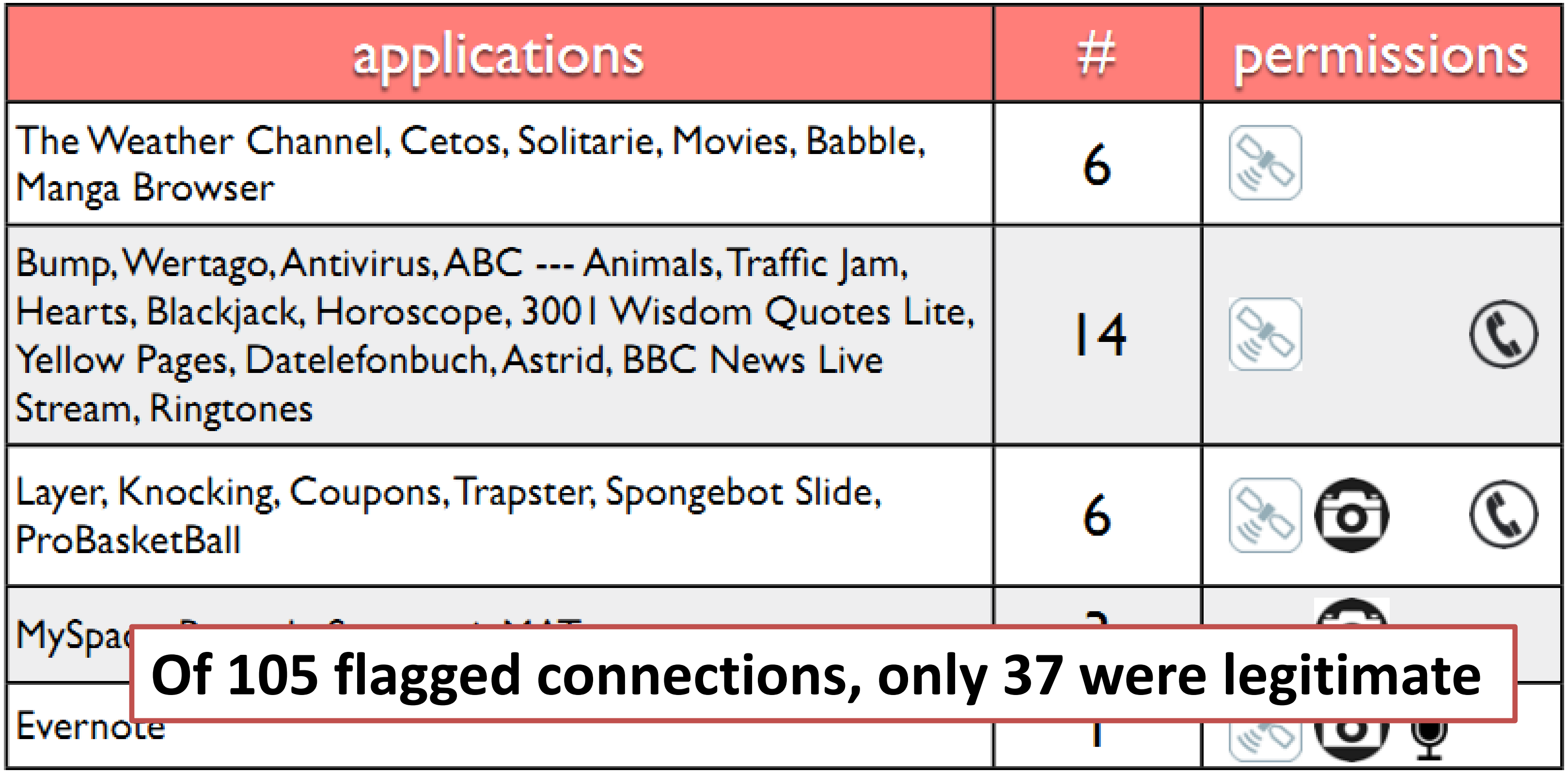
- It is not a simple question!
- Tradeoff between functionality
- Also important whom to disclose it to?
  - Relatives
  - Co-workers
  - Friends
- There have been studies about this
  - Not easy to classify
  - People want to disclose only what is useful

How is your data used by apps?

- Many “free” apps supported by ads
- Analytics: profiling users
- Our research: found it common for popular free apps to send location+device ID to advertising and analytics servers
- What can we do?
  - – More visibility into what app does with data once it reads it

Application Study

- 30 popular Android applications that access Internet, camera, location or microphone



applications	#	permissions
The Weather Channel, Cetos, Solitarie, Movies, Babble, Manga Browser	6	🛰
Bump, Wertago, Antivirus, ABC --- Animals, Traffic Jam, Hearts, Blackjack, Horoscope, 3001 Wisdom Quotes Lite, Yellow Pages, Datelefonbuch, Astrid, BBC News Live Stream, Ringtones	14	🛰 ☎
Layer, Knocking, Coupons, Trapster, Spongebot Slide, ProBasketBall	6	🛰 📷 ☎
MySpa	3	📷
Evernote	1	🛰 📷 🎤

Of 105 flagged connections, only 37 were legitimate

Findings - Location

- 15 of the 30 applications shared physical location with an ad server
- Most of this information was sent in the clear
- In no case was sharing obvious to user
    - Or written in the EULA
    - In some cases it occurred without app use!

Findings – Phone identifiers

- 7 applications sent device unique identifiers (IMEI) and 2 apps sent phone info (e.g. phone number) to a remote location without warning
  - One app’s EULA indicated the IMEI was sent
- Appeared to be sent to app developers

“There has been cases in the past on other mobile platforms where well-intentioned developers are simply over-zealous in their data gathering, without having malicious intent.” -- Lookout

What are the risks?

- Privacy
- Reputation
- Data
- Access
- Control
- Employment
- Legal Proceedings