

---

# What is Privacy?



---

# What is Privacy?

- Warren and Brandeis: “the right to be let alone”
- FIPS PUB 41: “The right of an entity . . . to determine the degree to which it will interact with its environment, including the degree to which the entity is willing to share its personal information with others”
- OSI: “The right of individuals to control or influence what information related to them may be collected and stored and by whom and to whom that information may be disclosed”

---

## Using versus Gathering

- The primary concern is how information is *used*
- Obtaining information is often much less of a concern
- Note, though, that a lot of personal information is considered private even from one other person

---

# Legal Foundations of Privacy

- Common law: “[T]he house of every one is to him as his castle and fortress.” Semayne’s Case, 5 C. Rep. 91a, 77 Eng. Rep. 194 (K.B. 1603)
- Doesn’t work as well in today’s interconnected world
- Types of privacy:
  - Bodily integrity
  - Decisional privacy
  - Information privacy
  - Communications privacy

---

## Why Violate Privacy?

- Thoughtlessness
- Efficiency, especially for marketing
- New markets (i.e., new location-based offerings)
- Public safety and national security

---

## How Do We Lose Privacy?

- Voluntarily
- Compulsion
- Reuse of data

👉 *This sort of secondary use is the source of most privacy violations*

---

## Voluntary Surrender of Data

- Social networking sites
- Purchases (Netflix, Amazon)
- Warranty cards

---

# Compulsion

- Various interactions with governments (marriage, property purchases, etc.)
- Boarding an airplane
- “Contracts” — e.g., getting a credit card in exchange for information



---

## Secondary Use

- We may not object — or object too much — to the initial collection of certain data
- Often, we benefit from the initial collection, and hence regard it as a fair trade
- When it is used for another purpose without our knowledge or consent, trouble often results

---

## Example: Bars and Drivers' Licenses

- Many bars use swipe readers to verify that the proffered license is genuine
- (Better-grade fakes have mag stripe data anyway...)
- But — the readers copy the data: name, address, gender, etc.

---

## What are the Privacy Violations?

- Using license data to establish age
- Using license data for marketing

---

## Data on a Driver's License

- Primary purpose: certification that you are legally allowed to drive
- Primary purpose of picture: assurance that the bearer is indeed the license holder
- Demographic data: accountability in event of violations
- *Not* intended for proof of age, *not* intended as an airplane boarding credential

---

## Age Verification

- Even if age verification is acceptable — and use of licenses for that is certainly accepted by the states — use of the additional data for marketing is not
- Resale of license data happens to be illegal, but not for that reason

---

## Example: MetroCard

- Primary purpose: paying subway or bus fare
- But — the MTA retains your trip information
- This data can be and has been used for criminal and divorce cases

# The London Oyster Card

Customer Reminder   Receipts   Your Bank   Call For Assistance

### View Oyster card usage

When	Added	Deducted	Balance	Description
18:34 Fri 19 Sep		2.00	1.20	Canary Whf - High St Ken
17:48 Fri 19 Sep	2.50		3.20	Pay as you go adjustment
13:53 Fri 19 Sep		4.00	0.70	Blackfriars - Uncompleted
10:59 Fri 19 Sep		1.50	4.70	High St Ken - Monument
21:09 Thu 18 Sep		1.30	6.20	St James Pk - High St Ken
14:35 Thu 18 Sep		2.00	7.50	Cutty Sark DLR - Tott Ct Rd
10:40 Thu 18 Sep		1.50	9.50	High St Ken - Tower Hill
15:12 Wed 17 Sep		1.50	11.00	Liverpl St - High St Ken

Current pay as you go balance: **£1.20**

View Oyster card Usage   Back Screen   Cancel

---

# Linkages

- Sometimes, items from two or more databases are linked
- Then possible to learn *much* more
- Prerequisite: common data item



---

## Linkages: MetroCard

- How did you pay for your last MetroCard? Credit card?
- That links the MetroCard to a person
- Query: who boarded the subway at 116th and Broadway between 3:30 and 3:45 AM last Tuesday?
- In principle, at least, that question may be answerable

---

## Deeper Linkages

- Correlate on patterns
- Example: assume a MetroCard is used infrequently, but at only two stops, Penn Station and 116th St
- Is there any one person who used a credit card to buy train tickets for just those days?
- (Note: I have no idea if that has actually been done)

---

# Identity

- Sometimes, anonymous data can be linked to a specific person
- Other times, behavior identifies you
- Linkages can be used to establish identity
- MetroCards are anonymous — but credit cards aren't

---

# Authentication

- If you're an authenticated user, your behavior can be tracked more easily over time
- (This includes Google, many media sites, etc.)
- Sometimes, even free accounts ask for demographic information, to improve profiles and ad targeting

---

# “On the Internet, Nobody Knows You’re a Dog”

- (Famous *New Yorker* cartoon)
- Often, what matters is not *who* you are, but what you do
- Example: for targeted ads, your identity doesn’t matter, your interests do

---

## Online and Offline

- You're profiled online and in the physical world
- Sometimes, the two are linked
- Profiling isn't new — but people have gotten a lot better at it

---

## Offline

- Credit reports
- Credit card purchases
- Loyalty card programs
- Magazine subscriptions
- Warranty cards
- Public data (e.g., mortgages)
- Zip code demographics

---

## Online

- Cookies
- “Flash local storage”
- Third-party ad sites



---

## What's a Cookie?

- “Small text file stored on your computer”
- Set by a site; sent back to it next time you visit
- True — but frequently used to track you
- Persistent identifier, retained across sessions
- Not necessarily linked to a particular person — but the same each time you come back
- Linked to particular sites; one site can't retrieve another site's cookies

---

## Good Uses for Cookies

- Login data
- Site preferences
- (Sometimes) shopping cart information

---

## What Your Browser Reveals

- Ordinary: `http://chadash.cs.columbia.edu/`
- Advanced: `http://panopticlick.eff.org/`
- (Please visit these sites on your own)

---

## Third-Party Ad Sites

- Most ads on the web come from third parties, not the site you're visiting
- These third-party sites have their own cookies, which they set and receive
- If an ad site places content on multiple pages, they'll know which of those pages you visit; this lets them build up a very complete profile of your interests
- Sometimes, sites pass extra information about you to the ad providers
- One of the biggest ad providers is Doubleclick, which is owned by Google. . .

---

## Media Sites

- Many media sites, including at least the *New York Times* and the *Wall Street Journal*, track what types of articles you read
- This information is used for targeted advertising

---

## Linking Online and Offline

- Online, it's easy to build a good profile of people
- If you buy something online, that site knows your name
- Use third-party cookies to associate your interest profile with a name

---

## Credit Cards

- Most people have only a few credit cards
- If you use the same card for online and offline purchases, your physical person in a store can be linked to online behavior
- Special features have been put into some network protocols to facilitate this

---

## Profiling: Good or Bad?

- Good: you see only ads you're interested in
- Bad: profiling is unpleasant. Besides, if you see interesting ads you're more likely to buy...



---

## Fair Information Practices

- First “code of fair information practices” developed in 1973 at HEW
- Basic rules for minimizing information collection, ensuring due process, protection against secret collection, provide security, ensure accountability
- Emphasize individual knowledge and consent
- Principles are broadly accepted, but individual principles not implemented uniformly

---

## Fair Information Principles and Practices

- Collection limitation
- Data quality
- Purpose specification
- Use limitation
- Security
- Openness/notice
- Individual participation
- Accountability