Introduction to Privacy

Michelle Mazurek

Some slides adapted from Lorrie Cranor, Elaine Shi, Christin Trask, and Yu-Xiang Wang
Today ...

• Defining privacy

• How privacy is protected
  – Laws and regulations
  – Self-regulation and competitive markets
  – Technical solutions
Privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all.”

Robert C. Post

Three Concepts of Privacy, 89 Geo. L.J. 2087 (2001)
Boundaries of self?

“The right to be let alone”
- Samuel D. Warren and Louis D. Brandeis,
  The Right to Privacy, 4 Harv. L. Rev. 193 (1890)

“Our concern over our accessibility to others: the extent to which we are known to others, the extent to which others have physical access to us, and the extent to which we are the subject of others’ attention.”
- Ruth Gavison,
  Privacy and the Limits of the Law, Yale Law Journal 89 (1980)
Control over information?

“Privacy is the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others. … Each individual is continually engaged in a personal adjustment process in which he balances the desire for privacy with the desire for disclosure and communication.”

- Alan Westin, Privacy and Freedom, 1967
Privacy definitions and goals

• Solitude, uninterrupted
• Unseen, unheard, unread
• Not talked about
• Not judged/misjudged
• Not profiled, targeted, treated differently
• Free to practice, make mistakes

• Being unknown
• Being forgotten
• Intimacy
• Control
• Boundaries

What do these mean in the digital age?
Privacy frameworks/axes

• Individual vs. communitarian
  – Principle vs. practice

• Data protection vs. personal privacy

Examples? Tensions between them?
Measuring privacy is hard, why?

- How much privacy does a person have?
- What is a person’s attitude toward privacy?
- What is the cost of a privacy invasion?
How privacy is protected

• Laws, self regulation, technology
  – Notice and access
  – Control over collection, use, deletion, sharing
  – Collection limitation
  – Use limitation
  – Security and accountability
Option 1: Privacy laws/regulations

• In the U.S., no explicit constitutional right
  – Some privacy rights inferred from constitution
• No general privacy law; some sector-specific
  – Health, financial, education, children, etc.
  – FTC jurisdiction over fraud, deceptive practices
  – FCC regulates telecomms
  – Some state and local laws
• Overall, relatively few protections
European Data Protection Directive

- EU countries must adopt comprehensive laws
- Privacy is a fundamental human right
- Privacy commissions in each county
- New(ish) “right to be forgotten”
  - http://www.stanfordlawreview.org/online/privacy-paradox/right-to-be-forgotten
OECD fair information principles

• Collection limitation
• Data quality
• Purpose specification
• Use limitation
• Security safeguards
• Openness
• Individual participation
• Accountability

http://oecdprivacy.org/
US government privacy reports

- U.S. FTC and White House reports released in 2012
- U.S. Department of Commerce multi-stakeholder process to develop enforceable codes of conduct
Option 2: Privacy self regulation
Notice and choice

Protect privacy by giving people control over their information

**Notice** about data collection and use

**Choices** about allowing their data to be collected and used in that way
We will talk about this again: Policies and notices
Requirements for meaningful control

• Individuals must:
  – Understand **what** options they have
  – Understand **implications** of their options
  – Have the **means** to exercise options

• Costs must be reasonable
  – Money, time, convenience, benefits
Why don’t we have a market for privacy?
Privacy concerns seem inconsistent with behavior

• People say they want privacy, but don’t always take steps to protect it (the “privacy paradox”)

• Many possible explanations
  – They don’t really care that much about privacy
  – They prefer immediate gratification to privacy protections that they won’t benefit from until later
  – They don’t understand the privacy implications of their behavior
  – The cost of privacy protection (including figuring out how to protect their privacy) is too high
Nobody wants to read privacy policies

“the notice-and-choice model, as implemented, has led to long, incomprehensible privacy policies that consumers typically do not read, let alone understand”

Cost of reading privacy policies

• What would happen if everyone read the privacy policy for each site they visited once each month?
  • Time = 244/hours year
  • Cost = $3,534/year
  • National opportunity cost for time to read policies: $781 billion

Requirements for meaningful control

• Individuals must:
  – Understand **what** options they have
  – Understand **implications** of their options
  – Have the **means** to exercise options

• Costs must be reasonable
  – Money, time, convenience, benefits
Option 2b: Computer reads for you

- Platform for Privacy Preferences (P3P)
- W3C specification for XML privacy policies
  - Proposed 1996
  - Adopted 2002
- Optional P3P compact policy HTTP headers to accompany cookies
- Goal: Your agent enforces your preferences
Criticisms of P3P

• Too complicated, hard to understand
• Lacks incentives for adoption
  – Only major companies?
PrivacyFinder: P3P search engine

• Checks each search result for computer-readable P3P privacy policy, evaluates against user’s preferences

• Composes search result page with privacy meter annotations and links to “Privacy Report”

• Allows people to comparison shop for privacy

• [http://privacyfinder.org/](http://privacyfinder.org/)
Install the Privacy Finder search bar!
Compatible with Firefox 2 and Internet Explorer 7

Frequently Asked Questions | Privacy Policy | Add A Site | Download Privacy Bird® Software | Privacy Finder 日本語版

Privacy Finder is a public service provided by CMU Usable Privacy and Security Laboratory
Discover a variety of athletic and casual shoes at Zappos.com.

Visit the official website of Nike to explore their products.

Explore the official online store of Onlineshoes.com for a variety of brands.

Shoebuy.com offers free shipping and return shipping on mens' and womens' shoes, sandals, boots, and sneakers.

Find a wide selection of footwear in the Yahoo! Directory.
Shoebuy.com, Inc. Privacy Practices

Privacy Policy Check

Shoebuy.com, Inc.'s privacy policy does not match your preferences:

- Site may contact you to interest you in other services or products and does not allow you to remove yourself from marketing/mailing list

Shoebuy.com, Inc. may share your information with:

- Companies that help this site fulfill your requests (for example, shipping a product to you), but these companies must not use your information for any other purpose
- Delivery companies that help this site fulfill your requests and who may also use your information in other ways

Privacy Policy Summary

Policy Statement 1 - Access log information

Our Web server collects access logs containing this information.

Policy Statement 2 - Cookies

Cookies are used to track visitors to our site, so we can better understand what portions of our site best serve you. We also use cookies to allow our server to maintain information about the contents of your shopping cart.

Policy Statement 3 - Transaction info (required)

Information we collect in order to process your purchase.

Policy Statement 4 - Transaction info (optional)
Impact on decisionmaking

- Online shopping study conducted at CMU lab
- Participants buy with their own credit cards
  - Bought batteries and a sex toy
- Pay them a fixed amount; keep the change
- Result: When information is accessible, many people will pay (a little) more for privacy

P3P in Internet Explorer

- Implemented in IE 6, 7, 8, 9, 10 ...
- “Compact policy” (CP)
- If no CP, reject third-party cookies
- Reject unsatisfactory third-party cookies
No P3P syntax checking in IE

• Accepts bogus tokens, nonsense policies
• Valid:

  CAO DSP COR CURa ADMa DEVa OUR IND PHY ONL UNI COM NAV INT DEM PRE

• Also accepted:

  AMZN

  Facebook does not have a P3P policy. Learn why here: http://fb.me/p3p

Microsoft uses a “self-declaration” protocol (known as “P3P”) dating from 2002 .... It is well known – including by Microsoft – that it is impractical to comply with Microsoft’s request while providing modern web functionality.
Can policy agents ever work?

• Simplify the practices enough?
• Require users to specify their preferences?
  – Learn them via AI?
• Incentives for broad adoption?
Requirements for meaningful control

• Individuals must:
  – Understand **what** options they have
  – Understand **implications** of their options
  – Have the **means** to exercise options

• Costs must be reasonable
  – Money, time, convenience, benefits
Option 3: The power of math

• Can we provide strong guarantees that don’t rely on good behavior from the data collector?
• Sort of!
• Differential privacy, invented by Cynthia Dwork
Privacy and Justin Bieber

• Suppose you are handed a survey:
  – Do you like listening to Justin Bieber?
  – How many Justin Bieber albums do you own?
  – What is your gender?
  – What is your age?

• After analysis, results will be released publicly
  – Do you feel safe submitting a survey?
  – Should you?
Brief notation

Q is the privatized query run on the data set, and R is the result released to the public.

\[ D_I = \{ d_i \mid i \in I \} \]
What do we want? (Privacy)

- My answer has no impact on the released results

- Any attacker looking at published $R$ can’t learn anything new about me personally (high probability)

- $Q(D_{(I-me)}) = Q(D_1)$

- $\Pr[\text{secret(me)} \mid R] = \Pr[\text{secret(me)}]$
Why can’t we have it?

- If individual answers had no impact, the results would be useless

- Trends in \( R \) may be true of me too. (If I am 15, do I like Justin Bieber?)

- By induction, \( Q(D_{(l)}) = Q(D_{\emptyset}) \)

- \( \Pr(\text{secret(me)} | \text{secret(Pop)} > \Pr(\text{secret(me)}) \)
Why can’t we have it?

If an attacker knows a function about me dependent on the general population:

- I’m 2x average age
- I’m the majority gender

Then the attacker knows things about me even if I don’t submit a survey!

- \( \text{age(me)} = 2 \times \text{mean}_\text{age} \)
- \( \text{gender(me)} = \text{mode}_{\text{gender}} \)
- \( \text{mean}_\text{age} = 16 \)
- \( \text{mode}_{\text{gender}} = F \)
- \( \text{age(me)} = 32 \text{ AND } \text{gender(me)} = F \)
What can we have instead?

- The chance that the released result will be $R$ is nearly the same, **regardless** of whether I submit a survey
- There is no (well, *almost* no) additional harm from submitting the survey
Differential privacy

\[
\frac{\Pr[Q(D_l) = R]}{\Pr[Q(D_{l\pm i}) = R]} \leq A, \quad \text{for all } l,i,R
\]

- If A=1, there is 0 utility (individuals have no effect)
- If A >> 1, there is little privacy
- A should be chosen by collector to be close to 1
What this means

• Probability of result is nearly the same, regardless of whether I submit a survey

\[ \Pr[R] = X \quad \text{and} \quad \Pr[R] = Y \]

\[ X \approx Y \]

• How can anyone guess which world is true?
Popular misconception

• The attacker can’t learn anything about me from the results (protection against all harms)

• NOPE: Background information still applies. Attackers can use aggregate results.
How to do it (high-level)

• Output perturbation: Return query answer plus some noise

• Input perturbation: Add noise to survey data before storing

• Perturbation of intermediate results

• Sample and aggregate
  – Ask Q over smaller samples; aggregate results
Challenges

• Utility / privacy tradeoffs
  – May require really large datasets
  – May result in ridiculous answers

• Privacy budget depletion
  – Each query reduces what else can be asked

• How can this fit in with personal privacy (as opposed to data protection?)

• What does a differential policy mean in practice?
Requirements for meaningful control

• Individuals must:
  – Understand **what** options they have
  – Understand **implications** of their options
  – Have the **means** to exercise options

• Costs must be reasonable
  – Money, time, convenience, benefits

How can/should these approaches (laws, notice/choice, technical solutions) be balanced?