

Privacy

Privacy \neq Security

- Security is about adversaries
 - Integrity
 - Confidentiality
- But privacy is about requirements
 - When you say "confidential" — what? and from whom?

Privacy Is Personal

- Different people have different privacy concerns
 - Job seekers
 - Early adopters may expect more risk
 - People in different countries have different expectations (Norway & Sweden have public tax returns)
 - Vulnerable populations in repressive regimes

Data Protection Principles (Sommerville)

- Awareness and control (users should know and control what data are collected)
- Purpose (tell users why data are being collected; only use data for those purposes)
- Consent (have consent before disclosing data to others)
- Data lifetime (only keep data for as long as needed)
- Secure storage
- Discovery and error correction (allow users to find out what data you store and correct errors)
- Location (don't store data where weaker protection laws apply)

Two Topics Today

- Differential privacy (injecting noise to enable privacy-preserving data analysis)
- Privacy policies

Protecting Sensitive Responses

- Survey students: ask if they cheated on the exam
 - But surely they'd refuse to tell the truth
- Ask each student to flip a coin
 - If heads, tell the truth
 - If tails, flip a second coin and report its result
- Even if personally identifiable results are published, still can't tell who cheated!

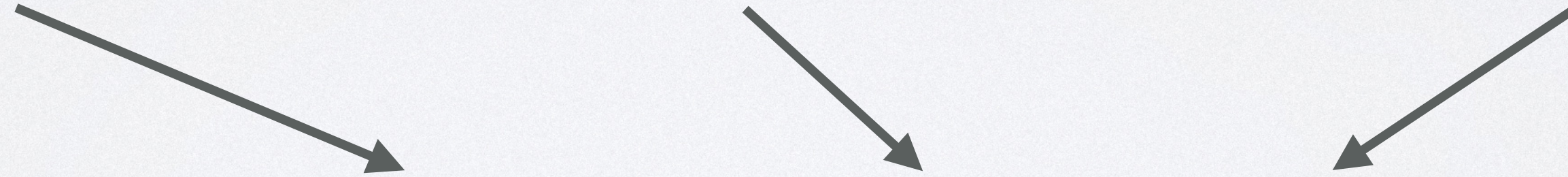
Estimating Cheating

- c is the fraction of students who cheated
- r is the number who report cheating

1/4 of non-cheaters
flip "tails" and then flip
"report cheating"

1/4 of cheaters also
flip "tails" and then flip
"report cheating"

1/2 of cheaters also
flip "heads" and then
report the truth


$$r = \frac{1}{4}(1 - c) + \frac{1}{4}c + \frac{1}{2}c$$

Estimating Cheating

$$r = \frac{1}{4}(1 - c) + \frac{1}{4}c + \frac{1}{2}c$$

$$r = \frac{1 - c + c + 2c}{4}$$

$$r = \frac{1 + 2c}{4}$$

$$\frac{4r - 1}{2} = c$$

$$c = 2r - \frac{1}{2}$$

Differential Privacy

- General principle: adding noise to data protects identities
- Question: how much noise is needed?
- Answer: depends on how safe you want to be

A Risky Data Set

- Assume: you don't want the world to know you have cancer
- Rows 1 and 3 are the same except for your name
- You're OK releasing this data set if no one will know YOUR NAME is in it
- Proposal: scramble the data *randomly* with mechanism M.
- Intuition: an adversary shouldn't be able to tell whether YOUR NAME was in there.

Name	Has Cancer
Someone	TRUE
Someone else	FALSE
YOU	TRUE

Differential Privacy Definition

- Database consists of a collection of rows.
- Goal: protect rows using mechanism M , which maps rows to randomly-chosen vectors
- M is ϵ -differentially private if adding or removing a row only affects the probability of any outcome by a small factor
- **Definition:** A randomized mechanism M is ϵ -differentially private if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq \text{Range}(M)$:

$$\Pr[M(D_1) \in S] \leq \exp(\epsilon) \times \Pr[M(D_2) \in S]$$

note: $\exp(\epsilon) \approx 1 + \epsilon$ for small ϵ

- $\Pr[M(D_1) \in S] \leq \exp(\varepsilon) \times \Pr[M(D_2) \in S]$

- Adversary sees $M(D_1)$ and tries to guess whether YOUR data is included in the original data set
- But if $M(D_1) = M(D_2)$, the adversary has no hope (no data leakage at all)
 - But maybe you can't do anything useful with the data
- Presumably $M(D_1) \neq M(D_2)$, but by how much?
- If $\varepsilon = 0$, there is no leakage (but maybe we can't come up with a good M)
- Larger ε means more leakage (but allows stronger data usage)

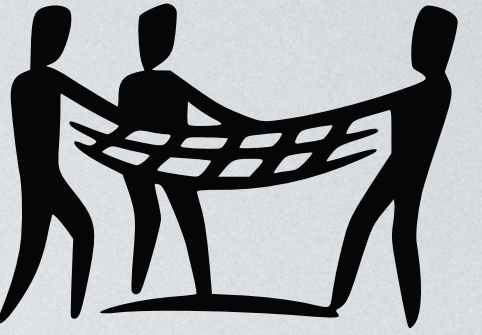
Privacy Budget

- Previous definition pertains to one query
- Each query leaks data
- So you have to be careful how many queries you allow!

Real World Usage

- U.S. Census
- Apple feature usage data
- Facebook

Which of the Following Best Describes Differential Privacy?



- A. A technique that ensures data is encrypted before being stored in a database.
- B. A process of anonymizing data by removing all personally identifiable information.
- C. A software tool used to detect and prevent unauthorized access to databases.
- D. A statistical method that adds noise to data to protect individual privacy while maintaining the utility of the dataset.

Activity: Privacy Policies

- What do you want to know about privacy (as a user)?
- What do privacy policies actually say?
- What will you put in your privacy policy?

When I see a privacy policy, I...

Read the whole thing carefully

0%

Skim it quickly

0%

Just accept it

0%

It depends (e.g. on what the privacy policy is for)

0%

What Do You Want To Know?

- Scenario
 - You are considering using a video streaming service.
 - Of course, the service will know which videos you have watched.
 - To sign up, link your Facebook account so the service can see who your friends are.
 - Or pay \$10/month for a version with ads.
 - What do you want to see in the privacy policy?

Research on Privacy Policies

- McDonald et al. compared three formats:
 - Layered (short, standardized form + full policy)
 - Privacy Finder report (standardized)
 - Conventional (non-standardized)
- Standardized formats improved speed

Search Engine: Google Yahoo! Preference Level:

Yahoo! Inc. Privacy Practices

[Privacy Policy Summary](#) | [Full Privacy Policy](#) | [Contact Site](#) | [P3P Policy](#)

Privacy Policy Check

Yahoo! Inc.'s privacy policy *does not match your preferences:*

- Site may use health or medical information for analysis or to make decisions that may affect what content or ads you see, etc.
- Site may use health or medical information for marketing
- Site may share health or medical information with other companies (other than those helping the site provide services to you)
- Site may share financial information or information about your purchases with other companies (other than those helping the site provide services to you)
- Site may contact you to interest you in other services or products and does not allow you to remove yourself from marketing/ mailing list
- Site may share information that personally identifies you with other companies (other than those helping the site provide services to you)
- This site collects data for an unknown purpose

Yahoo! Inc. may share your information with:

- People who may access your information from a public area, such as a bulletin board, chat room, or directory
- Other companies whose privacy policies are unknown to this site
- Companies that are accountable to this site, though their privacy policies may be different from this site's
- Companies that have privacy policies similar to this site's
- Delivery companies that help this site fulfill your requests and who may also use your information in other ways
- 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

Privacy Policy Summary

Policy Statement

[Show data collection, use, and sharing details...](#)

Access to your information

This site gives you access to your contact information and some of its other information identified with you

How to reach this site

Yahoo! Inc.
701 First Avenue
Sunnyvale, CA 94089 USA

<http://help.yahoo.com/help/privacy/privacy/index.html>

Privacy Nutrition Labels (CHI 2010)

- Standardized presentations had a positive effect on accuracy, speed, and enjoyment

Acme

information we collect

ways we use your information

information sharing

	provide service and maintain site	marketing	telemarketing	profiling	other companies	public forums
contact information		opt out	opt out			
cookies						
demographic information		opt out	opt out			
financial information						
health information						
preferences		opt out	opt out			
purchasing information		opt out	opt out			
social security number & gov't ID						
your activity on this site		opt out	opt out			
your location						

Access to your information

This site gives you access to your contact data and some of its other data identified with you

How to resolve privacy-related disputes with this site

Please email our customer service department

acme.com

5000 Forbes Avenue
Pittsburgh, PA 15213 United States
Phone: 800-555-5555
help@acme.com

Acme

information we collect

ways we use your information

information sharing

	provide service and maintain site	marketing	telemarketing	profiling	other companies	public forums
contact information		opt out	opt out			
cookies						
demographic information		opt out	opt out			
preferences		opt out	opt out			
purchasing information		opt out	opt out			
your activity on this site		opt out	opt out			

Information not collected or used by this site: social security number & government ID, financial, health, location.

Access to your information

This site gives you access to your contact data and some of its other data identified with you

How to resolve privacy-related disputes with this site

Please email our customer service department

acme.com

5000 Forbes Avenue
Pittsburgh, PA 15213 United States
Phone: 800-555-5555
help@acme.com



we will collect and use your information in this way



we will not collect and use your information in this way



by default, we will collect and use your information in this way unless you tell us not to by opting out



by default, we will not collect and use your information in this way unless you allow us to by opting in

Privacy Nutrition Labels

App Privacy

[See Details](#)

The developer, Apple, indicated that the app's privacy practices may include handling of data as described below. For more information, see the [developer's privacy policy](#).



Data Not Linked to You

The following data may be collected but it is not linked to your identity:



Health & Fitness



Usage Data



Location



Diagnostics

Privacy practices may vary, for example, based on the features you use or your age. [Learn More](#)

App Privacy

The developer, Apple, indicated that the app's privacy practices may include handling of data as described below. For more information, see the [developer's privacy policy](#).

Learn how the developer lets you [manage your privacy choices](#).

To help you better understand the developer's responses, see [Privacy Definitions and Examples](#).

Privacy practices may vary, for example, based on the features you use or your age. [Learn More](#)



Data Not Linked to You

The following data, which may be collected but is not linked to your identity, may be used for the following purposes:

Analytics



Health & Fitness

Health



Location

Coarse Location



Usage Data

Product Interaction

Other Usage Data



Diagnostics

Crash Data

Other Purposes



Health & Fitness

Health



Location

Coarse Location



Usage Data

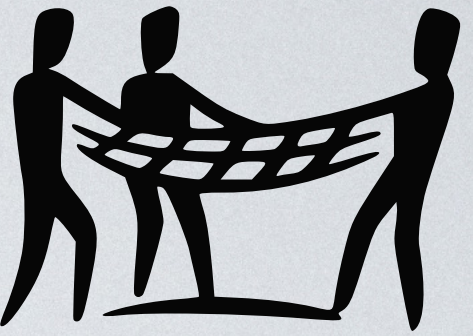
Product Interaction

Other Usage Data

More Results

- Privacy Finder formats were slightly more trusted
- Formats were equally pleasurable

Evidence Suggests That Privacy Policies May Be More Readable When...



- A. They include very clear legal language so they can be interpreted unambiguously.
- B. They are visually appealing.
- C. They follow a standardized format.
- D. They are formalized with math and logic to ensure software follows the advertised policies.

Exercise: Reading Privacy Policies

- Take Netflix's policy as an example (<https://help.netflix.com/legal/privacy>)
- Extract: what information can Netflix share, and with whom?