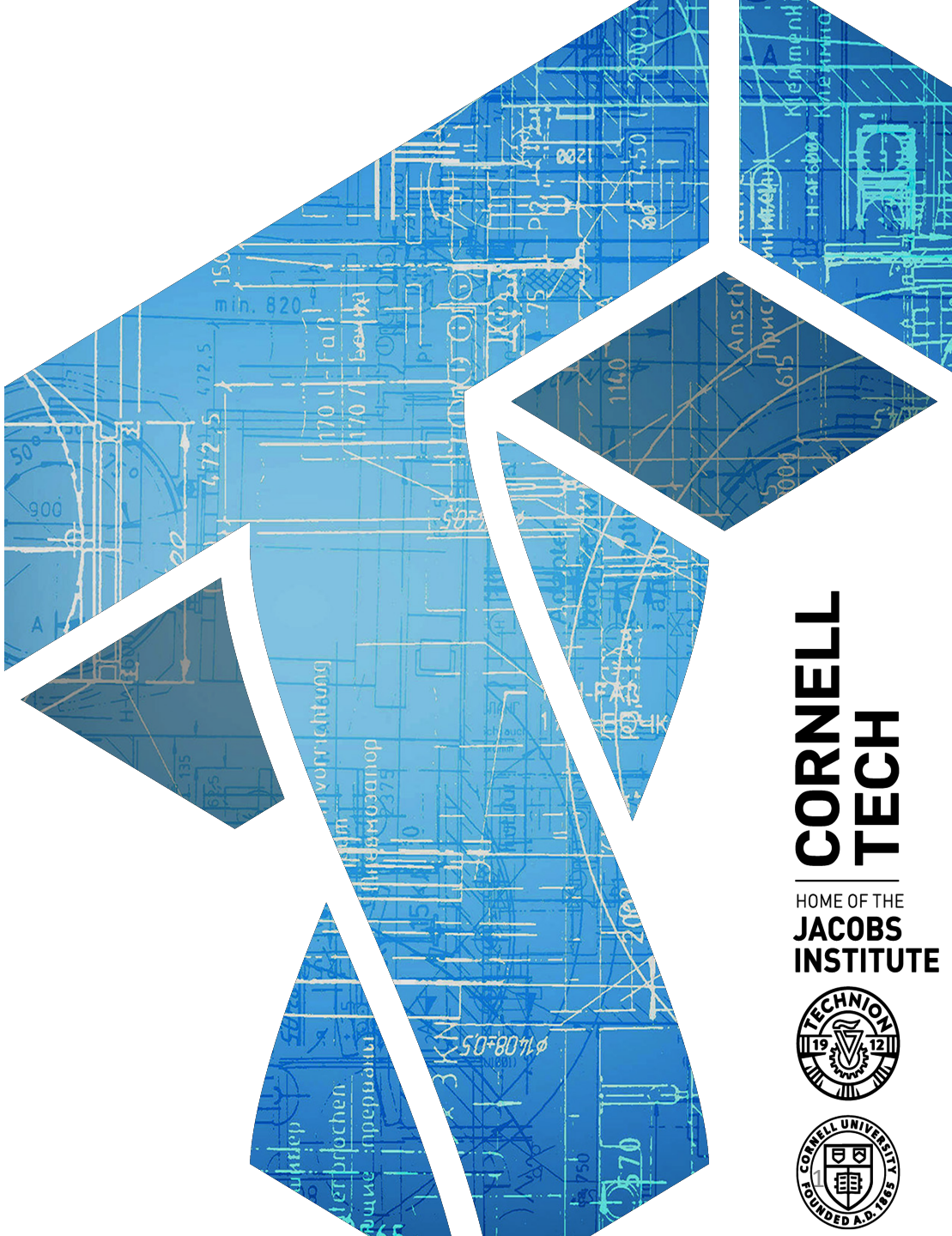
CS 6431: Re-identification



Instructor: Tom Ristenpart

Some slides borrowed from Shmatikov's lectures

https://github.com/tomrist/cs6431-fall2021

CORNELL TECH

HOME OF THE JACOBS INSTITUTE

Reviewing procedure

1. Turn in proposals
2. Bid for proposals (review preferences) by end of day Monday
   - Positively bid (e.g., 20) for proposals you'd like to review
   - Negatively bid (e.g., -20) for those you'd like to avoid reviewing
3. I'll make review assignments after bidding closes, using HotCRP's assignment algorithm, and manual adjustments
4. Have 2 weeks from when I make assignment to do your two reviews

Reviewing logistics

- We are doing unblinded reviewing for simplicity
  - Project authors will know who reviewers are
  - Reviewers will know who project authors are
- Non-standard, and a bad idea in most contexts
  - Double-blind is standard in security
  - Double-blind avoids well-documented (implicit/explicit) bias against authors; protects reviewers from retaliation by upset authors
- Didn't think it would work well here given small class size, and hope that reviewers will provide advice to project authors
  - Ok for authors to answer reviewer questions about projects, document if so

Reviewing goals

- Your project will ***not be*** graded based on reviews
- Reviews will be graded based on:
  - Editorial quality including tone (does it communicate well, and in a constructive manner?)
  - Thoroughness (did you complete the form?)
  - Insightfulness (does it provide a good assessment of project/proposal)
  - Positive contribution (is it likely to help project authors improve project?)

Reviewing proposals

Review template:

- Reviewer expertise
- Paper summary
- Strengths
- Weaknesses
- Comments for author

▶ **Reviewer expertise** *

Please rate how familiar you are with the domain of the project before you started your review. W
reviewers to gain some familiarity, enough to still help provide constructive feedback to authors:
feedback, the motivation for the project, and more.

- ○ **1.** No familiarity
- ○ **2.** Some familiarity
- ○ **3.** Knowledgeable

▶ **Paper summary**

(Text field)

▶ **Strengths**

What are the paper's important strengths? Just a couple sentences, please.

(Text field)

▶ **Weaknesses**

What are the paper's important weaknesses? Just a couple sentences, please.

(Text field)

▶ **Comments for author**

(Text field)

Suggested proposal reviewing approach

- Read over the proposal
- Write down summary of proposal ***in your own words***
- Look up 2-3 pieces of related work
  - References included in proposal & references not included (try Google scholar, general web searches)
  - Read them "lightly". How does proposed project fit with this prior work's claims
  - Go back and read them more thoroughly
- Write comments to author
  - Is the resulting project likely to be interesting to research community? How does it relate to prior work?
  - Are the methods suggested a good fit for questions being pursued? Do you foresee technical problems?

Tone is really important

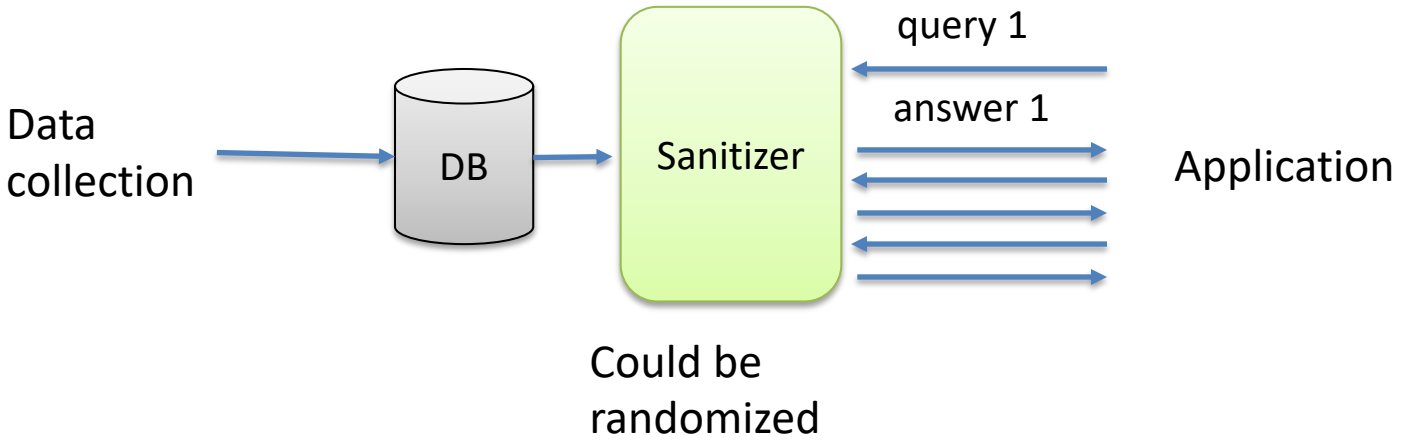
- Any negative about a paper/project can & should be phrased as a suggested improvement
- You are discussing the paper/project, not the ***authors***
  - "The authors seem to believe that X implies Y, which is dumb."
  - "The paper relies on the assumption that X implies Y, and I'm not sure when/where this holds. It would be great to see a discussion of when this assumption holds."
  - "I may be missing something, but I found the paper's statement that 2+2=5 to potentially be an error."
  - "The authors appear to not know basic arithmetic."

Questions?

ML Privacy and Dataset privacy

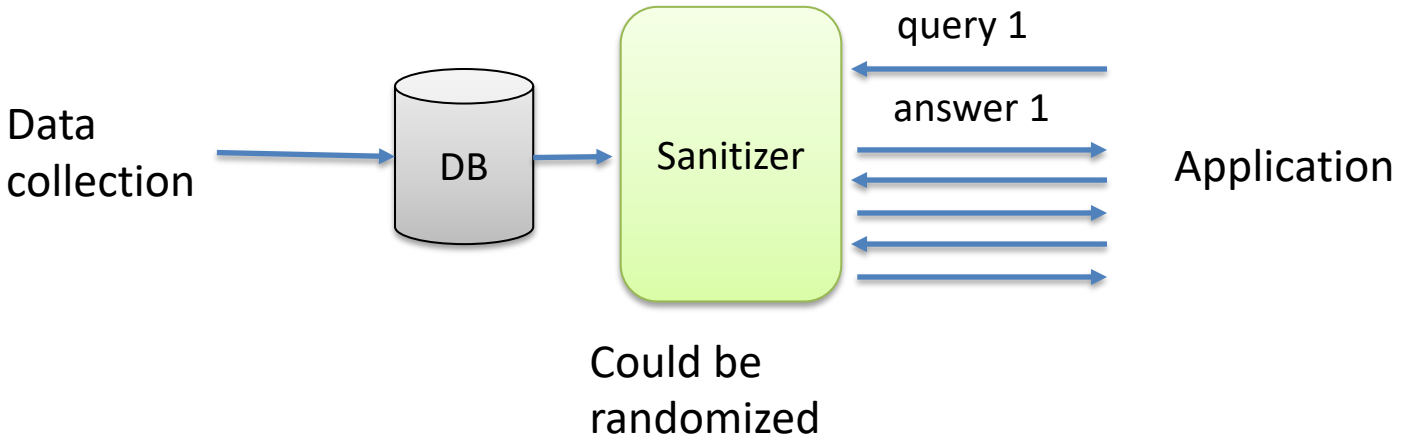
- Privacy of ML models viewed as ***membership privacy***
  - Nuanced relationship to other goals like feature/attribute privacy, training set data privacy, etc.
- Today we rewind a bit, and fill out some of the prior landscape that has impacted our framing of ML privacy
  - Dataset release and anonymization techniques long-studied topic

The sanitization model



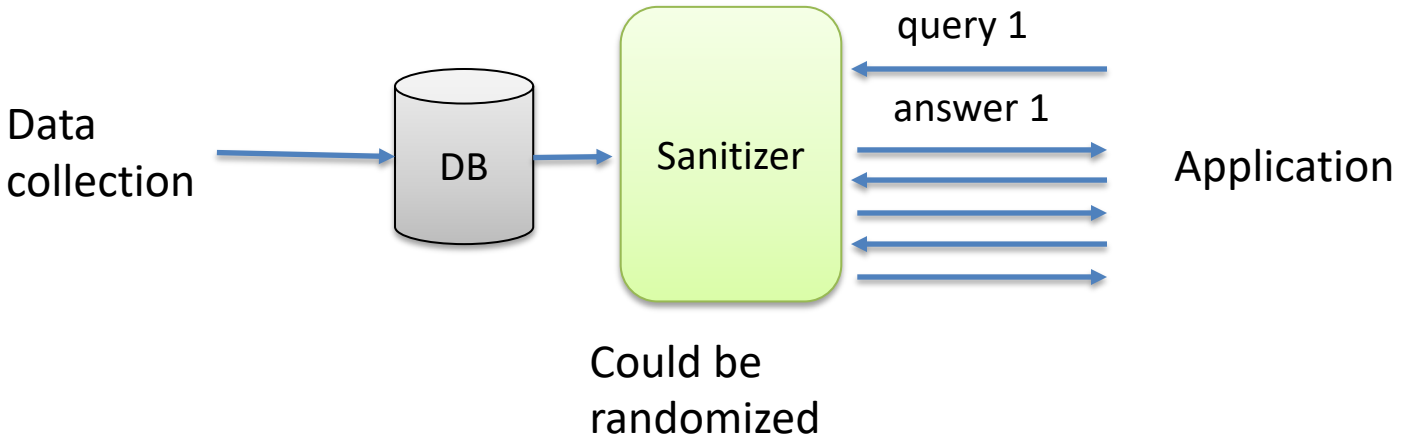
1. Sensitive data collected
2. Sanitize it somehow
3. Release sanitized version all at-once, or answer queries about it

The sanitization model



- ***Input perturbation***: add noise to DB and then release
- ***Summary statistics***: just release aggregate stats (means, variances, marginal totals, regression coefficients)
- ***Output perturbation***: add noise to statistics
- ***Interactive versions of above***: can decide how/if to answer queries

Personally identifiable information (PII)



PII is often discussed. IRBs use this as a legal term. Examples:

- Names, phone numbers, (email) address, social security number

PII has no precise definition

PII can be almost anything, depending on context

Sweeney's 2000 paper

De-identified medical dataset ("microdata")

Name	Zip	Sex	Birthdate	Problem
[redacted]	10001	m	May 2001	Shortness of breath
[redacted]	94517	f	Dec 2010	Obesity
[redacted]	47153	m	June 1980	Chest pains

Public voter registration dataset

Name	Zip	Sex	Birthdate
Adrian	47153	m	June 1980
Parker	10001	m	May 2001
Haven	94517	f	Dec 2010

Quasi-identifier:
Combination of columns that can serve to uniquely identify someone

Linkage re-identification attack:
Use quasi-identifier to do a join between de-identified and public datasets

Birthdate-sex-zipcode uniquely identifies 87% of Americans
(later work suggested 63%)

“Simple Demographics Often Identify People Uniquely”

- Fantastic title: simple, gets across main point
- What about the aesthetics of the paper?
- How does it compare to last paper’s style?

K-anonymity

- Sweeney suggested idea of k-anonymity to prevent re-identification type attacks
- Info for each person contained in dataset cannot be distinguished from at least k-1 others in dataset
- Any quasi-identifier present in dataset must appear in at least k records

- Lead to a ton of research about how to achieve this
- And how to break it

K-anonymity: techniques

- Generalization
  - Replace quasi-identifiers with less specific, but semantically equivalent ones
  - E.g.: 10035 -> 100**
- Suppression: just remove value completely

K-anonymity: limitations

- May not hide membership
  - Suppose 10 people with particular quasi-identifier (k=10)
  - Suppose exactly 10 people in population have this quasi-identifier
- May not (completely) hide sensitive features
  - Homogeneity attack
  - Background knowledge attack
  - Complementary dataset attack (k-anonymity doesn't compose)
- May not even prevent re-identification

K-anonymity: limitations

Homogeneity attack

Bob	
Zipcode	***Age***
47678	27

Background knowledge attack

Carl	
Zipcode	***Age***
47673	36

A 3-anonymous patient table

Zipcode	Age	Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
4790*	⩾40	Flu
4790*	⩾40	Heart Disease
4790*	⩾40	Cancer
476**	3*	Heart Disease
476**	3*	Cancer
476**	3*	Cancer

K-anonymity: limitations

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

(a)

K=4 anonymized

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*	Viral Infection
12	130**	≥35	*	Viral Infection

(b)

K=6 anonymized

Sensitive condition untouched

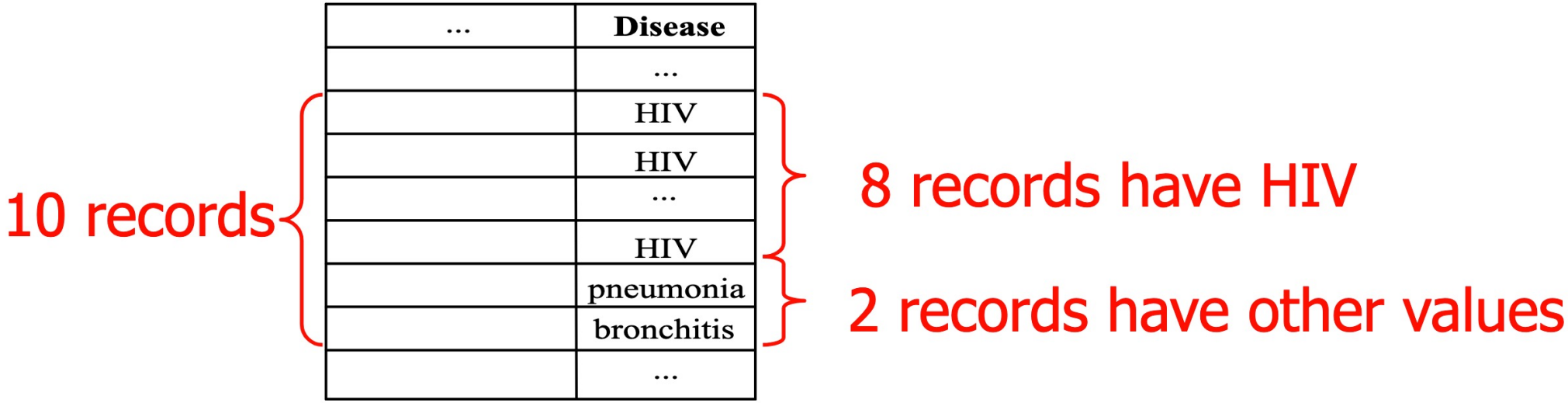
Alice: 28 yo, zip 13012, visited both hospitals

What condition does she have?

L-diversity

[Machanavajjhala, Gehrke, Kifer 2007]

- Each equivalence class (each pseudo-identifier) has at least L distinct values



- Variety of limitations.
- t-closeness, probabilistic L-diversity, entropy L-diversity, …

Syntactic approaches all have pitfalls

- How do we know at time of data release what are reasonable quasi-identifiers?
- Ignores semantics of sensitive attributes

Netflix dataset reconstruction attack [Naranyan, Shmatikov 2008]

- Netflix released de-identified dataset of customer ratings and their dates
  - Almost 500,000 customers, >100 million ratings
- Show new algorithms for re-identification
  - Inputs are “anonymized” dataset, auxiliary data about a person
  - Output best seeming matched row in dataset
  - Intuitively: finds good quasi-identifiers using some heuristics
- Why does it work?
  - High dimensional datasets mean nearest neighbors far away
- Reidentify two people using IMDB as public data

Membership inference vs. attribute recovery?

- All this sounds a lot like our ML discussions from Tueday
- Re-identification versus membership inference
- Model inversion versus attribute recovery

- Dataset is a model of data!
  - Really same stuff going on, different particulars

US Census Data

- Constitutionally mandated every-10 year release of data
- Releases demographic microdata used by many researchers
- Constitutional requirement for confidentiality
- Reconstruction / re-identification attacks considered a risk
- 2020 data will use differential privacy
  - Privacy