COS 125, Summer 2025

Privacy:

Time to Admit, We Don't Understand It

Some Distinctions

01 Confidential

02 Private

01 Autonomy

02 Personal Identity

Handout with all definitions for today

First Set of Distinctions

a. Confidentiality

An ethical duty that prevents certain people from sharing your information with others, unless they have your permission or a valid reason.

b. Privacy

Either, 1) A right to freedom from intrusion into your personal matters, and information about you that helps you to define who you are. Or, 2) An ethical duty of others to refrain from seeking certain personal information about you, or spreading it once they have it.

Second Set of Distinctions

a. Autonomy

A person's right of self-government, or the capacity to self-govern. A person's ability to self-govern is ordinarily thought to involve them choosing which actions to take or rules to follow. Some level of outside interference with that choice, or attempts to influence that choice, may diminish autonomy.

b. Personal Identity

The attributes that make you the same you over time, usually those properties to which you feel a special sense of attachment (as opposed to a government ID number, or our living cells, which "identify" us over time). My personal identity is those things I take to "define me as a person" or "make me the person I am."

K-Anonymity

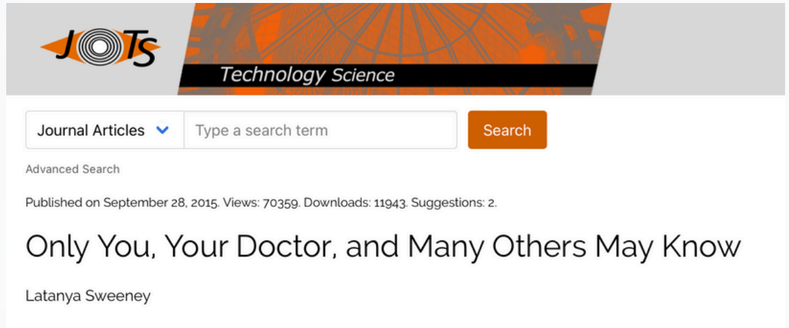
Prof. Latanya Sweeney
Harvard University

In the event of a re-identification, an adversary should be able to use "insensitive" data points to narrow the data set to only k people. The higher the more private.

So, for instance, a 2-anonymous dataset would allow an adversary to identify only 2 people in the data set who might share your "insensitive" data.

But what is "sensitive" and what is "insensitive" data?

- *We would need a theory to tell us that!* (Stay tuned.)



De-Anonymization

Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)

Arvind Narayanan and Vitaly Shmatikov

The University of Texas at Austin

Abstract

We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge.

We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix, the world's largest online movie rental service. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.

K-Differential Privacy

Cynthia Dwork's "Differential Privacy"

That we should be able to learn no more about a person from an analysis if the person's data was included in the analysis than if it was not included.

- Kearns and Roth, Chapter 1

What does “Autonomy” Protect?

Anything that touches upon your choice-making:

- Medical conditions, esp. if lose choices (pre-existing conditions)?
- Reproductive choice?
- Social and sexual identity?

Nothing that you have chosen to reveal:

- Nothing you've told others?
- Nothing others have asked of you?
- Nothing covered by a signed consent form?
    - The Henrietta Lacks Problem

What alternative theories could there be?

Preview

Alternative Theories

Number 1

Democratic Privacy

Number 2

"Privacy as Power"
by Carissa Veliz,
Oxford

Number3

"Privacy as Property"

Questioning the Alternatives

Number 1

Democratic Privacy

Protects only information about political participation. *Maybe* art, literature, science, social change

Number 2

"Privacy as Power"

What if we don't GAF?

What if no-one "uses" our data to control us?

Number3

"Privacy as Property"

Too strong, too weak

Can I "own" info that is about you, too?

Problems with GDPR

a. Too Strong

Protects even non-private information

b. Too Weak

Protects information only if the data subject asserts a right

Reasonable Expectation of Privacy (in US Law)

01 **Place Based**

The Telephone Booth Case

02 **Personality Based**

Louis Brandeis's 1890 Paper: "inviolate personality"

03 **Quality Based**

Rare in US Courts (as is the "personality" standard)

04 **Quantity Based**

GPS Tracker Case

A Test Case

Samir and Samuel are married, which everyone knows. They tell their neighbors over dinner that they have a very normal sex life, and are intimate with each other about as much as the average person. One of the neighbors at that dinner tells Samir's boss at a restaurant what Samir has said. Another of the neighbors tells Samuel's boss at a government agency.

Has either of these neighbors violated confidentiality? Write one sentence explaining your answer.

Have they violated privacy? Use a theory of privacy you learned to write one sentence about the case of Samir and one about the case of Samuel, explaining your answers.

COS 125, Summer 2025

Privacy:

Time to Admit, We Don't Understand It

Princeton University

Thank You

From Steven Kelts

Question and Answer...