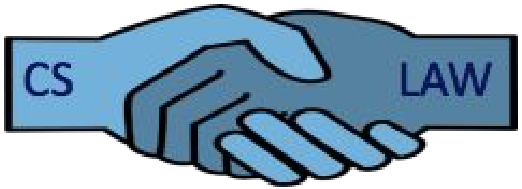
Bridging the Gap between Computer Science and Legal Approaches to Privacy



Alexandra Wood
Berkman Klein Center for Internet & Society at Harvard University

Privacy Semester Planning Workshop
May 24, 2017

An Interdisciplinary Collaboration

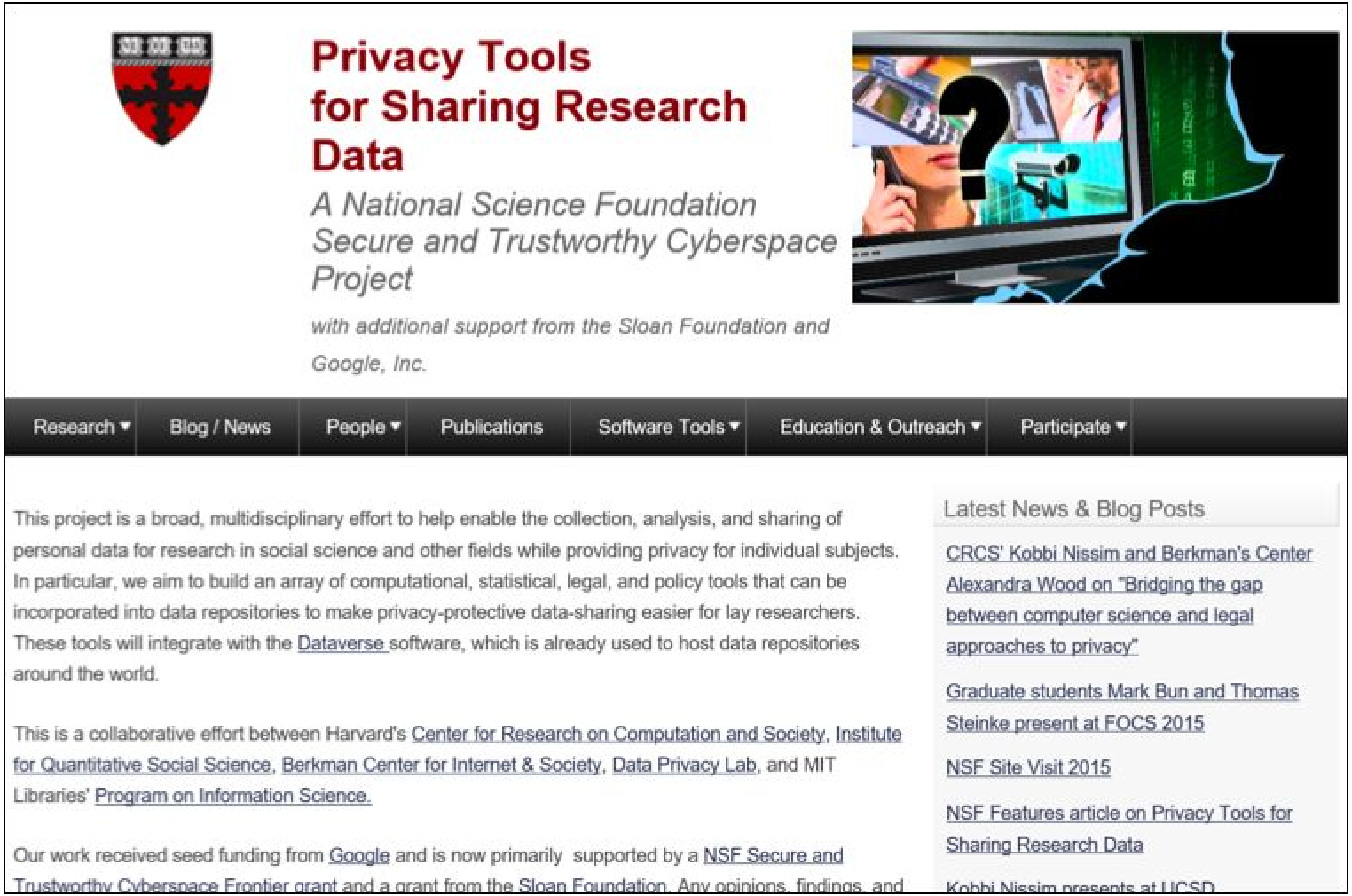
This work is the product of an *interdisciplinary working group* bringing together computer scientists and legal scholars

CRCS Center for Research on Computation and Society
at Harvard John A. Paulson School of Engineering and Applied Sciences

Kobbi Nissim, Aaron Bembenek, Mark Bun, Marco Gaboardi, Thomas Steinke, Salil Vadhan

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

Alexandra Wood, David O’Brien, Urs Gasser



Privacy Tools for Sharing Research Data

A National Science Foundation Secure and Trustworthy Cyberspace Project

with additional support from the Sloan Foundation and Google, Inc.

This project is a broad, multidisciplinary effort to help enable the collection, analysis, and sharing of personal data for research in social science and other fields while providing privacy for individual subjects. In particular, we aim to build an array of computational, statistical, legal, and policy tools that can be incorporated into data repositories to make privacy-protective data-sharing easier for lay researchers. These tools will integrate with the Dataverse software, which is already used to host data repositories around the world.

This is a collaborative effort between Harvard's Center for Research on Computation and Society, Institute for Quantitative Social Science, Berkman Center for Internet & Society, Data Privacy Lab, and MIT Libraries' Program on Information Science.

Our work received seed funding from Google and is now primarily supported by a NSF Secure and Trustworthy Cyberspace Frontier grant and a grant from the Sloan Foundation. Any opinions, findings, and

Latest News & Blog Posts

CRCS' Kobbi Nissim and Berkman's Center Alexandra Wood on "Bridging the gap between computer science and legal approaches to privacy"

Graduate students Mark Bun and Thomas Steinke present at FOCS 2015

NSF Site Visit 2015

NSF Features article on Privacy Tools for Sharing Research Data

Kobbi Nissim presents at UCSD

Motivation

Formal privacy models like **differential privacy** offer a solution for providing wide access to statistical information with guarantees that individual-level information will not be leaked inadvertently or due to an attack.

- Formal mathematical privacy concept that addresses weaknesses of traditional schemes (and more).
- Supported by a rich theoretical literature and now in initial stages of implementation and testing by industry and statistical agencies.

Motivation

Formal privacy models like **differential privacy** offer a solution for providing wide access to statistical information with guarantees that individual-level information will not be leaked inadvertently or due to an attack.

- Formal mathematical privacy concept that addresses weaknesses of traditional schemes (and more).
- Supported by a rich theoretical literature and now in initial stages of implementation and testing by industry and statistical agencies.

🚫 However, these tools cannot be used to share sensitive data with the general public unless they satisfy legal standards with some certainty.

Challenges

Demonstrating that formal privacy models satisfy applicable legal requirements is challenging due to the conceptual gaps between legal and technical approaches to defining privacy.

Notably, information privacy laws are generally:

- **context-specific**,
- **subject to interpretation**,
- allow for some **degree of flexibility**, and
- rely on **traditional, often heuristic, conceptions of privacy**,

which creates uncertainty for the implementation of more formal approaches.

Example Points of Mismatch

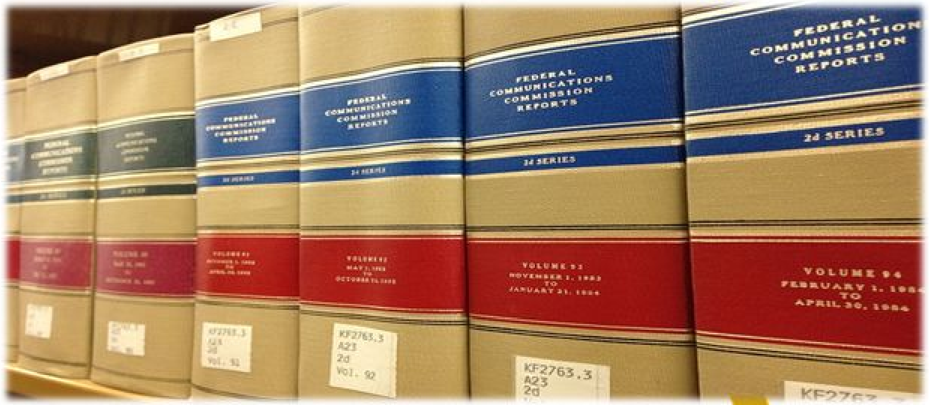
FERPA

- Applies to highly sector- and context-specific settings
- Contemplates a small set of specific types of privacy attacks
- Protects a small set of information (non-directory PII)
- Refers to the obvious extreme cases, not to more difficult “gray areas”
- Applies to releases of microdata and tabulations
- Imprecise, not rigorous/formal from a technical standpoint

Differential Privacy

- Offers general privacy protection
- Addresses a very large class of potential data misuses
- Protects any information contributed by an individual
- Applies to all analyses, does not leave “gray areas”
- Not limited to releases of microdata and tabulations
- A mathematically rigorous definition

Is it possible to bridge these very different languages?



$M: X^n \to T$ satisfies ϵ-differential privacy if

$\forall x, x' \in X^n$ s.t. $dist_H(x, x') = 1\ \forall S \subseteq T$,

$$\Pr_{\mathrm{M}}[M(x) \in S] \leq e^{\epsilon} \Pr_{\mathrm{M}}[M(x') \in S].$$

Approach

We seek a methodology for rigorously arguing that a technological privacy solution satisfies the requirements of a particular law.

Our approach has two components:

1. Extraction of a formal mathematical requirement of privacy based on a legal standard found in an information privacy law, and

2. Construction of a rigorous mathematical proof for establishing that a technological privacy solution satisfies the mathematical requirement derived from the law.

Example: FERPA and Differential Privacy

We illustrate this application of this approach with an example bridging between:

- a specific legal standard (**FERPA**) and
- a specific technical standard (**differential privacy**).

Introduction to FERPA

The Family Educational Rights and Privacy Act of 1974 requires the protection of personal information maintained in education records.

FERPA distinguishes between two categories of information:

- **Directory information**: information from education records that can be made public; is designated by each school (34 C.F.R. 99.37).
- **Non-directory personally identifiable information**: information that can only be disclosed without consent under certain exceptions (34 C.F.R. 99.31).

Personally identifiable information: “information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty” (34 C.F.R. 99.3)

Extracting a Formal Definition from FERPA

Our goal: To extract a formal model of the Department of Education's privacy desiderata for FERPA.

Our formal model is in the form of a game-based privacy definition, following a computer-science paradigm used for formalizing and analyzing security and privacy.

Goals of game-based definition:

1. Provides a concise and fairly intuitive abstraction of the requirements in FERPA.
2. Enables us to prove that if a formal model, such as differential privacy, satisfies the game-based definition, then we have a strong argument that it satisfies the requirements of FERPA.

Although FERPA is not written with a privacy game framework in mind, we claim (and demonstrate) that it is possible to extract a game that is based on its requirements.

Extracting a Formal Definition from FERPA

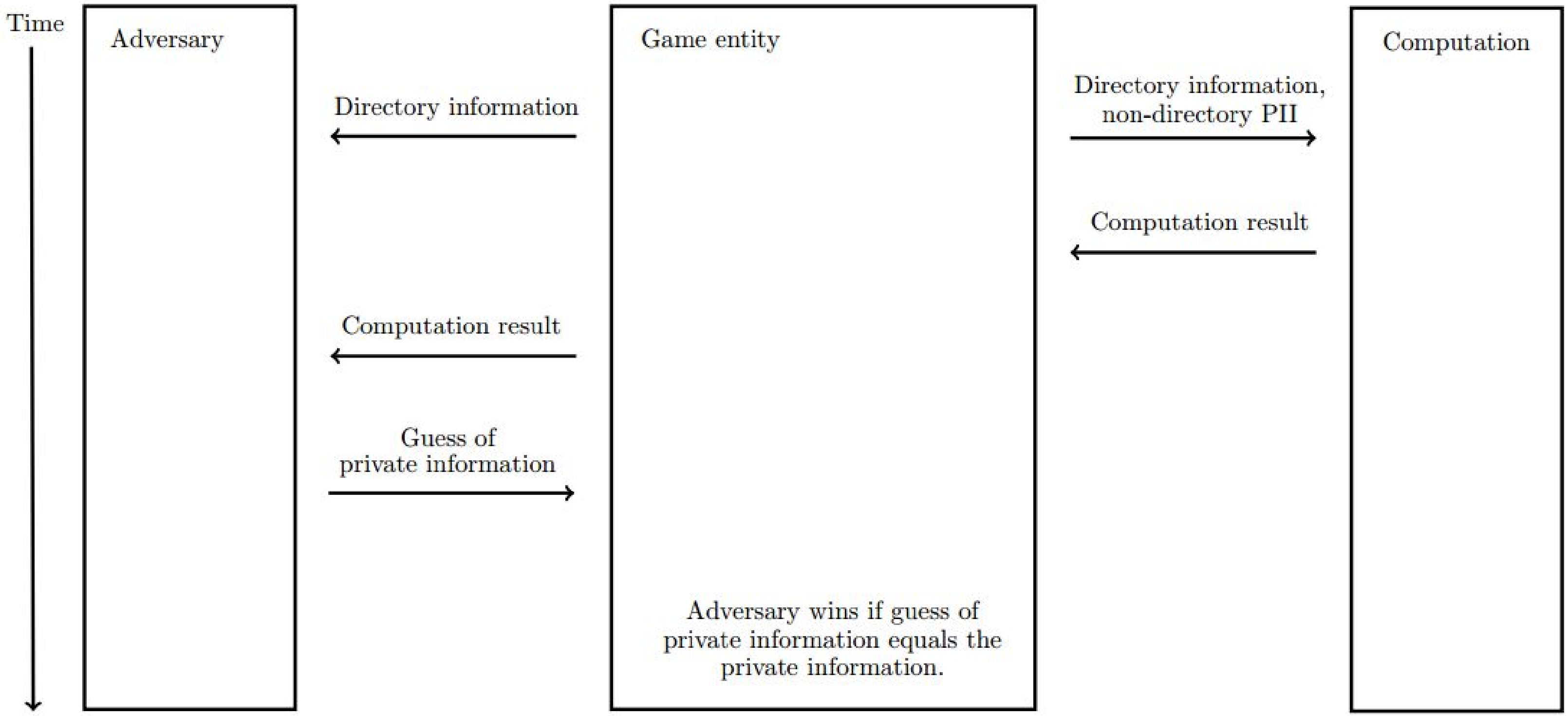
FERPA allows the release of **de-identified information** from education records

De-identification can be thought of in terms of a (statistical) **computation**; e.g., a legal provision requiring the removal of identifying attributes can be seen as requiring a computation to redact those identifiers from the input data.

This framing is useful for modeling a law's requirements using the formal language used in computer science. This modeling allows us to extract a mathematical definition for determining whether a computation meets the FERPA privacy standard.

But how do we know whether a given computation provides a sufficient level of privacy protection to meet the requirements of a statute or regulation?

Components of a FERPA Privacy Game


Time
Adversary
Game entity
Computation
Directory information
Directory information, non-directory PII
Computation result
Computation result
Guess of private information
Adversary wins if guess of private information equals the private information.

Modeling FERPA: Directory Information

The regulatory language is **ambiguous**, so we interpret the language as conservatively as reasonably possible. In other words, where there is ambiguity, we err on the side that is most beneficial for the adversary.

For example, the definition of **directory information** is ambiguous (e.g., the definition varies between schools).

- We could make assumptions in defining directory information in our model. However, new interpretations could call these assumptions into question.
- Instead, **we let the attacker to choose** what constitutes directory information.

Modeling FERPA: The Adversary

Personally identifiable information: “information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty.”

This is FERPA’s implicit adversary. Key points from guidance:

- We should not assume anything about the skill level of the adversary.
- Standard is based on the knowledge of a member of the school community, which is stronger than one based on the knowledge of any reasonable person.
- The adversary can have both high-level knowledge (e.g., knowledge of general demographics of school) and “insider” knowledge about specific individuals in local community.

Modeling FERPA: The Adversary’s Knowledge

The adversary clearly has (potentially a lot of) knowledge, but by definition does not have “personal knowledge of the relevant circumstances.”

In our model, the adversary has access to any information that is publicly available, but has some uncertainty about private student information.

We model the adversary’s knowledge via probability distributions. Adversary associates with each student a probability distribution that represents her knowledge about the private information of that student. We allow the adversary to choose these statistics.

Example: If Alice comes from a school where 50% of the students failed the state math proficiency exam, then adversary might associate with Alice a distribution that has her failing the exam with a probability of 0.5.

Proving Differential Privacy Satisfies FERPA

Developing a formal definition of privacy protection based on the requirements of FERPA allows us to reason, with high confidence, about whether the use of a privacy technology satisfies FERPA.

For instance, we can prove mathematically that any computation that is differentially private meets this definition, and (since the requirements of this definition are likely stricter than that of FERPA) thus satisfies the privacy requirements of FERPA.

Conclusion

An illustration of an approach to answering a broader question: As a new technology emerges, can we claim it satisfies existing regulations?

Satisfying legal standards is crucial for DP to be used in practice with sensitive personal information -- but this should not be taken to mean we think current legal standards are adequate.

Demonstrating how to make a combined mathematical-legal formal claim that differential privacy satisfies FERPA:

- Extracting a formal, conservative attacker model for the regulation
- Using an argument that is both legally and mathematically rigorous
- Leveraging mathematical tools in order to deal with inherent ambiguity in legal standards