Integrating Differential Privacy and Contextual Integrity

Sebastian Benthall, New York University Law School

Rachel Cummings, Columbia Engineering

PEPR ‘22
USENIX

Differential Privacy (DP)

A parameterized notion of algorithmic privacy for databases.

It bounds the impact of any one data entry on the result of analysis of the database.

$$\Pr[M(X) \in \mathcal{S}] \leq e^{\epsilon} \Pr[M(X) \in \mathcal{S}] + \delta$$

The parameters (here ϵ, δ) encapsulate trade-offs between privacy and accuracy.

DP provides no guidance about the choice of parameters.

We see this as a challenge for practitioners, and one emblematic of the state of privacy enhancing technologies (PETs) more generally.

Contextual Integrity (CI) - Contexts

A social theory of privacy for interdisciplinary research. (Nissenbaum, 2009)

(a) Privacy is *appropriate information flow:*

(+ appropriate flow) and (- inappropriate flow)

(b) *Appropriateness* refers to *information norms* that inhere in a social context, e.g.: health care, education, etc.

(c) Social contexts have a *purpose*, defined *roles* that people fill, and relevant information *attributes*

Contextual Integrity (CI) - Norms

(a) *Information norms* are parameterized in terms of:

Sender, Receiver, Subject, Attribute, Transmission Principle

Example: *Radiologist*, *General Doctor*, *Patient*, *X-rays*, *Confidentiality*

(b) Information norms are legitimized by how they balance contextual purposes (e.g. a healthy society) with individual ends (doctors limiting liability)

CI is used to analyze privacy norms in legal and ethical analysis, as well as technical design.

Why integrate DP and CI?

CI is a rubric for collecting contextual information that is needed to make normative decisions about information flow.

This information can then be used to tune DP parameters:

Tune parameters to optimize *appropriate information flow* given contextual purposes.

We can also contribute back to CI refinements and insights from PET practice.

Information properties: modulating information flow. I.e “with Gaussian noise”.

Contributions: Privacy Theory

- **New formalization of CI**. Based on a systematic review of previous computer science implementations of CI (Benthall et al., 2017) and our use case of tuning and communicating PETs parameters.
- **Integrated rubric for privacy analysis**.
  Normative and Descriptive; Contexts and Flow.

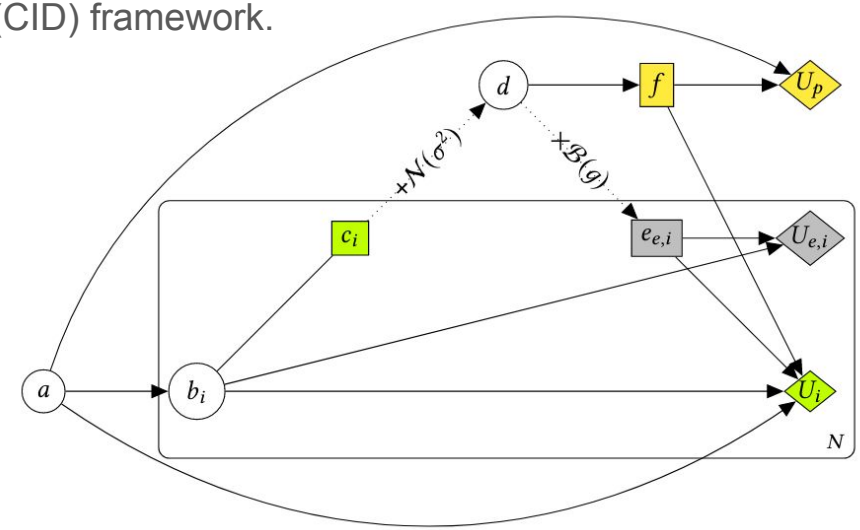
Transmission Properties	Transmission Principles	Situation	Sphere
Flow	Flow	Context	Context
Descriptive	Normative	Descriptive	Normative
Flow with no PET	Consent	With N population	Nationwide
With Gaussian noise	Reciprocity	Bounds on adversary	Interpersonal
With Laplace noise	Disclosure	With X auxiliary information	Health
Encrypted	With a warrant		Financial
Securely Aggregated	Minimized		Educational

Table 1. Elements of continuous information design combining CI and DP.

Contributions: Parameter Tuning Procedure

- **Privacy Modeling**. Components of integrated privacy rubric combine into contextualized model of information flows and threats.
  - Potential PETs and parameters are represented in the model
  - Modeling built on Causal Influence Diagram (CID) framework.

- **Parameter tuning as optimizing appropriate information flow**. Contextualized model operationalizes purposes and appropriateness as equations for the optimization problem.



Case study: U.S. Census

Purpose:

- Allocate seats for Congress.
- Social science research

Roles:

- U.S. Census Bureau (sender)
- U.S. residents (subject)
- Researchers (receiver)
- General public (receiver)

Attributes:

- PL 94-171 (redistricting dataset);
- Public-Use Microdata Sample
- Restricted-Use Data: detailed information on U.S. persons

Information Norms:

- Redistricting dataset - produced from Decennial Census survey data
  - with PET use.
- Public-Use Microdata Sample - produced from the American Communities Survey data
  - with PET use.
- Restricted-Use Data: produced from the American Communities Survey data
  - with PET use.
  - Available only to “qualified researchers with approved projects”
  - Access in secure Federal Statistical Research Data Centers (RDC) with no data export.

Other use cases: federated learning with smartphone data, interstate medical data sharing, …

Thank you! Contact: spb413@nyu.edu ; rac2239@columbia.edu