CONSTRAINT-BASED DIFFERENTIAL PRIVACY

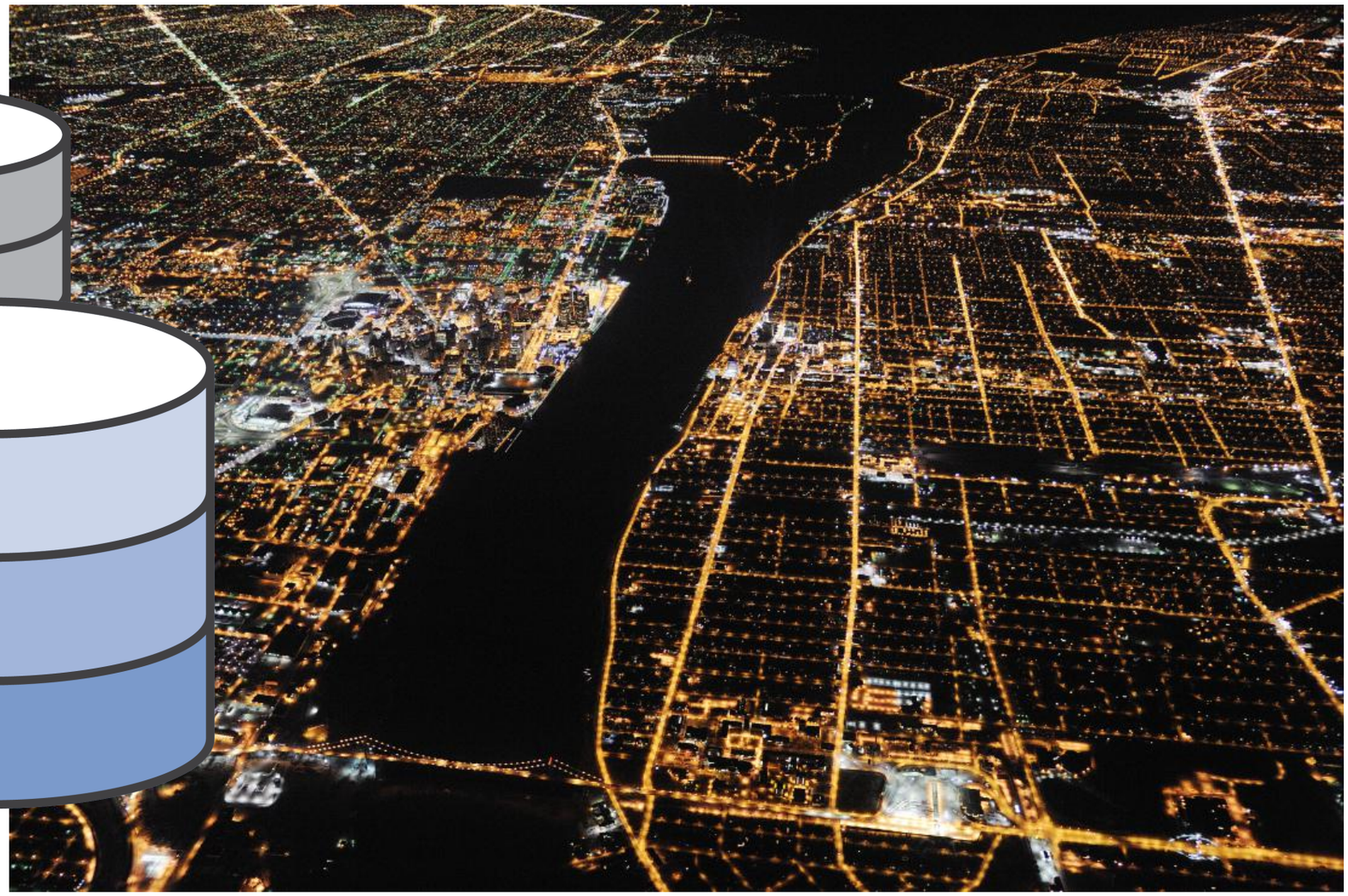
Releasing Optimal Power Flow Benchmarks Privately

Ferdinando Fioretto & Pascal Van Hentenryck

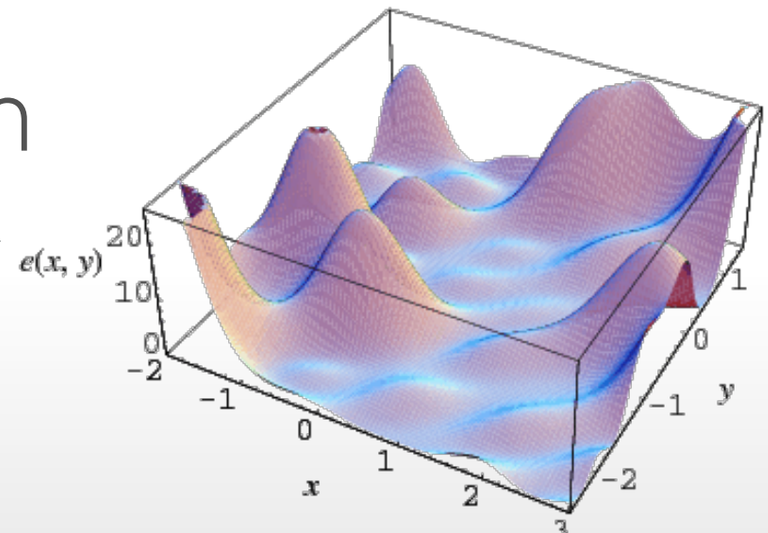
University of Michigan

CPAIOR 2018

Customers Loads

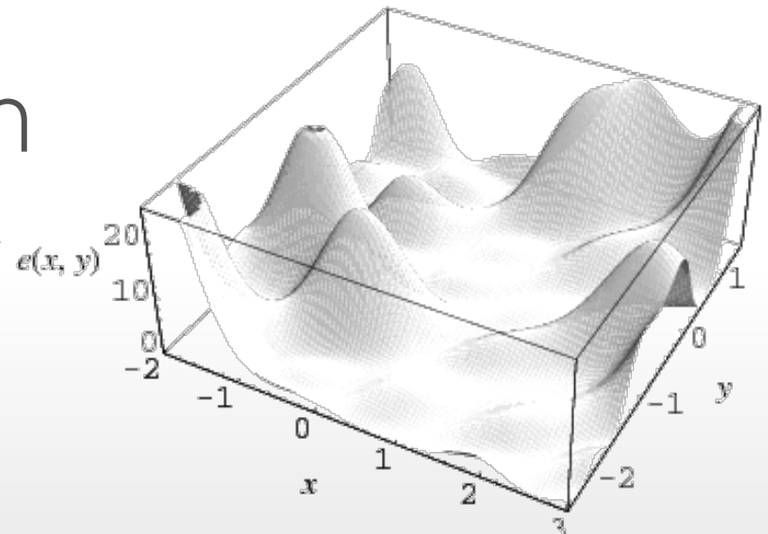


optimization



Customers Loads

optimization



Content

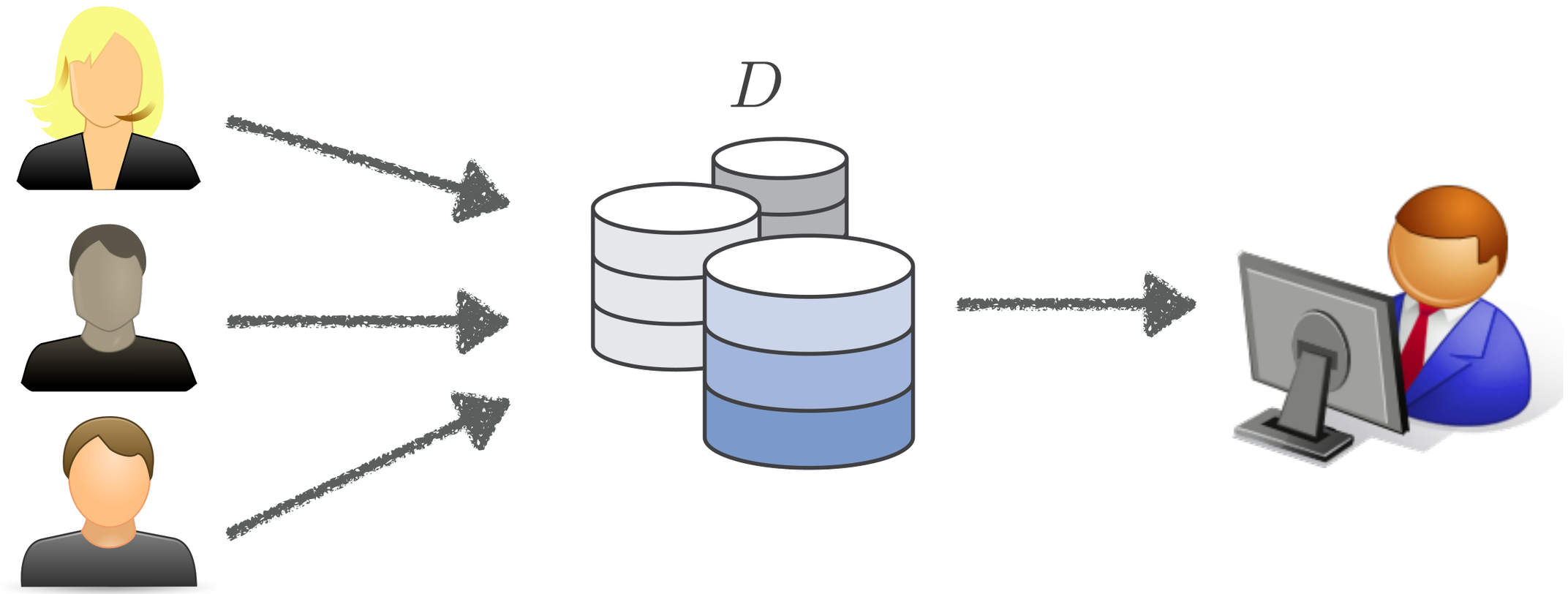
- **Private Data Release and Differential Privacy**
- Optimal Power Flow Problem
- The CBDP Mechanism
- Experimental Analysis on Private OPF

Private data-release

contributors

data curator

data analyst

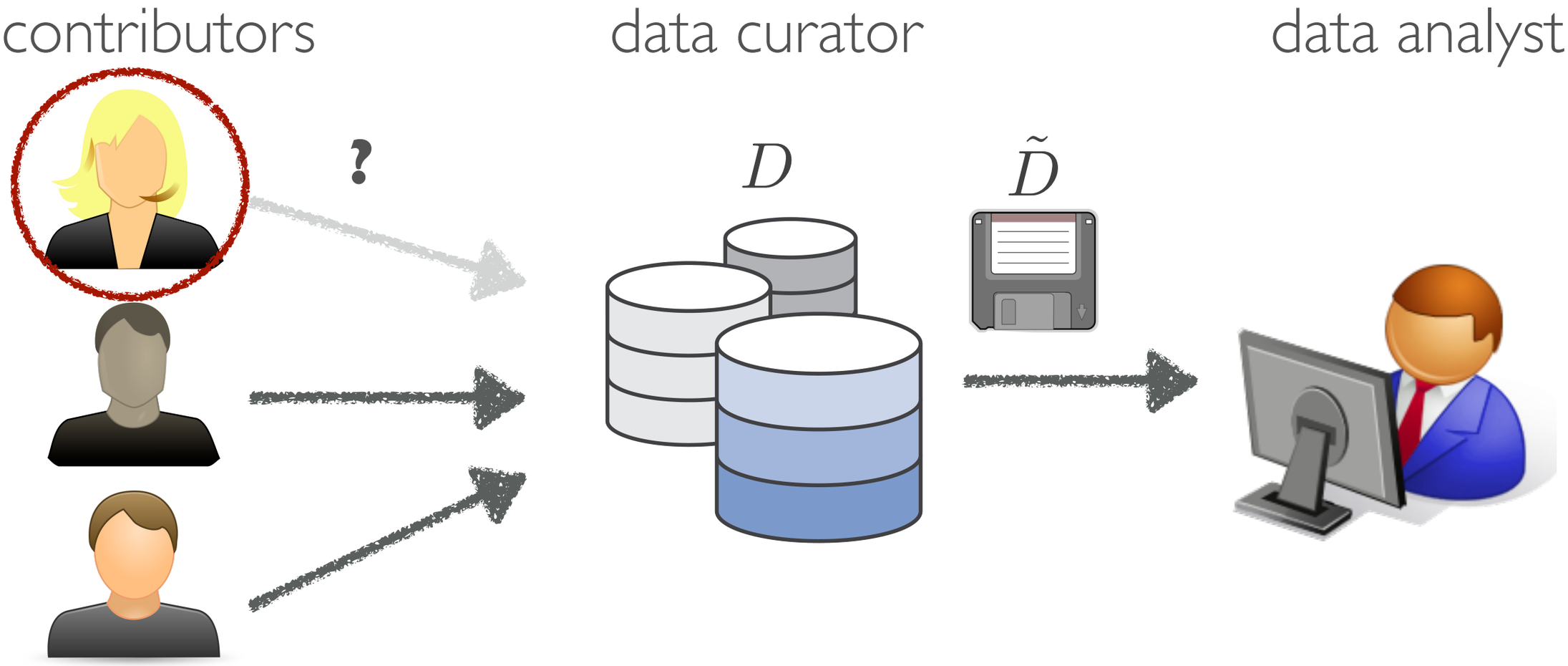


Goal 1: Protect the privacy of the contributors

Goal 2: The data analyst receives *useful* data

Differential Privacy

(Informal)

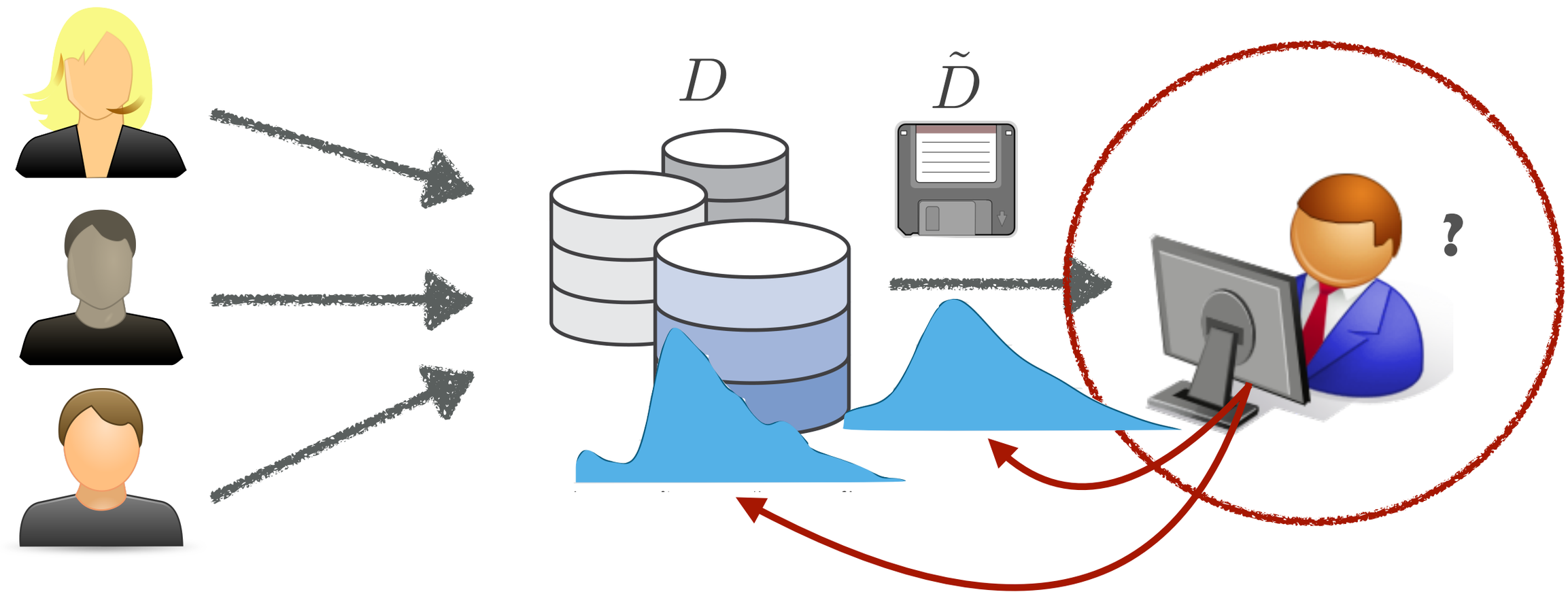


Contributor: Small participation risk (privacy loss)

Data analyst: Analysis on original and modified data are very similar (data distributions)

Differential Privacy

(Informal)



Contributor: Small participation risk (privacy loss)

Data analyst: Analysis on original and modified data are very similar (data distributions)

Differential Privacy

- *Two datasets D_1, D_2 are said neighbors ($D_1 \sim_\alpha D_2$) if they differ by α in at most one tuple*

D_1

name	load
Alice	21.2
Bob	30.1
Carl	17.4
Diana	20.5
…	…

D_2

name	load
Alice	21.2
Bob	30.1
Carl	27.4
Diana	20.5
…	…

$$\alpha = 10$$

Differential Privacy

- *In statistical databases, often, the l -hamming distance is used:*

$$D_1 \sim D_2 \Leftrightarrow \|D_1 - D_2\|_1 \leq 1$$

D_1

name	age	gender
Alice	21	F
Bob	39	M
Carl	17	M
Diana	25	F
…	…	….

D_2

name	age	gender
Alice	21	F
Bob	39	M
Carl	17	M
Diana	25	F
Emily	26	F
…	…	….

[Dwork:06]

Differential Privacy

- *Two datasets D_1, D_2 are said neighbors ($D_1 \sim_\alpha D_2$) if they differ by α in at most one tuple*

A randomized mechanism $\mathcal{M} : \mathscr{D} \to \mathscr{R}$ is ϵ-differentially private if, for any pair $D_1, D_2 \in \mathscr{D}$ of neighboring datasets and any output $O \in \mathscr{R}$:

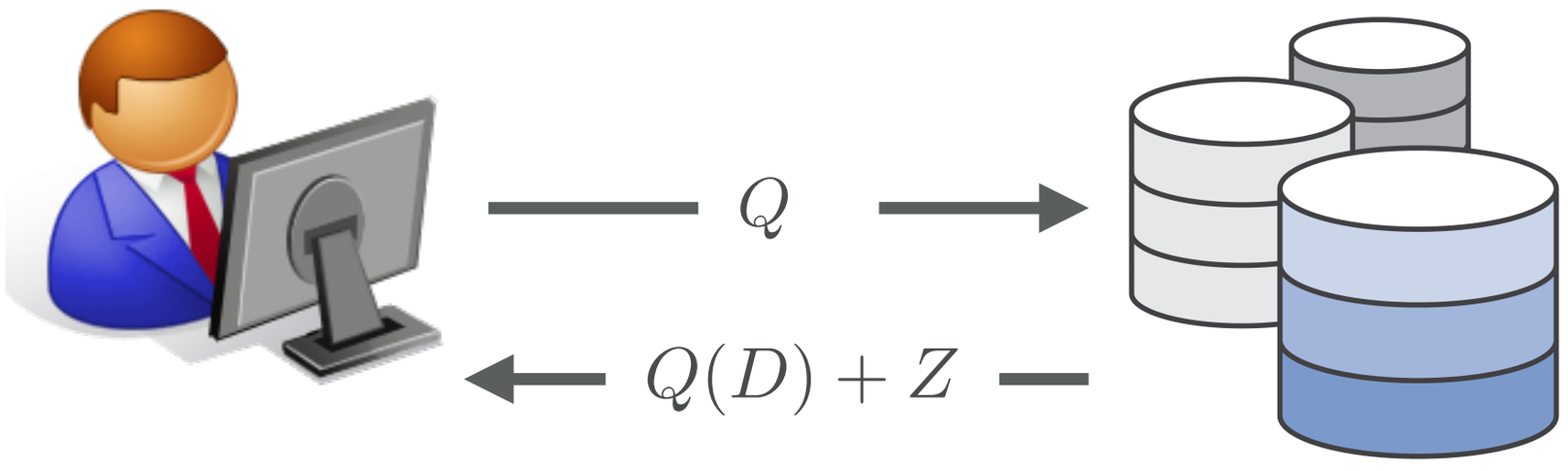
$$\frac{\Pr[\mathcal{M}(D_1) = O]}{\Pr[\mathcal{M}(D_2) = O]} \leq \exp(\epsilon), \quad (\epsilon > 0)$$

- *The risk of a user to join the dataset or to change her value by at most α is bounded (by ε)*

How Can we Achieve DP?

[Dwork:06]

The Laplace Mechanism



$Q(D)$ = true answer

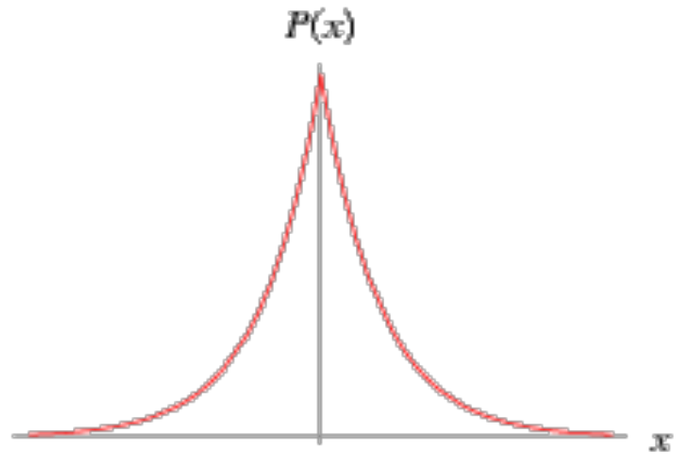
$Z \sim \text{Laplace}(\Delta_Q/\epsilon)$

Theorem (Laplace Mechanism)

Let $Q : \mathscr{D} \to \mathscr{R}$ be a numerical query. The Laplace mechanism $\mathcal{M}(D; Q, \epsilon) = Q(D) + Z$, where $Z \sim Lap(\frac{\Delta_Q}{\epsilon})$ achieves ϵ-differentially privacy.

$$b = (\Delta_Q/\epsilon)$$

How much does the output of Q changes if we add/remove one tuple (or α) from D ?



$$f(x \mid \mu = 0, b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$$

PDF

Differential Privacy

Notable Properties

- *No linkage attack: Adversary knows arbitrary auxiliary information*

- Composability: *If M_1 enjoys ε_1-differential privacy and M_2 enjoys ε_2 differential privacy, then, their composition $M_1(D)$, $M_2(D)$ enjoys $\varepsilon_1+\varepsilon_2$-differential privacy*

- *Post-Processing immunity: If M enjoys ε-differential privacy and g is an arbitrary mapping, $g \circ M$ is ε-differential private*

Content

- Private Data Release and Differential Privacy
- **Optimal Power Flow Problem**
- The CBDP Mechanism
- Experimental Analysis on Private OPF

Optimal Power Flow (OPF)

The AC Optimal Power Flow Problem (AC-OPF)

variables: $S_i^g, V_i \;\; \forall i \in N, \;\; S_{ij} \;\; \forall (i,j) \in E \cup E^R$

minimize:
$$\sum_{i \in N} \boldsymbol{c}_{2i}(\Re(S_i^g))^2 + \boldsymbol{c}_{1i}\Re(S_i^g) + \boldsymbol{c}_{0i}$$

→ generators' cost

subject to:

$$\angle V_{\boldsymbol{r}} = 0, \;\; r \in N$$

$$\boldsymbol{v}_i^{\boldsymbol{l}} \leqslant |V_i| \leqslant \boldsymbol{v}_i^{\boldsymbol{u}} \;\; \forall i \in N$$

$$-\boldsymbol{\theta}_{ij}^{\boldsymbol{\Delta}} \leqslant \angle(V_i V_j^*) \leqslant \boldsymbol{\theta}_{ij}^{\boldsymbol{\Delta}} \;\; \forall (i,j) \in E$$

→ engineering limits

$$\boldsymbol{S}_i^{\boldsymbol{gl}} \leqslant S_i^g \leqslant \boldsymbol{S}_i^{\boldsymbol{gu}} \;\; \forall i \in N$$

$$|S_{ij}| \leqslant \boldsymbol{s}_{ij}^{\boldsymbol{u}} \;\; \forall (i,j) \in E \cup E^R$$

$$S_i^g - \boldsymbol{S}_i^d = \textstyle\sum_{(i,j) \in E \cup E^R} S_{ij} \;\; \forall i \in N$$

→ demands are met

$$S_{ij} = \boldsymbol{Y}_{ij}^* |V_i|^2 - \boldsymbol{Y}_{ij}^* V_i V_j^* \;\; \forall (i,j) \in E \cup E^R$$

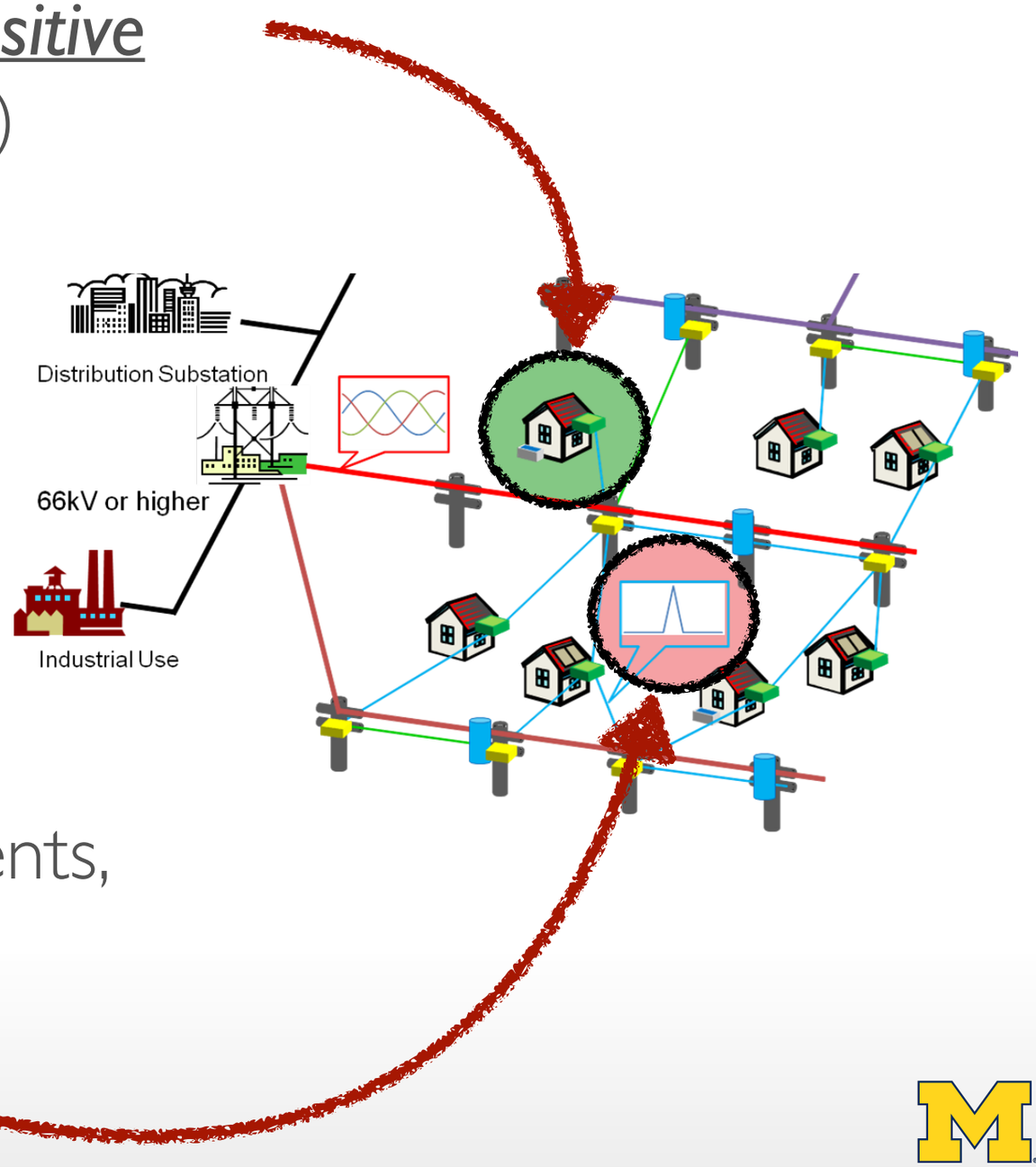
→ conservation of flow

Optimal Power Flow (OPF)

- AC-OPF Relaxations:
  - SOC Relaxation [*Jabr 2006*]
    Relaxes the product of voltage variables with second-order cone constraints
  - QC Relaxation [*Hijazi, Coffrin, and Van Hentenryck 2017*]
    Relaxes voltage constraints by taking tight convex envelops of their nonlinear terms
  - DC Relaxation [*Wood and Wollenberg 1996*]
    Relates real power to voltage phase angles, ignore reactive power, and assume voltages are colse to their nominal values

Differential Privacy Challenge for OPF

- Privacy in OPF test cases:
  - **Hide user participation:** ***not sensitive***
    (load location is typically known)
  - **Load magnitude:** ***sensitive***
    - Associated with customer's activity
    - May reveal strategic investments, decreases in sales, etc.

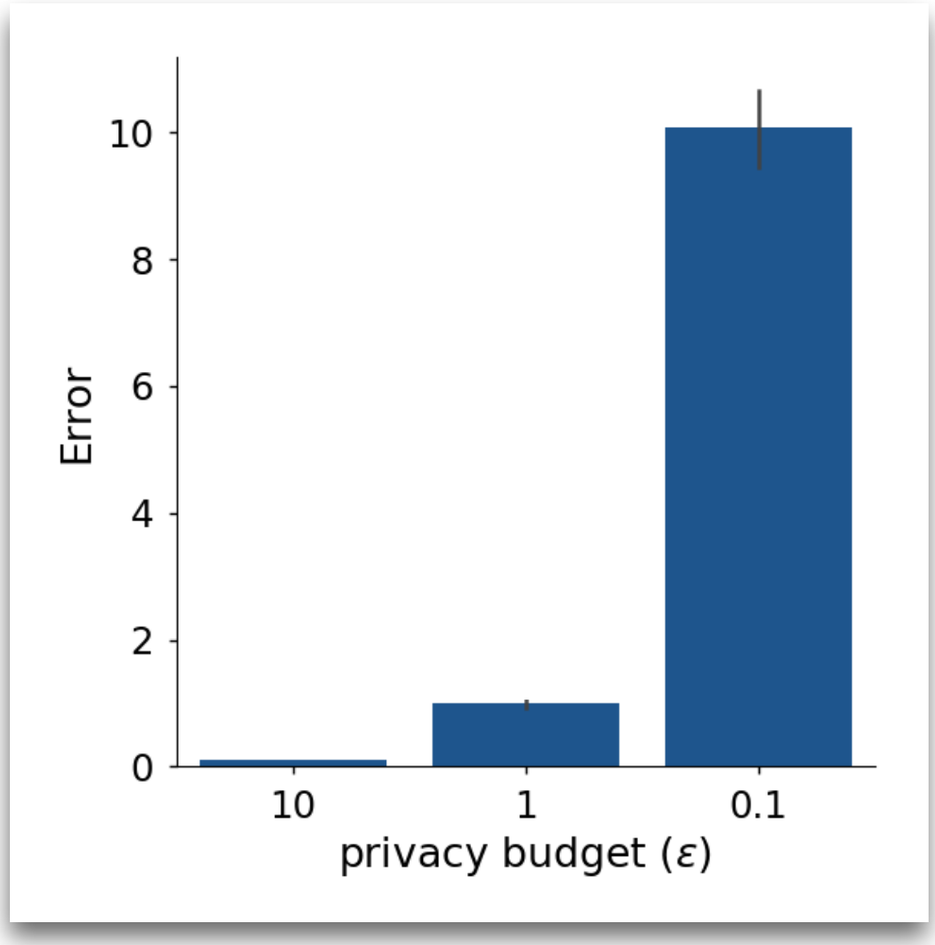


The Laplace mechanism for private OPF

- Undesirable outcomes when applied to protect load profiles

- Significant higher loads than the actual demand

- Recall: Larger privacy budget = less noise

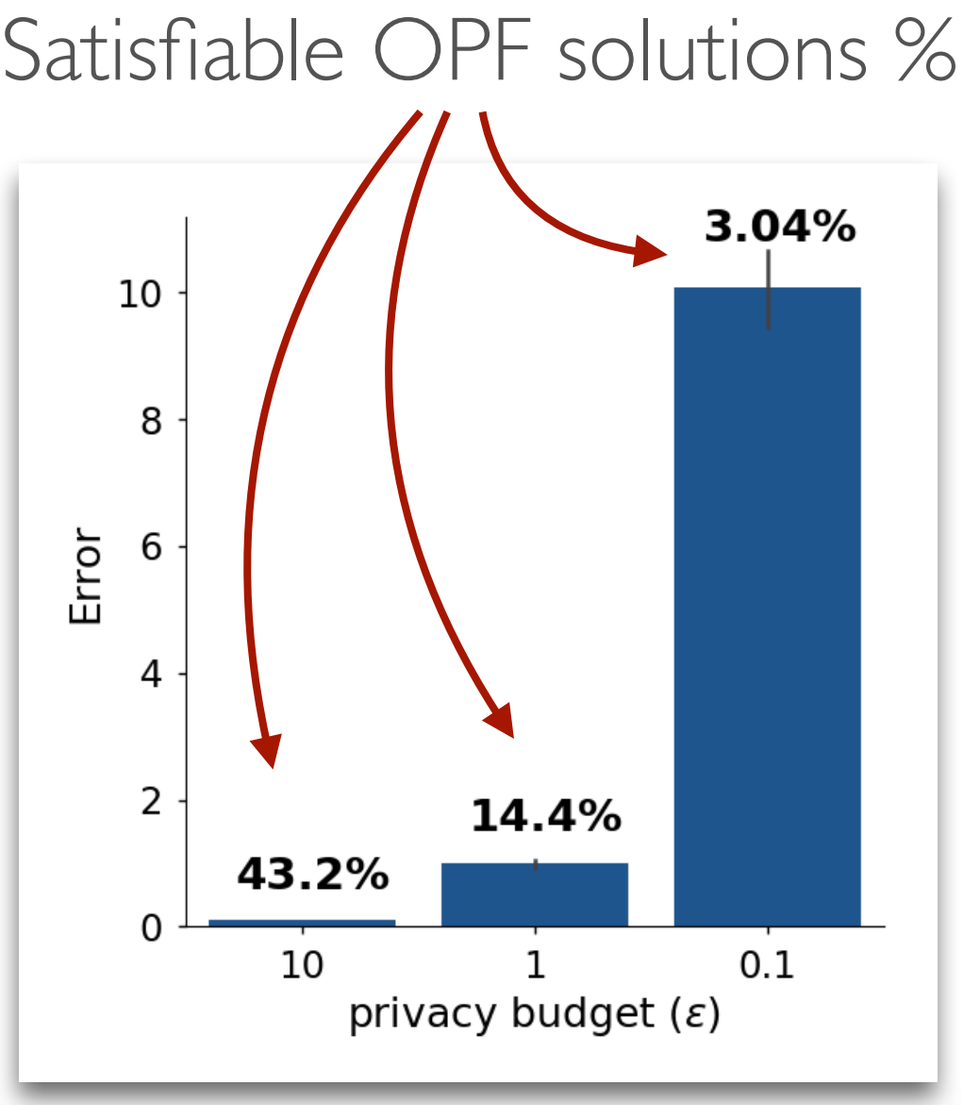


Average L1 error

The Laplace mechanism for private OPF

- The Laplace mechanism is oblivious to the structure of the dataset and the constraints and objective of the optimization problem

- It produces private datasets that are not representative for the actual OPF



Average L1 error

Content

- Private Data Release and Differential Privacy
- Optimal Power Flow Problem
- **The CBDP Mechanism**
- Experimental Analysis on Private OPF

DP for Complex Optimization Problems

- Consider a generic optimization problem

$$\begin{aligned} \text{minimize}_{\mathbf{x}\in\mathbb{R}^n} \quad & f(D,\mathbf{x}) \\ \text{subject to} \quad & g_i(D,\mathbf{x}) \leqslant 0, \quad i=1,\ldots,p \end{aligned}$$

  where D is the data whose privacy we want to protect.

- Desiderata:
  - Data privacy
  - Faithfulness to the optimal objective value
  - The private data must satisfy the problem constraints

Constraint-Based Differential Privacy

- Consider a generic optimization problem

$$\begin{aligned} \text{minimize}_{\mathbf{x}\in\mathbb{R}^n} \quad & f(D,\mathbf{x}) \\ \text{subject to} \quad & g_i(D,\mathbf{x}) \leqslant 0, \quad i=1,\ldots,p \end{aligned}$$

Definition 3 ((ϵ,β)-CBDP). *Given $\epsilon > 0, \beta \geqslant 0$, a DP-data-release mechanism $\mathcal{M} : \mathscr{D} \to \mathscr{D}$ is (ϵ,β)-CBDP iff, for each private database $\hat{D} = \mathcal{M}(D)$, there exists a solution $\mathbf{x}$ such that*

1. ϵ-privacy*: $\mathcal{M}$ satisfies ϵ-DP;*
2. β-faithfulness*: $|f(\hat{D},\mathbf{x}) - f(D,\mathbf{x}^*)| \leqslant \beta$;*
3. Consistency*: Constraints $g_i(\hat{D},\mathbf{x}) \leqslant 0$ $(i=1,\ldots,p)$ are satisfied.*

The CBDP Mechanism

1. Uses the Laplace mechanism to query each dimension of D:

$$\mathcal{M}_{\mathrm{Lap}}(D, Q, \epsilon) = \tilde{D} = D + \mathrm{Lap}(1/\epsilon)^n$$

where $\tilde{D} = (\tilde{c}_1, \ldots, \tilde{c}_n)$ is the vector of noisy values

2. Solves the following optimization problem:

$$\begin{aligned} &\text{minimize}_{\hat{D}, \mathbf{x} \in \mathbb{R}^n} \|\hat{D} - \tilde{D}\|_2^2 \\ &\quad \text{subject to} \quad |f(\hat{D}, \mathbf{x}) - f^*| \leqslant \beta \\ &\qquad\qquad\quad g_i(\hat{D}, \mathbf{x}) \leqslant 0, i = 1, \ldots, p \end{aligned}$$

3. Releases $\hat{D}$

The CBDP Mechanism

1. Uses the Laplace mechanism to query each dimension of D:

$$\mathcal{M}_{\mathrm{Lap}}(D, Q, \epsilon) = \tilde{D} = D + \mathrm{Lap}(1/\epsilon)^n$$

where $\tilde{D} = (\tilde{c}_1, \ldots, \tilde{c}_n)$ is the vector of noisy values

2. Solves the following optimization problem:

$$\begin{aligned} \text{minimize}_{\hat{D}, \mathbf{x} \in \mathbb{R}^n} \; & \|\hat{D} - \tilde{D}\|_2^2 \\ \text{subject to} \quad & |f(\hat{D}, \mathbf{x}) - f^*| \leqslant \beta \\ & g_i(\hat{D}, \mathbf{x}) \leqslant 0, i = 1, \ldots, p \end{aligned}$$

Decision variables: post-processed loads

3. Releases $\hat{D}$

The CBDP Mechanism

1. Uses the Laplace mechanism to query each dimension of D:

$$\mathcal{M}_{\mathrm{Lap}}(D, Q, \epsilon) = \tilde{D} = D + \mathrm{Lap}(1/\epsilon)^n$$

where $\tilde{D} = (\tilde{c}_1, \ldots, \tilde{c}_n)$ is the vector of noisy values

2. Solves the following optimization problem:

$$\begin{aligned} &\text{minimize}_{\hat{D}, \mathbf{x} \in \mathbb{R}^n} \; \|\hat{D} - \tilde{D}\|_2^2 \\ &\text{subject to} \quad |f(\hat{D}, \mathbf{x}) - f^*| \leqslant \beta \\ &\qquad\qquad g_i(\hat{D}, \mathbf{x}) \leqslant 0, i = 1, \ldots, p \end{aligned}$$

Decision variables: optimization problem

3. Releases $\hat{D}$

The CBDP Mechanism

1. Uses the Laplace mechanism to query each dimension of D:

$$\mathcal{M}_{\mathrm{Lap}}(D, Q, \epsilon) = \tilde{D} = D + \mathrm{Lap}(1/\epsilon)^n$$

where $\tilde{D} = (\tilde{c}_1, \ldots, \tilde{c}_n)$ is the vector of noisy values

2. Solves the following optimization problem:

$$\begin{aligned} &\text{minimize}_{\hat{D}, \mathbf{x} \in \mathbb{R}^n} \; \|\hat{D} - \tilde{D}\|_2^2 \\ &\text{subject to} \quad |f(\hat{D}, \mathbf{x}) - f^*| \leqslant \beta \\ &\qquad\qquad g_i(\hat{D}, \mathbf{x}) \leqslant 0, i = 1, \ldots, p \end{aligned}$$

Differential Privacy

3. Releases $\hat{D}$

The CBDP Mechanism

1. Uses the Laplace mechanism to query each dimension of D:

$$\mathcal{M}_{\mathrm{Lap}}(D, Q, \epsilon) = \tilde{D} = D + \mathrm{Lap}(1/\epsilon)^n$$

where $\tilde{D} = (\tilde{c}_1, \ldots, \tilde{c}_n)$ is the vector of noisy values

2. Solves the following optimization problem:

$$\begin{aligned} &\text{minimize}_{\hat{D}, \mathbf{x} \in \mathbb{R}^n} \; \|\hat{D} - \tilde{D}\|_2^2 \\ &\text{subject to} \quad |f(\hat{D}, \mathbf{x}) - f^*| \leqslant \beta \\ &\qquad\qquad g_i(\hat{D}, \mathbf{x}) \leqslant 0, i = 1, \ldots, p \end{aligned}$$

Faithfulness to the objective

3. Releases $\hat{D}$

The CBDP Mechanism

1. Uses the Laplace mechanism to query each dimension of D:

$$\mathcal{M}_{\mathrm{Lap}}(D, Q, \epsilon) = \tilde{D} = D + \mathrm{Lap}(1/\epsilon)^n$$

where $\tilde{D} = (\tilde{c}_1, \ldots, \tilde{c}_n)$ is the vector of noisy values

2. Solves the following optimization problem:

$$\begin{aligned} &\text{minimize}_{\hat{D}, \mathbf{x} \in \mathbb{R}^n} \; \|\hat{D} - \tilde{D}\|_2^2 \\ &\quad \text{subject to} \quad |f(\hat{D}, \mathbf{x}) - f^*| \leqslant \beta \\ &\qquad\qquad g_i(\hat{D}, \mathbf{x}) \leqslant 0, i = 1, \ldots, p \end{aligned}$$

Constraint consistency

3. Releases $\hat{D}$

The CBDP Mechanism

Properties

- Thm. (Privacy): It achieves (ε, β)-CBDP
  - By composition, post-processing, and noticing that a solution to the optimization model (step 2) always exists

- Thm. (Accuracy): The optimal solution $\langle \hat{D}^+, x^+ \rangle$ to the optimization model of CBDP satisfies:

$$\|\hat{D}^+ - D\|_2 \leq 2\|\tilde{D} - D\|_2$$

- Cor. CBDP is at most a factor 2 away from optimality

Content

- Private Data Release and Differential Privacy
- Optimal Power Flow Problem
- The CBDP Mechanism
- **Experimental Analysis on Private OPF**

Experimental analysis

Settings

- Data sets: *NESTA power network test cases (44 networks). Number of buses: 3 - 9241*

- *OPF Models: AC, QC, SOC, DC*

- *Privacy budget:* $\epsilon \in \{0.1, 1.0, 10.0\}$

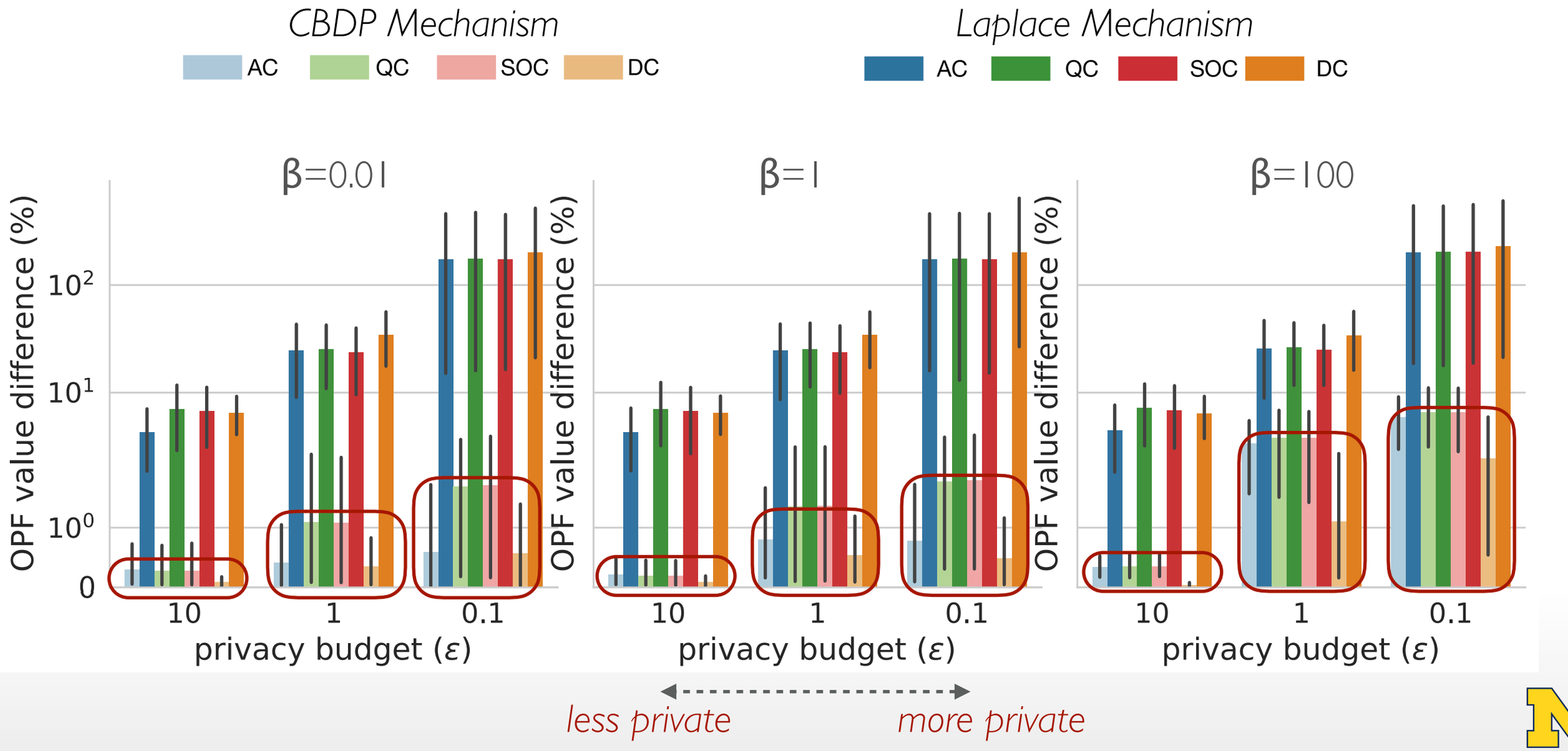
- *Faithfulness parameter:* $\beta \in \{0.01, 1.0, 100.0\}$

Experimental Analysis

Analysis of OPF cost

Summary:

- 1-2 order of magnitude improvements, for all **ε** and **β**
- Difference of OPF values between CBDP and original data is <10%

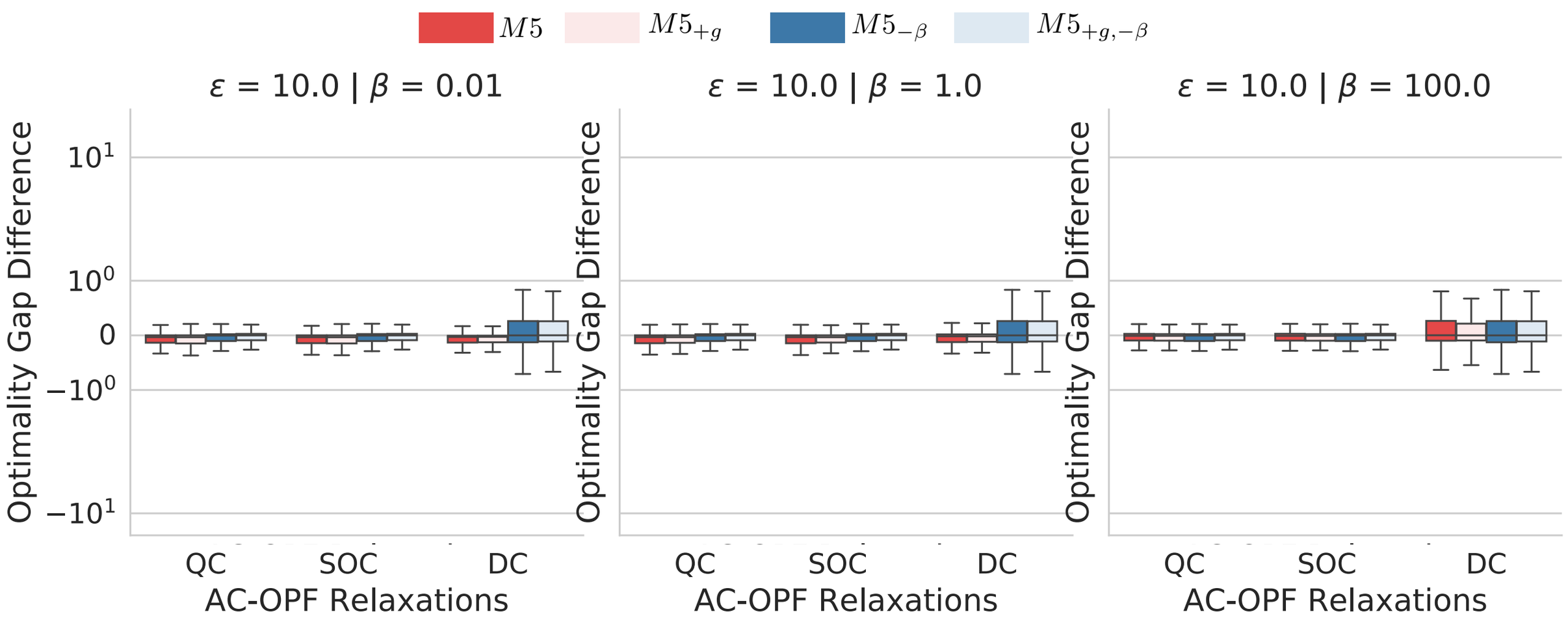


Experimental Analysis

Analysis of Optimality Gap

Summary:

- CBDP (M5) preserves the optimality gaps very closely (<1% for $\varepsilon \geq 1$ and < 3% for $\varepsilon < 1$).

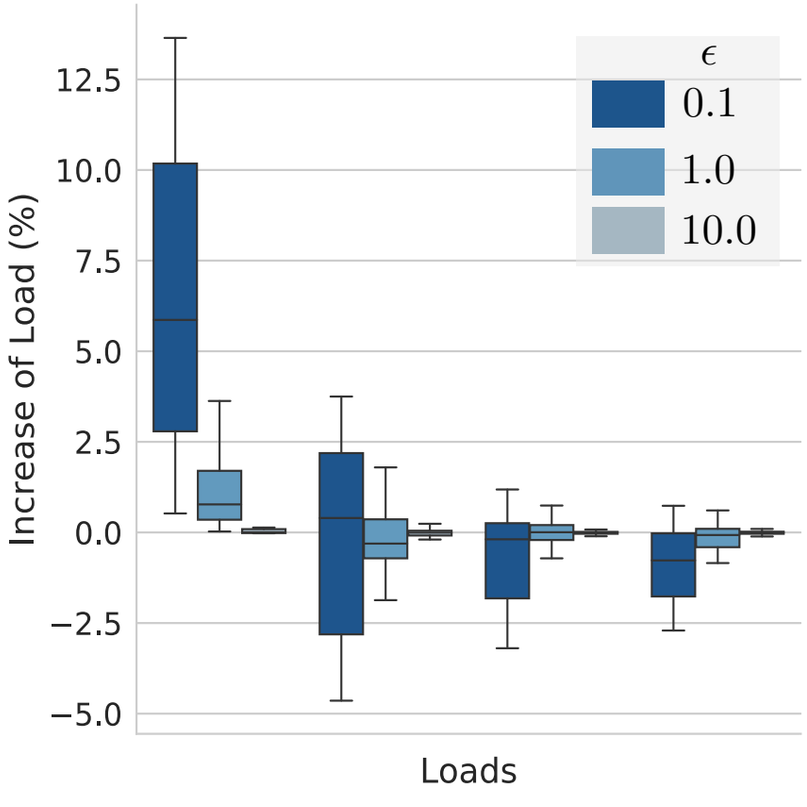


Experimental Analysis

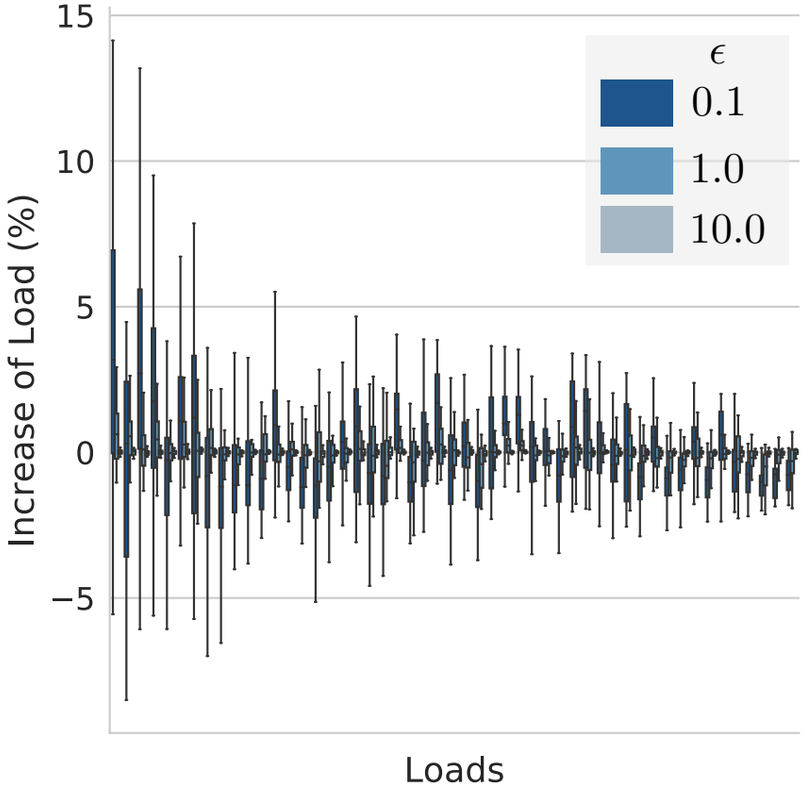
Analysis of the Private Network Loads

Summary:

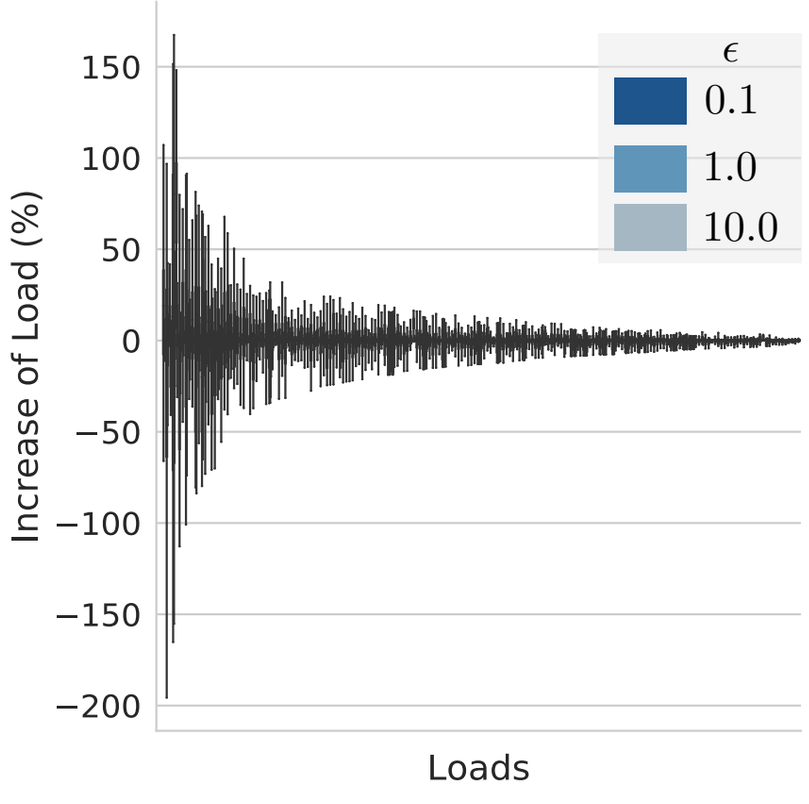
- Load variation is often significant for a portion of the loads
- Yet, the CBDP mechanism preserves the problem structure accurately



4-bus



73-bus



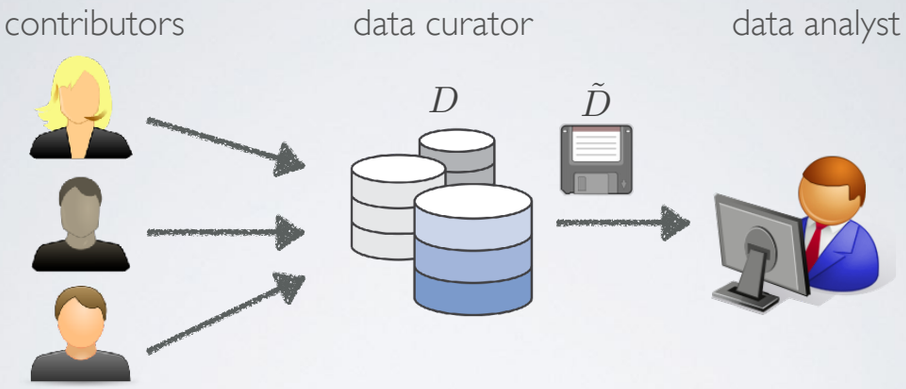
300-bus

Conclusions

- Motivated by the Differential Privacy Challenge for OPF
- Proposed a CBDP Mechanism which ensures:
  1. Differential private data release
  2. Faithfulness to the optimal objective value
  3. Constraint consistency
- We have applied CBDP to Optimal Power Flows
- CBDP improves the accuracy of the Laplace Mechanism by orders of magnitude while preserving salient computational features of the test cases

Thank you

PRIVATE DATA-RELEASE

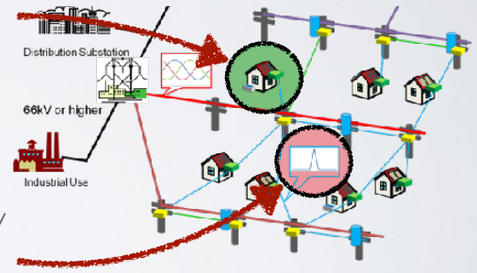


Publish a modified version of the data such that:

1. The contributors' privacy is adequately protected
2. The modified data is useful for the intended purpose

DIFFERENTIAL PRIVACY CHALLENGE FOR OPF

- Privacy in OPF test cases:
  - **Hide user participation:** ***not sensitive*** (load location is typically known)
  - **Load magnitude:** ***sensitive***
    - Associated with customer's activity
    - May reveal strategic investments, decreases in sales, etc.



CBDP MECHANISM

1. Uses the Laplace mechanism to query each dimension of the dataset:

$$\mathcal{M}_{\text{Lap}}(D, Q, \epsilon) = \tilde{D} = D + \text{Lap}(1/\epsilon)^n$$

   where $\tilde{D} = (\tilde{c}_1, \ldots, \tilde{c}_n)$ is the vector of noisy values.

2. Solves the following optimization problem:

$$\begin{aligned} &\text{minimize}_{\hat{D}, \mathbf{x} \in \mathbb{R}^n} \|\hat{D} - \tilde{D}\|_2^2 \\ &\text{subject to} \quad |f(\hat{D}, \mathbf{x}) - f^*| \leqslant \beta \\ &\qquad\qquad g_i(\hat{D}, \mathbf{x}) \leqslant 0, i = 1, \ldots, p \end{aligned}$$

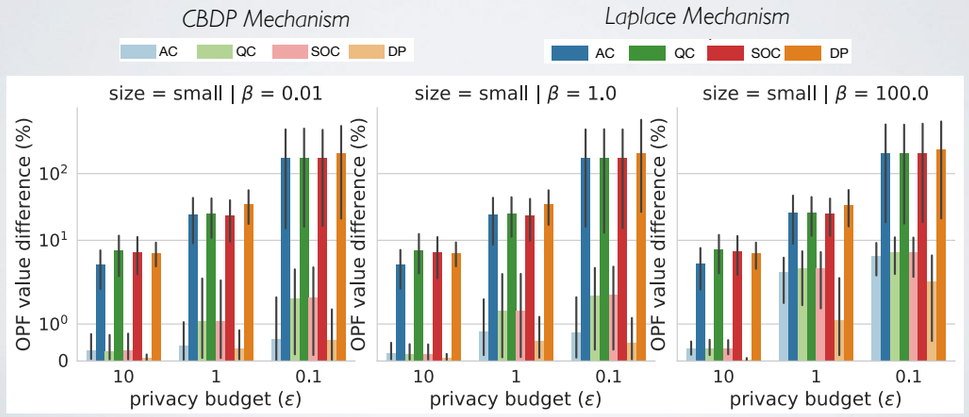
3. Releases $\hat{D}$

EXPERIMENTAL ANALYSIS

Analysis of OPF cost

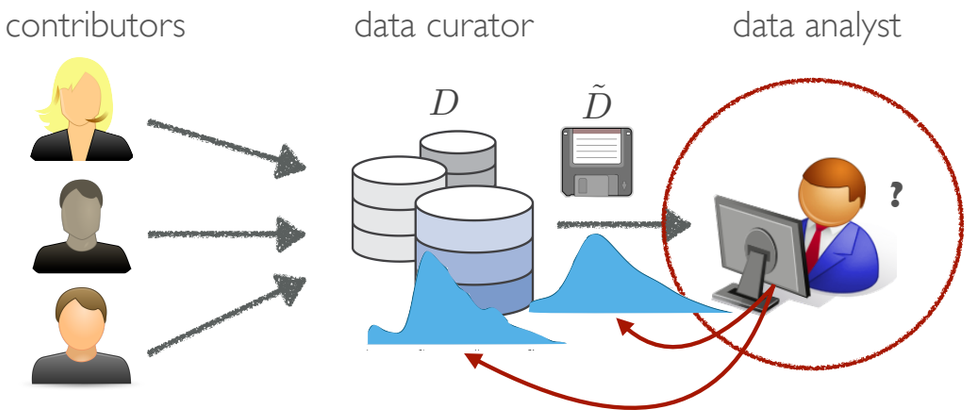
Summary:

- 1-2 order of magnitude improvements, for all **ε** and **β**
- Difference of OPF values between CBDP and original data is <10%



Thank you

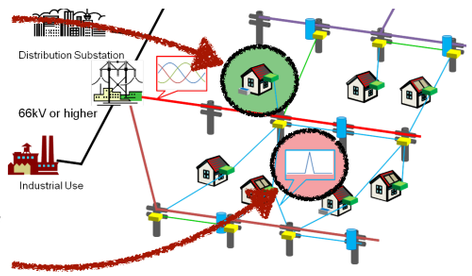
Differential Privacy
(Informal)



Contributor: Small participation risk (privacy loss)
Data analyst: Analysis on original and modified data are very similar (data distributions)

Differential Privacy Challenge for the OPF

- Privacy in OPF test cases:
  - **Hide user participation:** ***not sensitive*** (load location is typically known)
  - **Load magnitude:** ***sensitive***
    - Associated with customer's activity
    - May reveal strategic investments, decreases in sales, etc.



The CBDP Mechanism

- The main idea: Let's exploit the problem structure
- Ask additional queries on aggregate counts (problem *features*), e.g.,
  - The total number trips
  - The number of trips in each zone
  - The number of trips made with each combination of transportation modes
- Use this (noisy) information to redistribute the noise introduced on the individual trips counts
- Also, enforce consistency!

$$\textbf{minimize: } \|\mathbf{x} - \tilde{\mathbf{c}}\|_{2,w}^2 = \sum_{i=1}^{k} \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \tilde{c}_{ij})^2 \qquad \text{(O1)}$$

subject to:

$$\forall i', i : \mathbf{F}_{i'} \prec \mathbf{F}_i, j \in [n_i] : x_{ij} = \sum_{l : \mathbf{d}_{i'l} \subseteq \mathbf{d}_{ij}} x_{i'l} \qquad \text{(O2)}$$

$$\forall i, j : x_{ij} \geq 0. \qquad \text{(O3)}$$

Lil'bit of shameless ad:

I am in the faculty job market!

fioretto@umich.edu

References

[Jabr 2006] R. Jabr. Radial distribution load flow using conic programming. Power Systems, IEEE Transactions on, 21(3):1458–1459, Aug 2006.

[Hijazi, Coffrin, and Van Hentenryck 2017] H. Hijazi, C. Coffrin, and P. Van Hentenryck. Convex Quadratic Relaxations of Nonlinear Programs in Power Systems. Mathematical Programming Computation, 32(5): 3549–3558, 2017.

[Wood and Wollenberg 1996] A. J. Wood and B. F. Wollenberg. Power Generation, Operation, and Control. Wiley-Interscience, 1996.