Constrained-based Differential Privacy
Releasing Optimal Power Flow Benchmarks Privately

Ferdinando Fioretto and Pascal Van Hentenryck

University of Michigan, Ann Arbor, MI, USA
{fioretto,pvanhent}@umich.edu

Abstract. This paper considers the problem of releasing optimal power flow benchmarks that maintain the privacy of customers (loads) using the notion of *Differential Privacy*. It is motivated by the observation that traditional differential-privacy mechanisms are not accurate enough: The added noise fundamentally changes the nature of the underlying optimization and often leads to test cases with no solution. To remedy this limitation, the paper introduces the framework of *Constraint-Based Differential Privacy (CBDP)* that leverages the post- processing immunity of differential privacy to improve the accuracy of traditional mechanisms. More precisely, CBDP solves an optimization problem to satisfies the problem-specific constraints by redistributing the noise. The paper shows that CBDP enjoys desirable theoretical properties and produces orders of magnitude improvements on the largest set of test cases available.

1 Introduction

In the last decades, scientific advances in artificial intelligence and operations research have been driven by competitions and collections of test cases. The MIPLIB library for mixed-integer programming and the constraint-programming, planning, and SAT competitions have significantly contributed to advancing the theoretical and experimental branches of the field. Recent years have also witnessed the emergence of powerful platforms (such as Kaggle [1]) to organize competitions between third-parties. Finally, the release of data sets may become increasingly significant in procurement where third-parties compete to demonstrate their capabilities. The desire to release data sets for scientific research, competitions, and procurement decisions is likely to accelerate. Indeed, with ubiquitous connectivity, many organizations are now collecting data at an unprecedented scale, often on large socio-technical systems such as energy networks. This data is often used as input to complex optimization problems.

The release of such rich data sets however raises some fundamental privacy concerns. For instance, the electrical load of an industrial customer in the power grid typically reflects its production and may reveal sensitive information on its economic strategy.

Differential Privacy (DP) [5, 6] is a general framework that addresses the sensitivity of such information and can be used to generate privacy-preserving data sets. It introduces carefully calibrated noise to the entries of a data set to prevent the disclosure of information related to those providing it. However, when these private data sets are used as inputs to complex optimization algorithms, they may produce results that are fundamentally different from those obtained on the original data set. For instance, the noise added by differential privacy may make the optimization problem infeasible or much easier to solve. As a result, despite its strong theoretical foundations, adoptions of differential privacy in industry and government have been rare. Large-scale practical deployments of differential privacy have been limited to big-data owners such as Google [7] and Apple [10]. In their applications, however, only internal users can access the private data by evaluating a pre-defined set of *queries*, e.g., the count of individuals satisfying certain criteria. This is because constructing a private version of the database is equivalent to simultaneously answering all possible queries and thus requires a large amount of noise.

This paper is motivated by the desire of releasing *Optimal Power Flow* (OPF) benchmarks that maintain the privacy of customers loads and the observation that traditional differential-privacy mechanisms are not accurate enough: The added noise fundamentally changes the nature of the underlying optimization and often leads to test cases with no solution. The paper proposes the framework of *Constraint-Based Differential Privacy* (CBDP) that leverages the post-processing immunity of DP to redistribute the noise introduced by a standard DP-mechanism, so that the private data set preserves the salient features of the original data set. More precisely, CBDP solves an optimization problem that minimizes the distance between the post-processed and original data, while satisfying constraints that capture the essence of the optimization application.

The paper shows that the CBDP has strong theoretical properties: It achieves ϵ-differential privacy, ensures that the released data set can produce feasible solutions for the optimization problem of interest, and is a constant factor away from optimality. Finally, experimental results show that the CBDP mechanism can be adopted to generate private OPF test cases: On the largest collection of OPF test cases available, it improves the accuracy of existing approaches of at least one order of magnitude and results in solutions with similar optimality gaps to those obtained on the original problems.

2 Differential Privacy

A *data set* D is a multi-set of elements in the *data universe* $\mathscr{U}$. The set of every possible data set is denoted $\mathscr{D}$. Unless stated otherwise, $\mathscr{U}$ is a cross product of multiple *attributes* $U_1, \ldots, U_n$ and has *dimension* n. For example, $\mathscr{U} = \mathbb{R}^n$ is the numeric data universe consisting of n-dimensional real-vectors. A *numeric query* is a function that maps a data set to a result in $\mathscr{R} \subseteq \mathbb{R}^r$.

Two data sets $D_1, D_2 \in \mathscr{D}$ are called *neighbors* (written $D_1 \sim D_2$) if D_1 and D_2 differ by at most one element, i.e., $|(D_1 - D_2) \cup (D_2 - D_1)| = 1$.

Definition 1 (Differential Privacy [5]). *A randomized mechanism $\mathcal{M} : \mathscr{D} \to \mathscr{R}$ with domain $\mathscr{D}$ and range $\mathscr{R}$ is* ϵ-differentially private *if, for any event $\mathcal{S} \subseteq \mathscr{R}$*

and any pair $D_1, D_2 \in \mathscr{D}$ of neighboring data sets:

$$Pr[\mathcal{M}(D_1) \in \mathcal{S}] \leqslant \exp(\epsilon) Pr[\mathcal{M}(D_2) \in \mathcal{S}], \tag{1}$$

where the probability is calculated over the coin tosses of $\mathcal{M}$.

A Differential Privacy (DP) mechanism maps a data set to distributions over the output set. The released DP output is a single random sample drawn from these distributions. The level of privacy is controlled by the parameter $\epsilon \geqslant 0$, called the *privacy budget*, with values close to 0 denoting strong privacy.

DP satisfies several important properties. Composability ensures that a combination of differentially private mechanisms preserve DP [6].

Theorem 1 (Composition). *Let $M_i : \mathscr{D} \to \mathscr{R}_i$ be an ϵ_i-differentially private mechanism for $i = 1, \ldots, k$. Then, their composition, defined as $\mathcal{M}(D) = (\mathcal{M}_i(D), \ldots, \mathcal{M}_k(D))$, is $(\sum_{i=1}^{k} \epsilon_i)$-differentially private.*

Post-processing immunity ensures that privacy guarantees are preserved by arbitrary post-processing steps [6].

Theorem 2 (Post-Processing Immunity). *Let $\mathcal{M} : \mathscr{D} \to \mathscr{R}$ be a mechanism that is ϵ-differentially private and $g : \mathscr{R} \to \mathscr{R}'$ be an (arbitrary) mapping. The mechanism $g \circ \mathcal{M}$ is ϵ-differentially private.*

The Laplace Distribution with 0 mean and scale b has a probability density function $\text{Lap}(x|b) = \frac{1}{2b} e^{-\frac{|x|}{b}}$. The *sensitivity* of a query Q, denoted by $\varDelta_Q$, is defined as $\varDelta_Q = \max_{D_1 \sim D_2} \|Q(D_1) - Q(D_2)\|_1$. The following theorem gives a differentially private mechanism for answering numeric queries [5].

Theorem 3 (Laplace Mechanism). *Let $Q : \mathscr{D} \to \mathscr{R}$ be a numerical query. The Laplace mechanism, defined as $\mathcal{M}_{Lap}(D; Q, \epsilon) = Q(D) + z$ where $z \in \mathscr{R}$ is a vector of i.i.d. samples drawn from Lap$(\frac{\varDelta_Q}{\epsilon})$ achieves ϵ-differential privacy.*

The Laplace mechanism is a particularly useful building block for DP [6]. Koufogiannis et al. [16] proved its optimality by showing that it minimizes the mean-squared error for both the L_1 and L_2 norms among all private mechanisms that use additive and input-independent noise. In the following, $\text{Lap}(\lambda)^n$ denotes the i.i.d. Laplace distribution over n dimensions with parameter λ.

Lipschitz privacy The concept of Lipschitz privacy is appropriate when a data owner desire to protect individual quantities rather than individual participation in a data set. Application domains where Lipschitz privacy has been successful include location-based systems [3, 20] and epigenetics [4].

Definition 2 (Lipschitz privacy [16]). *Let $\mathscr{D}$ be a metric space. A randomized mechanism $\mathcal{M} : \mathscr{D} \to \mathscr{R}$ is ϵ-Lipschitz differentially private if:*

$$Pr[\mathcal{M}(D_1) \in \mathcal{S}] \leqslant \exp(\epsilon \|D_1 - D_2\|) Pr[\mathcal{M}(D_2) \in \mathcal{S}],$$

for any $\mathcal{S} \subseteq \mathscr{R}$ and any two inputs $D_1, D_2 \in \mathscr{D}$.

The Laplace mechanism with parameter ϵ achieve ϵ-Lipschitz DP.

N	The set of nodes in the network	θ^{Δ}	Phase angle difference limits
E	The set of *from* edges in the network	$S^d = p^d + \boldsymbol{i}q^d$	AC power demand
E^R	The set of *to* edges in the network	$S^g = p^g + \boldsymbol{i}q^g$	AC power generation
$\boldsymbol{i}$	imaginary number constant	c_0, c_1, c_2	Generation cost coefficients
I	AC current	$\Re(\cdot)$	Real component of a complex number
$S = p + \boldsymbol{i}q$	AC power	$\Im(\cdot)$	Imaginary component of a complex number
$V = v\angle\theta$	AC voltage	$(\cdot)^*$	Conjugate of a complex number
$Y = g + \boldsymbol{i}b$	Line admittance	$\lvert \cdot \rvert$	Magnitude of a complex number
$W = w^R + \boldsymbol{i}w_I$	Product of two AC voltages	$\angle$	Angle of a complex number
s^u	Line apparent power thermal limit	x^l, x^u	Upper and lower bounds of x
θ_{ij}	Phase angle difference (i.e., $\theta_i - \theta_j$)	$\boldsymbol{x}$	A constant value

Table 1. Power Network Nomenclature.

3 Optimal Power Flow

Optimal Power Flow (OPF) is the problem of determining the best generator dispatch to meet the demands in a power network. A power network is composed of a variety of components such as buses, lines, generators, and loads. The network can be viewed as a graph (N, E) where the set of buses N represent the nodes and the set of lines E represent the edges. Note that E is a set of directed arcs and E^R is used to denote those arcs in E but in reverse direction. Table 1 reviews the symbols and notation adopted in this paper. Bold-faced symbols are used to denote constant values.

The AC Model The AC power flow equations are based on complex quantities for current I, voltage V, admittance Y, and power S. The quantities are linked by constraints expressing Kirchhoff's Current Law (KCL), i.e., $I^g_i - \boldsymbol{I}^d_i = \sum_{(i,j)\in E\cup E^R} I_{ij}$, Ohm's Law, i.e., $I_{ij} = \boldsymbol{Y}_{ij}(V_i - V_j)$,, and the definition of AC power, i.e., $S_{ij} = V_i I^*_{ij}$. Combining these three properties yields the AC Power Flow equations, i.e.,

$$S^g_i - \boldsymbol{S}^d_i = \sum_{(i,j)\in E\cup E^R} S_{ij} \quad \forall i \in N$$

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

These non-convex nonlinear equations are a core building block in many power system applications. Practical applications typically include various operational constraints on the flow of power, which are captured in the AC OPF formulation in Model 1. The objective function (2) captures the cost of the generator dispatch. Constraint (3) sets the reference angle for some arbitrary $r \in N$, to eliminate numerical symmetries. Constraints (4) and (5) capture the voltage and phase angle difference operational constraints. Constraints (6) and (7) enforce the generator output and line flow limits. Finally, Constraints (8) capture KCL and constraints (9) capture Ohm's Law.

Notice that this is a non-convex nonlinear optimization problem and is NP-Hard [17, 23]. Therefore, significant attention has been devoted to finding convex relaxations of Model 1.

Model 1 The AC Optimal Power Flow Problem (AC-OPF)

$$\textbf{variables: } S^g_i, V_i \ \ \forall i \in N, \ \ S_{ij} \ \ \forall (i,j) \in E \cup E^R$$

$$\textbf{minimize: } \sum_{i \in N} \boldsymbol{c}_{2i}(\Re(S^g_i))^2 + \boldsymbol{c}_{1i}\Re(S^g_i) + \boldsymbol{c}_{0i} \tag{2}$$

$$\textbf{subject to: } \angle V_{\boldsymbol{r}} = 0, \ \ r \in N \tag{3}$$

$$\boldsymbol{v}^{\boldsymbol{l}}_i \leqslant |V_i| \leqslant \boldsymbol{v}^{\boldsymbol{u}}_i \ \ \forall i \in N \tag{4}$$

$$-\boldsymbol{\theta}^{\boldsymbol{\Delta}}_{ij} \leqslant \angle(V_i V_j^*) \leqslant \boldsymbol{\theta}^{\boldsymbol{\Delta}}_{ij} \ \ \forall (i,j) \in E \tag{5}$$

$$\boldsymbol{S}^{\boldsymbol{gl}}_i \leqslant S^g_i \leqslant \boldsymbol{S}^{\boldsymbol{gu}}_i \ \ \forall i \in N \tag{6}$$

$$|S_{ij}| \leqslant \boldsymbol{s}^{\boldsymbol{u}}_{ij} \ \ \forall (i,j) \in E \cup E^R \tag{7}$$

$$S^g_i - \boldsymbol{S}^d_i = \textstyle\sum_{(i,j) \in E \cup E^R} S_{ij} \ \ \forall i \in N \tag{8}$$

$$S_{ij} = \boldsymbol{Y}^*_{ij}|V_i|^2 - \boldsymbol{Y}^*_{ij} V_i V_j^* \ \ \forall (i,j) \in E \cup E^R \tag{9}$$

The SOC Relaxation The SOC relaxation [14] lifts the product of voltage variables $V_i V_j^*$ into a higher dimensional space (i.e., the W-space):

$$W_i = |V_i|^2 \ \ i \in N \tag{10a}$$

$$W_{ij} = V_i V_j^* \ \ \forall (i,j) \in E \tag{10b}$$

It takes the absolute square of each constraint (10b), refactors it, and relaxes the equality into an inequality:

$$|W_{ij}|^2 \leqslant W_i W_j \ \ \forall (i,j) \in E \tag{11}$$

Constraint (11) is a second-order cone constraint, which is widely supported by industrial strength convex optimization tools (e.g., Gurobi [11], CPlex [13], Mosek [21]). The SOC relaxation of (AC-OPF) is presented in Model 2 (SOC-OPF). The constraints for the generator output limits (6), line flow limits (7), and KCL (8), are identical to those in the (AC-OPF) model. Constraints (12) and (13) capture the voltage and phase angle difference operational constraints. Constraints (14) and (15) capture the line power flow in the W-space. Finally, constraints (16) strengthen the relaxation with second-order cone constraints for voltage products.

The Quadratic Convex (QC) Relaxation The QC relaxation was introduced to preserve stronger links between the voltage variables [12]. It represents the voltages in polar form (i.e., $V = v\angle\theta$) and links these real variables to the W variables using the following equations:

$$W_{ii} = v_i^2 \ \ i \in N \tag{17a}$$

$$\Re(W_{ij}) = v_i v_j \cos(\theta_i - \theta_j) \ \ \forall (i,j) \in E \tag{17b}$$

$$\Im(W_{ij}) = v_i v_j \sin(\theta_i - \theta_j) \ \ \forall (i,j) \in E \tag{17c}$$

The QC relaxation relaxes these equations by taking tight convex envelopes of their nonlinear terms, exploiting the operational limits for $v_i, v_j, \theta_i - \theta_j$. In particular, it uses the convex envelopes for the square $\langle x^2 \rangle^T$ and product $\langle xy \rangle^M$

Model 2 The SOC Relaxation of AC-OPF (SOC-OPF)

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

minimize:(2)

subject to: (6), (7), (8)

$$(\boldsymbol{v}_i^{\boldsymbol{l}})^2 \leqslant W_i \leqslant (\boldsymbol{v}_i^{\boldsymbol{u}})^2 \;\; \forall i \in N \tag{12}$$

$$\tan(-\boldsymbol{\theta}_{ij}^{\boldsymbol{\Delta}})\Re(W_{ij}) \leqslant \Im(W_{ij}) \leqslant \tan(\boldsymbol{\theta}_{ij}^{\boldsymbol{\Delta}})\Re(W_{ij}) \;\; \forall (i,j) \in E \tag{13}$$

$$S_{ij} = \boldsymbol{Y}_{ij}^* W_i - \boldsymbol{Y}_{ij}^* W_{ij} \;\; (i,j) \in E \tag{14}$$

$$S_{ji} = \boldsymbol{Y}_{ij}^* W_j - \boldsymbol{Y}_{ij}^* W_{ij}^* \;\; (i,j) \in E \tag{15}$$

$$|W_{ij}|^2 \leqslant W_i W_j \;\; \forall (i,j) \in E \tag{16}$$

Model 3 The QC Relaxation of AC-OPF (QC-OPF)

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

minimize: (2)

subject to: (3)–(8), (14)–(16)

$$W_{ii} = \langle v_i^2 \rangle^T \;\; i \in N \tag{18}$$

$$\Re(W_{ij}) = \langle \langle v_i v_j \rangle^M \langle \cos(\theta_i - \theta_j) \rangle^C \rangle^M \;\; \forall (i,j) \in E \tag{19}$$

$$\Im(W_{ij}) = \langle \langle v_i v_j \rangle^M \langle \sin(\theta_i - \theta_j) \rangle^S \rangle^M \;\; \forall (i,j) \in E \tag{20}$$

of variables, as defined in [19]. Under the assumption that the phase angle difference bound is within $-\boldsymbol{\pi}/2 \leqslant \boldsymbol{\theta}_{ij}^{\boldsymbol{l}} \leqslant \boldsymbol{\theta}_{ij}^{\boldsymbol{u}} \leqslant \boldsymbol{\pi}/2$, relaxations for sine $\langle \sin(x) \rangle^S$ and cosine $\langle \cos(x) \rangle^C$ are given in reference [12]. Convex envelopes for equations (17a)–(17c) can be obtained by composing the convex envelopes of the functions for square, sine, cosine, and the product of two variables, i.e.,

$$W_{ii} = \langle v_i^2 \rangle^T \;\; i \in N \tag{21a}$$

$$\Re(W_{ij}) = \langle \langle v_i v_j \rangle^M \langle \cos(\theta_i - \theta_j) \rangle^C \rangle^M \;\; \forall (i,j) \in E \tag{21b}$$

$$\Im(W_{ij}) = \langle \langle v_i v_j \rangle^M \langle \sin(\theta_i - \theta_j) \rangle^S \rangle^M \;\; \forall (i,j) \in E \tag{21c}$$

The QC relaxation also proposes to strengthen these convex envelopes with a second-order cone constraint from the SOC relaxation ((10a), (10b), (11)). The complete QC relaxation is presented in Model 3.

The DC model The DC model is an extensively studied linear approximation to the AC power flow [24]. The DC load flow relates real power to voltage phase angle, ignores reactive power, and assumes voltages are close to their nominal values (1.0 in per unit notation). The DC OPF is presented in Model 4. Constraints (22) capture the phase angles operational constraints. Constraints (23) and (24) enforce the generator output and line flow limits. Constraints (25) captures the KCL and constraints (26) the Ohm's Law.

Model 4 The DC Relaxation of the AC OPF (DC-OPF)

$$\textbf{variables:} \quad \Re(S^g_i), \theta_i \;\; \forall i \in N, \;\; S_{ij} \;\; \forall (i,j) \in E \cup E^R$$

$$\textbf{minimize:} \quad (2)$$

$$\textbf{subject to:} \quad (3)$$

$$|\theta_i| \leqslant \boldsymbol{\theta^u_i} \;\; \forall i \in N \tag{22}$$

$$\Re(\boldsymbol{S^{gl}_i}) \leqslant \Re(S^g_i) \leqslant \Re(\boldsymbol{S^{gu}_i}) \;\; \forall i \in N \tag{23}$$

$$\Re(|S_{ij}|) \leqslant \Re(\boldsymbol{s^u_{ij}}) \;\; \forall (i,j) \in E \cup E^R \tag{24}$$

$$\Re(S_{ij}) = -\boldsymbol{b}_{ij}(\theta_i - \theta_j) \;\; \forall (i,j) \in E \cup E^R \tag{25}$$

$$\Re(S^g_i) - \Re(\boldsymbol{S^d_i}) = \textstyle\sum_{(i,j)\in E \cup E^R} S_{ij} \;\; \forall i \in N \tag{26}$$

4 The Differential Privacy Challenge for OPF

When releasing private OPF test cases, it is not critical to hide user participation: The location of a load is public knowledge. However, the magnitude of a load is sensitive: It is associated with the activity of a particular customer (or group of customers) and may indirectly reveal production levels and hence strategic investments, decreases in sales, and other similar information. Indirectly, it may also reveal how transmission operators operate their networks, which should not be public information. As a result, the concept of *Lipschitz differential privacy* is particularly suited to the task.

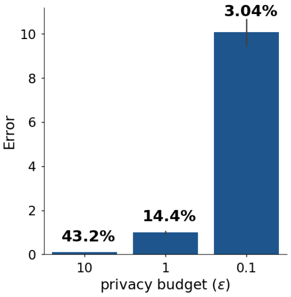

Fig. 1. Average L_1 error reported by the Laplace Mechanism. The percentages express the AC-OPF instances with satisfiable solution.

As mentioned in Section 2, the Laplace mechanism can be used to achieve Lipschitz DP. However, its application on load profile queries results in a new output vector of loads which produces undesirable outcomes when used as input to an OPF problem. Indeed, Figure 1illustrates the average error (measured as the L_1 distance) between the original load and the private load for a set of 44 networks.[1] With a privacy budget of $\epsilon = 0.1$, the average error is about 10–implying a significantly higher load than the actual demand. The numbers reported on each bar represent the percentage of feasible private instances for the AC OPF problem: It reveals severe feasibility issues with the private instances.

These results highlight the challenges that arise when traditional differential privacy is applied to inputs of complex optimization tasks. For instance, the Laplacian mechanism is oblivious to the structure of the data set (e.g., the generation capabilities should be large enough to serve the load) and the constraints and objectives of the optimization application (e.g., the transmission network should have the ability to transport electricity from generators to loads). As a

[1] The experimental settings are reported in all details in Section 7.

result, it produces private data sets that are typically not useful and not representative of actual OPFs. *What is needed is a differential-privacy mechanism that preserves the structure of the optimization model and its computational properties such as the optimality gap between the solutions produced by MINLP solvers and convex relaxations.*

5 Constrained-Based Differential Privacy

This section introduces Constraint-Based Differential Privacy (CBDP) to remedy the limitations identified in the previous section. It considers an optimization problem $\mathcal{O}(D)$:

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

where $f : \mathscr{D} \times \mathbb{R}^n \to \mathbb{R}$ is the *objective function* to minimize over variables $\mathbf{x}$ and $g_i(D,x) \leqslant 0$ $(i = 1,\dots,p)$ are the problem constraints.

This paper studies the following setting. The data owner desires to release a private data set $\hat{D}$ such that the optimization problems $\mathcal{O}(D)$ and $\mathcal{O}(\hat{D})$ are closely related. In particular, the optimal objective value of $\mathcal{O}(\hat{D})$ must be *close* to the optimal value of the original problem $f(D,\mathbf{x}^*)$ (which is a public information), where $\mathbf{x}^* \in \mathbb{R}^n$ is the optimal solution of the original optimization problem. Hence the private data set must satisfy the following desiderata: (1) *data privacy*: The data set to be released must be private; (2) *faithfulness*: The private data must be faithful to the objective function; (3) *consistency*: The private data must satisfy the constraints arising from the data and/or from the problem of interest. To address such challenges the following definition is introduced.

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 $g_i(\hat{D},\mathbf{x}) \leqslant 0$ $(i = 1,\dots,p)$ are satisfied.*

The Input To balance between utility and privacy, the mechanism takes as input the data set D, as well as two non-negative real numbers: ϵ which determines the *privacy value* of the private data and β which determines the required *faithfulness* of the optimization problem over the private data. Additionally, the data owner provides the optimization problem and the optimal objective value $f^* = f(D,\mathbf{x}^*)$, which are typically considered public information in competitions. For simplicity, this section assumes that $\mathscr{D} = \mathbb{R}^n$.

$$\begin{aligned} &\text{minimize}_{\hat{D},\mathbf{x}\in\mathbb{R}^n} \|\hat{D} - \tilde{D}\|_2^2 && \text{(O1)}\\ &\text{subject to} \quad |f(\hat{D},\mathbf{x}) - f^*| \leqslant \beta && \text{(O2)}\\ &\qquad\qquad g_i(\hat{D},\mathbf{x}) \leqslant 0, i = 1,\dots,p && \text{(O3)} \end{aligned}$$

Fig. 2. The CBDP Post-Processing Step.

The Mechanism The CBDP mechanism first injects Laplace noise with privacy parameter ϵ to each query on each dimension of the dataset:

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

where $\tilde{D} = (\tilde{c}_1, \dots, \tilde{c}_n)$ is the vector of noisy values. These values are then post-processed by the optimization algorithm specified in Figure 2 to obtain a value vector $\hat{D} = (\hat{x}_1, \dots, \hat{x}_n) \in \mathbb{R}^n$. Finally, the CBDP mechanism outputs $\hat{D}$.

The CBDP mechanism thus solves a constrained optimization problem whose decision variables include vectors of the form $\hat{D} = (\hat{x}_1, \dots, \hat{x}_n)$ that correspond to the post-processed result of the private query on each dimension of the universe. In other words, each original data (that must remain private) is replaced by a decision variable representing its post-processed private counterpart. The objective (O1) minimizes the L_2-norm between the private query result $\tilde{D}$ and its post-processed version $\hat{D}$. Constraint (O2) forces the post-processed values to be β-faithful with the respect to the objective value, and Constraints (O3) enforce the optimization constraints of the original model.

Additional constraints capturing auxiliary public information about the data can be integrated in this model. For instance, in the OPF problem, the total power load is public. Thus, an additional constraint on the sum of values of the variables $(\hat{x}_1, \dots \hat{x}_n)$ can be enforced to be equal to this public information.

The CBDP post-processing can be thought as redistributing the noise of the Laplace mechanism to obtain a data set which is consistent with the problem constraints and objective. It searches for a feasible solution that satisfies the problem constraints $g_i(\hat{D}, \mathbf{x}) \leqslant 0$ and the β-faithfulness constraint. *A feasible solution always exists, since the original values D trivially satisfy all constraints.*

It is important to notice that the post-processing step of CBDP uses exclusively the private data set $\tilde{D}$ and additional public information (i.e., the optimization problem and its optimal solution value). Its privacy guarantees are discussed below.

5.1 Theoretical Properties

Theorem 4. *The mechanism above is (ϵ, β)-CBDP.*

Proof. Each $\tilde{c}_i$ obtained from the Laplace mechanism is ϵ-differentially-private by Theorem 3. The combination of these results $(\tilde{c}_1, \dots, \tilde{c}_n)$ is ϵ-differentially-private by Theorem 1. The β-faithfulness and the consistency properties is satisfied by constraint (O2) and (O3) respectively. The result follows from post-processing immunity (Theorem 2).

Model 5 The CBDP mechanism for the AC-OPF

$$\textbf{variables:}\quad S^g_i, V_i, \dot{S}^l_i \ \ \forall i \in N, \ \ S_{ij} \ \ \forall (i,j) \in E \cup E^R$$

$$\textbf{minimize:}\quad \|\dot{S}^l - \tilde{\boldsymbol{S}}^{\boldsymbol{l}}\|_2^2 \tag{s^l_1}$$

$$\textbf{subject to:}\quad (3) - (9)$$

$$|\sum_{i \in N} \boldsymbol{c}_{2i}(\Re(S^g_i))^2 + \boldsymbol{c}_{1i}\Re(S^g_i) + \boldsymbol{c}_{0i} - f^*| \leqslant \beta \tag{s_2}$$

$$\sum_{i \in N} \dot{S}^l_i = L \tag{s_3}$$

As mentioned earlier, additional constraints can be enforced, e.g., to ensure the consistency that the sums of individual quantities equals their associated aggregated quantity. In this case, the aggregated quantities must return private counts and a portion of the privacy budget must be used to answer such queries.

Theorem 5. *The optimal solution $\langle \hat{D}^+, \mathbf{x}^+ \rangle$ to the optimization model (O1–O3) satisfies $\|\hat{D}^+ - D\|_2 \leqslant 2\|\tilde{D} - D\|_2$,*

Proof. We have

$$\|\hat{D}^+ - D\|_2 \leqslant \|\hat{D}^+ - \tilde{D}\|_2 + \|\tilde{D} - D\|_2 \tag{27}$$

$$\leqslant 2\|\tilde{D} - D\|_2. \tag{28}$$

where the first inequality follows from the triangle inequality on norms and the second inequality follows from

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

by optimality of $\langle \hat{D}^+, \mathbf{x}^+ \rangle$ and the fact that $\langle D, \mathbf{x}^* \rangle$ is a feasible solution to constraints (O2) and (O3).

The following result follows from the optimality of the Laplace mechanism [16].

Corollary 1. *The CBDP mehanism is at most a factor 2 away from optimality.*

6 Application to the Optimal Power Flow

The CBDP optimization model for the (AC-OPF) is presented in Model 5. In addition to the variables of Model 5, it takes as inputs the variables $\dot{S}^l_i$ representing the post-processed values of the loads for each bus in $i \in N$. The optimization model minimizes the L_2-norm between the variables $\dot{S}^l \in \mathbb{R}^n$ and the noisy loads $\tilde{\boldsymbol{S}}^l \in \mathbb{R}^n$ resulting from the application of the Laplace mechanism to the original load values. Model 5 is subject to the same constraint of Model 1, with the addition of the β-faithfulness constraint (s_2) and the constraint enforcing consistency of the aggregated load values $L \in \mathbb{R}$ (s_3), which is typically public knowledge.

7 Experimental Results

This section presents an evaluation of the CBDP mechanism on the case study. It first presents the experimental setup and then compares the CBDP mechanism with the Laplace mechanism.

Data sets and Experimental Setup The experimental results concern the *NESTA* power network test cases (https://gdg.engin.umich.edu). The test cases comprise 44 networks whose number of buses ranges from 3 to 9241. This section categorizes them in *small* (networks with up to 100 buses), *medium* (networks with more than 100 buses and up to 2000 buses) and *large* (networks with more than 2000 buses).

In the following, D denotes the original data set and $\tilde{D}$ its private version (i.e., the data set resulting through the application of a DP mechanism). Moreover, $\mathbf{c}_m(D)$ and $\mathbf{c}_m(\tilde{D})$ denote the cost of the dispatch obtained by an OPF given model m (i.e., AC, QC, SOC, or DC) on the original data set D and on its private version $\tilde{D}$, respectively.

The results obtained by the AC-OPF and its relaxations/approximations (QC, SOC, and DC) are evaluated using both the original and the private data sets, analyzing the dispatch cost (c) and the optimality gap, i.e., the ratio $G_R(D) = \frac{|\mathbf{c}_R(D) - c_{AC}(D)|}{c_{AC}(D)}$, where $c_{AC}(D)$ and $c_R(D)$ denote the best-known solution cost of the problem instance and of the relaxation R over data set D.

The baseline $\mathcal{M}_{\text{Lap}}$ is the Laplace mechanism applied to each load of the network. To obtain a private version of the loads, $\mathcal{M}_{\text{Lap}}$ is first used to construct a private value for the active loads $p^l_i = \Re(S^l_i)$ as $\tilde{p}^l_i = p^l_i + \text{Lap}(100/\epsilon)$, $(\forall i \in N)$ where 100 is the change in MWs protected by Lipschitz DP. The reactive load $q^l_i = \Im(S^l_i)$ is set as $\tilde{q}^l_i = \tilde{p}^l_i\ r_i$, with $r_i = q^l_i / p^l_i$ is the power load factor and is considered to be public knowledge, as is natural in power systems. The CBDP mechanism $\mathcal{M}_C$ uses the output of the Laplace mechanism and the post-processing step to obtain the private loads $\dot{S}^l_i$ for all $i \in N$.

The mechanisms are evaluated for privacy budgets $\epsilon \in \{0.1, 1.0, 10.0\}$ and faithfulness parameter $\beta \in \{0.01, 1.0, 100.0\}$. Smaller values for ϵ increase privacy guarantees at the expense of more noise introduced by the Laplace mechanism. All experimental results are reported as the average of 30 runs. This gives a total of 47,520 experiments, which are analyzed below.

Error Analysis on the OPF Cost The first experimental result measure the error introduced by a mechanism as the average distance, in percentage, between the exact and the private costs of the OPF dispatches. The reported error is expressed as $\frac{|\mathbf{c}_m(D) - \mathbf{c}_m(\tilde{D})|}{\mathbf{c}_m(D)} \cdot 100$, for $m = \{AC, QC, SOC, DC\}$.

Figure 3 illustrates the error of the private mechanisms for varying privacy budgets. Rows show the results for the different network sizes (small, medium, and large), while columns show the results for the different faithfulness level values ($\beta = 0.01, 1.0, 100.0$). The results are shown in log scale. Each sub-figure also presents the results for three privacy budgets $(10, 1, 0.1)$.

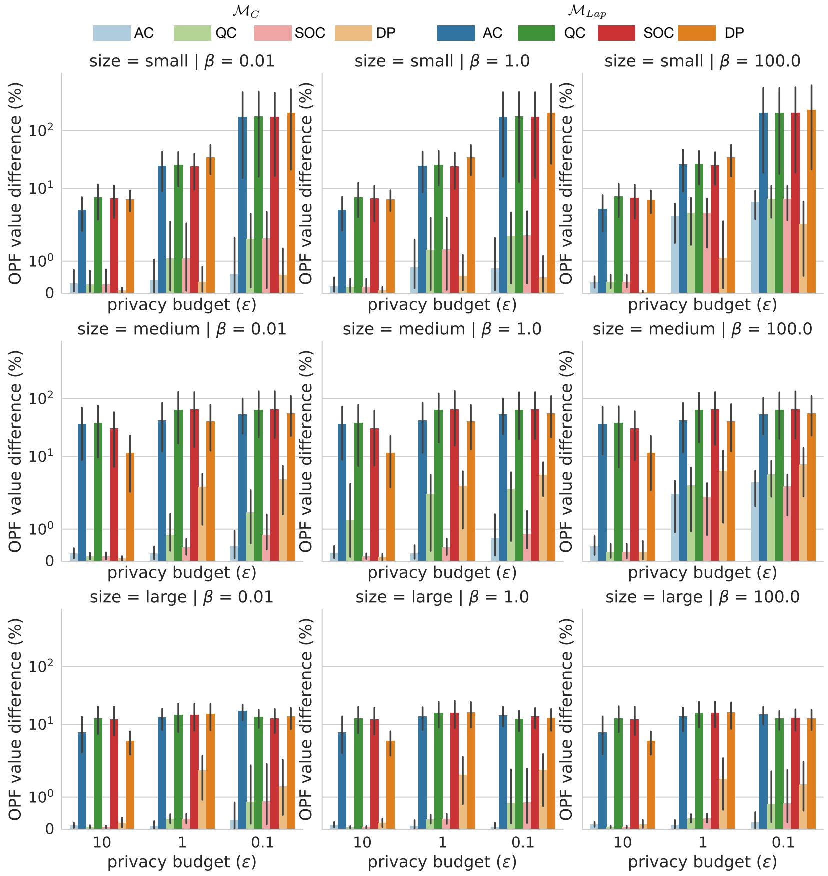

Fig. 3. Average OPF Objective Differences for the CBDP Mechanism (light colors) and the Laplace Mechanism (dark colors).

For all privacy budgets and all faithfulness-levels, the CBDC mechanism outperforms the Laplace mechanism by one to two orders of magnitude. For every OPF model, the Laplace mechanism produces OPF values which are, in general, more than 10% away (and exceeding 100% in many instances) from the values reported by the model ran on the original data, with the exception for the largest privacy budget, whose results produces differences slightly below 10%. In contrast, CBDP produces OPF values close (within 10%) to those produced on the original data, with all the OPF models adopted.

The errors reported by the Laplace mechanism on small networks are larger than those reported on medium and large networks. This is due to the fact that the dispatch costs of larger networks are typically much higher than those of small networks, and thus the relative distance of the error accumulated by the DP mechanism is more pronounced for the smaller test cases. Despite this, the CBDP mechanism produces solutions with small error costs, even for the small network instances.

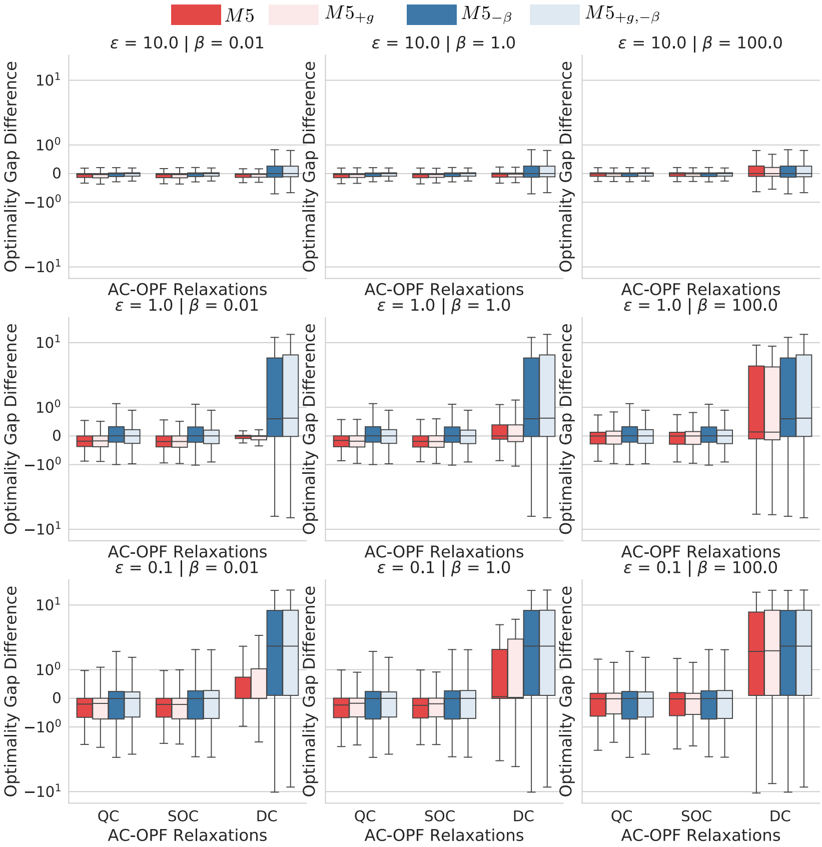

Fig. 4. Optimality GAP error for the QC, SOC, and DC, relaxations of the AC-OPF, on different CBDP post-processing models.

The CBDP mechanism preserves the objectives of the AC OPF problems accurately (within 1%), demonstrating its benefits for small beta values. The OPF value differences increase as the faithfulness parameter β increases.

Analysis of the Optimality Gap In a competition setting, it is also critical to preserve the computational difficulty of the original test case. The next set of results show how well CBDP preserves the optimality gap of the instance and its relationship to well-known approximations such as the DC model.

Figure 4 shows the differences $G_R(\tilde{D}) - G_R(D)$, for the AC OPF relaxation/approximation models $R = \{QC, SOC, DC\}$, where the private data set $\tilde{D}$ is produced by the CBDP mechanism. It compares four CBDP post-processing models: $M5$, which solves the CBDP of Model 5; $M5_{+g}$, which extends the Model 5 by modifying the objective (s_1^l) by adding the terms $\|S^g - \boldsymbol{S}^g\|_2^2$ to minimize the distance from the generator setpoints; $M5_{-\beta}$, which solves the Model 5without the beta faithfulness constraint (s_2); and $M5_{+g,-\beta}$, which excludes constraint (s_2) but includes the terms $\|S^g - \boldsymbol{S}^g\|_2^2$ into its objective. Rows show the results

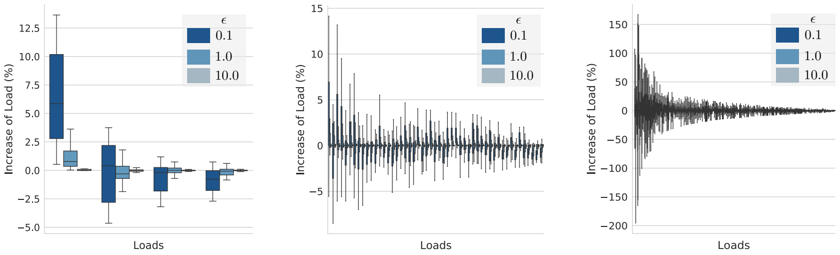

Fig. 5. Percentage of load increase in the 4-bus (left) 73-bus (center), and 300-bus (right) NESTA test cases.

for the different privacy faithfulness levels ($\beta = 0.01, 1.0, 100.0$), while columns show the results for the different privacy budgets $(10, 1, 0.1)$. The results are shown in log scale.

For all settings, $M5$ produces instances whose optimality gaps are close to those of the original ones (their distance is < 1 for $\epsilon = 10.0$, and 1.0, and < 3 for $\epsilon = 0.1$). $M5_{+g}$ produces very similar optimality gaps, showing that the CBDP mechanism does not need to take into account the generator setpoint values. For the relaxations, both $M5_{-\beta}$ and $M5_{+g,-\beta}$ produce quite similar results to $M5$. It is only for the DC-approximation that the faithfulness constraint is important. Moreover, note that the faithfulness constraint must be relatively tight even for $M5$ to preserve the result of the DC model. The DC model ignores many aspects of the power systems and hence it is not a surprise that it is more brittle. *These results indicate that the CBDP mechanism is capable of preserving the optimality gap of relaxations (and the quality of the DC approximation) with high fidelity.*

Analysis of the Private Network Loads The last set of results reports the effect of the CBDP mechanism on the network load profiles. Figure 5 depicts the percentage of load increase when applying the CBDP mechanism on three example networks with 4-buses (left), 73 buses (center), and 300 buses (right) for various privacy budgets. The results illustrate that load variation is often significant for a portion of the loads, although the CBDP mechanism preserves the problem structure accurately. Moreover, some of the loads exhibit a positive or negative bias. A detailed examination of these test cases reveals that this is due to the underlying network characteristics. For example, in the 4-bus test case case (Figure 5(left)), the first load (load 1) tends to be higher than its original value. This is explained by the fact that such load resides on the same bus as the cheaper generator which also has a very high generation capacity [9](pp. 337–338). As a result, the CBDP mechanism has significant flexibility to increase this load to redistribute the noise appropriately.

8 Related Work

There is rich literature on theoretical results of DP (see e.g., [6, 22]). The literature on DP applied to energy systems includes considerably fewer efforts. Ács and Castelluccia [2] exploited a direct application of the Laplace mechanism to hide user participation in smart meter data sets, achieving ϵ-DP. Zhao et al. [25] studied a DP schema that exploits the ability of households to charge/discharge a battery to hide the real energy consumption of their appliances. Liao et al. [18] introduce Di-PriDA, a privacy-preserving mechanism for appliance-level peak-time load balancing control in the smart grid, aimed at masking the consumption of the top-k appliances of a household.

Karapetyan et al. [15] conduct an empirical study on quantifying the trade-off between privacy and utility in demand response systems. The authors analyze the effects of a simple Laplace mechanism on the objective value of the demand response optimization problem. Their experiments on a 4-bus micro-grid show drastic results: the optimality gap approaches nearly 90% in some cases.

A DP schema that uses constrained post-processing was recently introduced by Fioretto et al. [8] and adopted to release private mobility data. In contrast, the proposed CBDP schema proposed in this work releases the private data set through a mechanism that imposes constraints to ensure the problem solution cost is close to the solution cost of the original problem, and that the underlying optimal power flow constraints are satisfiable.

9 Conclusions

This paper introduced the Constraint-Based Differential Privacy (CBDP) mechanism, an approach to Differential Privacy (DP) which aims at releasing optimal power flow benchmarks that retain the privacy of the customers (loads). CBDP leverages the post-processing immunity of DP to cast the production of a private data set as an optimization problem that redistributes the noise introduced by a randomized mechanism to satisfy problem-specific constraints.

The proposed mechanism enjoys desirable theoretical properties: It achieves ϵ-DP, ensures that the released data set can produce feasible solutions for the optimization problem of interest, and is a constant factor away from optimality. CBDP has been evaluated on the largest collection of OPF test cases available. Experimental results show that CBDP improves the accuracy of traditional approaches (e.g., the Laplace mechanism) by orders of magnitude and preserves some salient computational features of the test cases, such as the optimality gap. These results are significant and indicate that CBDP has the potential to become an important tool to release data sets for competition settings.

Although the paper focused on the applicability of CBDP to OPF problems, the proposed mechanism is general and can be used for other applications where a private data set is the input to a complex optimization problem.

Acknowledgments. The authors would like to thank the anonymous reviewers for their valuable comments. This research is partly funded by the ARPA-E Grid Data Program under Grant 1357-1530. The views and conclusions contained in this document are those of the authors only.

Bibliography

[1] Kaggle: Your home for data science. https://www.kaggle.com.

[2] G. Ács and C. Castelluccia. I have a dream!(differentially private smart metering). In *Information hiding*, volume 6958, pages 118–132. Springer, 2011.

[3] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 901–914. ACM, 2013.

[4] M. Backes, P. Berrang, A. Hecksteden, M. Humbert, A. Keller, and T. Meyer. Privacy in epigenetics: Temporal linkability of microrna expression profiles. In *USENIX Security Symposium*, pages 1223–1240, 2016.

[5] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, volume 3876, pages 265–284. Springer, 2006.

[6] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Theoretical Computer Science*, 9(3-4):211–407, 2013.

[7] G. Fanti, V. Pihur, and Ú. Erlingsson. Building a rappor with the unknown: Privacy-preserving learning of associations and data dictionaries. *Proceedings on Privacy Enhancing Technologies*, 2016(3):41–61, 2016.

[8] F. Fioretto, C. Lee, and P. V. Hentenryck. Constrained-based differential privacy for private mobility. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, page (to appear), 2018.

[9] J. J. S. Grainger, W. D. J. J. Grainger, and W. D. Stevenson. *Power system analysis*. 1994.

[10] A. Greenberg. Apples differential privacyis about collecting your databut not your data. *Wired, June*, 2016.

[11] Gurobi. Gurobi software. http://www.gurobi.com/.

[12] 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.

[13] IBM. ILOG CPLEX software. http://www.ibm.com/.

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

[15] A. Karapetyan, S. K. Azman, and Z. Aung. Assessing the privacy cost in centralized event-based demand response for microgrids. *CoRR*, abs/1703.02382, 2017.

[16] F. Koufogiannis, S. Han, and G. J. Pappas. Optimality of the laplace mechanism in differential privacy. *arXiv preprint arXiv:1504.00065*, 2015.

[17] K. Lehmann, A. Grastien, and P. Van Hentenryck. Ac-feasibility on tree networks is np-hard. *IEEE Transactions on Power Systems*, 99:1–4, March 2015.

[18] X. Liao, P. Srinivasan, D. Formby, and A. R. Beyah. Di-prida: Differentially private distributed load balancing control for the smart grid. *IEEE Transactions on Dependable and Secure Computing*, 2017.

[19] G. McCormick. Computability of global solutions to factorable nonconvex programs: Part i - convex underestimating problems. *Mathematical Programming*, 10:146–175, 1976.

[20] D. J. Mir, S. Isaacman, R. Cáceres, M. Martonosi, and R. N. Wright. Dp-where: Differentially private modeling of human mobility. In *Big Data, 2013 IEEE International Conference on*, pages 580–588. IEEE, 2013.

[21] MOSEK ApS. *The MOSEK optimization toolbox.*, 2015.

[22] S. Vadhan. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*, pages 347–450. Springer, 2017.

[23] A. Verma. *Power grid security analysis: An optimization approach.* PhD thesis, Columbia University, 2009.

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

[25] J. Zhao, T. Jung, Y. Wang, and X. Li. Achieving differential privacy of data disclosure in the smart grid. In *INFOCOM, 2014 Proceedings IEEE*, pages 504–512. IEEE, 2014.