Differential Privacy of Aggregated DC Optimal Power Flow Data

Fengyu Zhou, James Anderson, and Steven H. Low

Abstract— **This paper introduces a differential privacy model for releasing optimal power flow (OPF) data for power systems. As an application, we consider an aggregation mechanism that provides only area-wide summary statistics of individual load and generation data. We define the notion of monotonicity of an OPF operator and used it to determine the amount of noise needed to preserve differential privacy under the aggregation mechanism. We also show that the level of differential privacy is independent of the number of aggregation regions and the number of buses in a region.**

I. INTRODUCTION

Realistic and publicly available power network models based on real data are important for the research community. One of the difficulties in developing such a model is that grid operators are reluctant to disclose consumer data or any information that may be commercially sensitive. Differential privacy, first developed in [1], [2], [3], has been widely used to evaluate the privacy loss for individual users in a dataset. It has recently been used by the researchers in the power systems community for use in applications such as distributed algorithms for EV charging [4] and power system data release [5].

In our work, we consider the differential privacy of power systems induced by an Optimal Power Flow (OPF) problem. In this context, the optimal generation can be viewed as a function of the loads. Typically generation data are publicly available. In contrast, load data can reveal consumer habits and other commercially sensitive information, and thus we aim to keep them private. We aim to prevent changes in generation data from disclosing sensitive load data. Instead of proposing new mechanisms and algorithms, we derive how much noise is required to be added to the data in order to achieve a certain level of differential privacy for a given query. Using the *aggregation mechanism* as an example, we introduce the concept of (δ,ε)-monotonicity, a metric that is central to our differential privacy analysis. We also show how it is affected under different system topologies. Finally we present the examples of three systems with different topologies and thus different monotonic characterizations, i.e., different (δ,ε) parameters. For each system we show that to preserve the same level of differential privacy, the required amounts of noise implied by our theorem are very different for each example. We hope that such theoretical guarantees will not only guide the design of differentially private power systems, but also encourage greater data sharing and cooperation between grid operators and academia in the future.

This work is funded by NSF grants CCF 1637598, ECCS 1619352, CNS 1545096, ARPA-E through grant DE-AR0000699 and the GRID DATA program, and DTRA through grant HDTRA 1-15-1-0003.

Fengyu Zhou is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, 91125. Email: f.zhou@caltech.edu

James Anderson is with the Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, 91125. Email: james@caltech.edu

Steven H. Low is with Department of Electrical Engineering and the Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, 91125. Email: slow@caltech.edu

II. BACKGROUND

In this section we define the power network model and the optimal power flow problem we consider in this paper. We introduce some assumptions on allowable parameter sets and show that these assumptions are mild.

Notation

Vectors and matrices are typically written in bold while scalars are not. Given two vectors $\mathbf{a},\mathbf{b}\in\mathbb{R}^n$, $\mathbf{a}\geq\mathbf{b}$ denotes the element-wise partial order $\mathbf{a}_i\geq\mathbf{b}_i$ for $i=1,\ldots,n$. For a scalar k, we define the projection operator $[k]^-:=\min\{0,k\}$. We define $\|\mathbf{x}\|_0$ as the number of non-zero elements of the vector $\mathbf{x}$. For $\mathbf{X}\in\mathbb{R}^{n\times m}$, the restriction $\mathbf{X}_{\{1,3,5\}}$ denotes the $3\times m$ matrix composed of stacking rows $1,3$, and 5 on top of each other. We will frequently use a set to describe the rows we wish to form the restriction from, in this case we assume the elements of the set are arranged in increasing order. We will use $\mathbf{e}_m$ to denote the standard base for the m^{th} coordinate, its dimension will be clear from the context. Finally, let $[m]:=\{1,2,\ldots,m\}$ and $[n,m]:=\{n,n+1,\ldots,m\}$.

A. System Model

Consider a power network modeled by an undirected graph $\mathcal{G}(\mathcal{V},\mathcal{E})$, where $\mathcal{V}:=\mathcal{V}_{\mathrm{G}}\cup\mathcal{V}_{\mathrm{L}}$ denotes the set of buses which can be further classified into generators in set $\mathcal{V}_{\mathrm{G}}$ and loads in set $\mathcal{V}_{\mathrm{L}}$, and $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ is the set of all branches linking those buses. We will later use the terms (graph, vertex, edge) and (power network, bus, branch) interchangeably. Suppose $\mathcal{V}_{\mathrm{G}}\cap\mathcal{V}_{\mathrm{L}}=\emptyset$ and there are $|\mathcal{V}_{\mathrm{G}}|=:N_{\mathrm{G}}$ generator and $|\mathcal{V}_{\mathrm{L}}|=:N_{\mathrm{L}}$ loads, respectively. For simplicity, let $\mathcal{V}_{\mathrm{G}}=[N_{\mathrm{G}}]$, $\mathcal{V}_{\mathrm{L}}=[N_{\mathrm{G}}+1,N_{\mathrm{G}}+N_{\mathrm{L}}]$. Let $N=N_{\mathrm{G}}+N_{\mathrm{L}}$. Without loss of generality, $\mathcal{G}$ is a connected graph with $|\mathcal{E}|=:E$ edges labelled as $1,2,\ldots,E$. Let $\mathbf{C}\in\mathbb{R}^{N\times E}$ be the signed incidence matrix. Suppose edge e connects vertices u and v $(u<v)$, then $\mathbf{C}_{u,e}=1$, $\mathbf{C}_{v,e}=-1$, and all other entries in the e^{th} column of $\mathbf{C}$ are 0. Let $\mathbf{B}=\mathrm{diag}(b_1,b_2,\ldots,b_E)$, where $b_e>0$ is the susceptance for branch e. As we adopt a DC power flow model, all branches are assumed lossless. Further, we denote the generation and load as $\mathbf{s}^g\in\mathbb{R}^{N_{\mathrm{G}}}$, $\mathbf{s}^l\in\mathbb{R}^{N_{\mathrm{L}}}$, respectively. Thus $\mathbf{s}^g_i$ refers to the generation on bus i while $\mathbf{s}^l_i$ refers to the load on bus $N_\mathrm{G}+i$. We will refer to bus $N_\mathrm{G}+i$ simply as load i for simplicity. The power flow on branch $e \in \mathcal{E}$ is denoted as $\mathbf{p}_e$, and $\mathbf{p} := [\mathbf{p}_1, \ldots, \mathbf{p}_E]^\mathsf{T} \in \mathbb{R}^E$ is the vector of all branch power flows. The following assumption is made to simplify the analysis.

Assumption 1: There are no buses in the network that are both loads and generators. Formally, $\mathcal{V}_\mathrm{G} \cap \mathcal{V}_\mathrm{L} = \emptyset$.

The above assumption is not restrictive. We can always split a bus with both a generator and a load into a bus with only the generator connected to another bus with only the load, and connect all the neighbors of the original bus to that load bus.

B. Optimal Power Flow

We focus on the DC OPF problem with a linear cost function [6]. That is to say, the voltage magnitudes are assumed to be fixed and known. Without loss of generality, we assume all the voltage magnitudes to be 1. The decision variables are the voltage angles denoted by vector $\boldsymbol{\theta} \in \mathbb{R}^N$ and power generations $\mathbf{s}^g$, given loads $\mathbf{s}^l$. The DC OPF takes the following form:

$$\begin{aligned}
\underset{\mathbf{s}^g,\boldsymbol{\theta}}{\text{minimize}} \quad & \mathbf{f}^\mathsf{T}\mathbf{s}^g && \text{(1a)}\\
\text{subject to} \quad & \boldsymbol{\theta}_1 = 0 && \text{(1b)}\\
& \mathbf{CBC}^\mathsf{T}\boldsymbol{\theta} = \begin{bmatrix} \mathbf{s}^g \\ -\mathbf{s}^l \end{bmatrix} && \text{(1c)}\\
& \underline{\mathbf{s}}^g \le \mathbf{s}^g \le \overline{\mathbf{s}}^g && \text{(1d)}\\
& \underline{\mathbf{p}} \le \mathbf{BC}^\mathsf{T}\boldsymbol{\theta} \le \overline{\mathbf{p}}. && \text{(1e)}
\end{aligned}$$

Here, $\mathbf{f} \in \mathbb{R}^{N_\mathrm{G}}$ is the unit cost for each generator, and bus 1 is selected as the slack bus with fixed voltage angle 0. In (1c), we let the injections for generators be positive while the injections for loads be the negation of $\mathbf{s}^l$. The upper and lower limits for the generation are set as $\overline{\mathbf{s}}^g$ and $\underline{\mathbf{s}}^g$, respectively, and $\overline{\mathbf{p}}$ and $\underline{\mathbf{p}}$ are the limits for branch power flow. We assume that (1) is well posed, i.e. $\overline{\mathbf{s}}^g > \underline{\mathbf{s}}^g$, $\overline{\mathbf{p}} > \underline{\mathbf{p}}$.

Let $\boldsymbol{\tau} \in \mathbb{R}^{N+1}$ be the vector of Lagrangian multipliers associated with equality constraints (1b), (1c), and $(\boldsymbol{\lambda}_+, \boldsymbol{\lambda}_-)$ and $(\boldsymbol{\mu}_+, \boldsymbol{\mu}_-)$ be the Lagrangian multipliers associated with inequalities (1d) and (1e) respectively. As (1) is a linear program [7], the following KKT condition holds at an optimal point when (1) is feasible.

$$\begin{aligned}
& \text{(1b)} - \text{(1e)} && \text{(2a)}\\
& \mathbf{0} = \mathbf{M}^\mathsf{T}\boldsymbol{\tau} + \mathbf{CB}(\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-) && \text{(2b)}\\
& -\mathbf{f} = -[\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_{N_\mathrm{G}}]^\mathsf{T} + \boldsymbol{\lambda}_+ - \boldsymbol{\lambda}_- && \text{(2c)}\\
& \boldsymbol{\mu}_+, \boldsymbol{\mu}_-, \boldsymbol{\lambda}_+, \boldsymbol{\lambda}_- \ge 0 && \text{(2d)}\\
& \boldsymbol{\mu}_+^\mathsf{T}(\mathbf{BC}^\mathsf{T}\boldsymbol{\theta} - \overline{\mathbf{p}}) = \boldsymbol{\mu}_-^\mathsf{T}(\underline{\mathbf{p}} - \mathbf{BC}^\mathsf{T}\boldsymbol{\theta}) = 0 && \text{(2e)}\\
& \boldsymbol{\lambda}_+^\mathsf{T}(\mathbf{s}^g - \overline{\mathbf{s}}^g) = \boldsymbol{\lambda}_-^\mathsf{T}(\underline{\mathbf{s}}^g - \mathbf{s}^g) = 0, && \text{(2f)}
\end{aligned}$$

where

$$\mathbf{M} := \begin{bmatrix} \mathbf{CBC}^\mathsf{T} \\ 1\ 0\ 0\ \cdots\ 0 \end{bmatrix}$$

is an $(N+1)$-by-N matrix with rank N. Condition (2a) corresponds to primal feasibility, condition (2d) corresponds to dual feasibility, conditions (2e), (2f) correpond to complementary slackness, and conditions (2b), (2c) correspond to stationarity [8].

C. OPF Operator

Henceforth we fix the topology and susceptances of the power network. Let $\boldsymbol{\xi} := [(\overline{\mathbf{s}}^g)^\mathsf{T}, (\underline{\mathbf{s}}^g)^\mathsf{T}, \overline{\mathbf{p}}^\mathsf{T}, \underline{\mathbf{p}}^\mathsf{T}]^\mathsf{T} \in \mathbb{R}^{2N_\mathrm{G}+2E}$ be the vector of system limits, and let Ω_para be the set of all $\boldsymbol{\xi}$ such that:[1]

- $\exists \mathbf{s}^l \in \mathbb{R}^{N_\mathrm{L}}$, such that (1b)-(1e) are feasible;
- $\underline{\mathbf{s}}^g \ge 0$.

We assume closure(interior(Ω_para)) $=$ closure(Ω_para).[2] For each $\boldsymbol{\xi} \in \Omega_\mathrm{para}$, let $\Omega_{\mathbf{s}^l}(\boldsymbol{\xi})$ be the corresponding set of $\mathbf{s}^l$ such that

- (1b)-(1e) are feasible.[3]

Assume that closure(interior($\Omega_{\mathbf{s}^l}$)) $=$ closure($\Omega_{\mathbf{s}^l}$) and $\Omega_{\mathbf{s}^l}$ is convex and nonempty.[4] When we fix $\boldsymbol{\xi}$ and there is no confusion, we use $\Omega_{\mathbf{s}^l}$ instead. We now define the operator $\mathcal{OPF}$, which will be used throughout the rest of the paper.

Definition 1: Let the set valued operator $\mathcal{OPF} : \Omega_{\mathbf{s}^l} \to 2^{\mathbb{R}^{N_\mathrm{G}}}$ be the mapping such that $\mathcal{OPF}(\mathbf{x})$ is the set of optimal solutions to (1) with parameter $\mathbf{s}^l = \mathbf{x}$.

Now fix $\mathbf{B}, \mathbf{C}$ and $\boldsymbol{\xi}$, let $\Omega_\mathbf{f}$ be the set of $\mathbf{f}$ such that $\forall \mathbf{s}^l \in \Omega_{\mathbf{s}^l}$,

- (1) has a unique solution;
- all solutions to (2) satisfy

$$\|\boldsymbol{\mu}_+\|_0 + \|\boldsymbol{\mu}_-\|_0 + \|\boldsymbol{\lambda}_+\|_0 + \|\boldsymbol{\lambda}_-\|_0 \ge N_\mathrm{G} - 1. \quad (3)$$

Proposition 1: $\Omega_\mathbf{f}$ is dense in $\mathbb{R}^{N_\mathrm{G}}$.

Proof: See Appendix A. ∎

Proposition 1 shows that for a fixed network, it is easy to find an objective vector $\mathbf{f}$ such that (1) will always have a unique solution for a reasonable $\mathbf{s}^l$.

Assumption 2: The objective vector $\mathbf{f}$ is in $\Omega_\mathbf{f}$, i.e, $\mathbf{f}$ always guarantees the uniqueness of the solution to (1) for all $\mathbf{s}^l \in \Omega_{\mathbf{s}^l}$.

When Assumption 2 does not hold, Proposition 1 implies that we can always perturb $\mathbf{f}$ a little such that the assumption is valid.

Remark 1: Under Assumption 2, the value of $\mathcal{OPF}$ is always a singleton, so we can consider $\mathcal{OPF}(\mathbf{x})$ as the function mapping from $\mathbf{x}$ to the unique optimal solution of (1) with parameter $\mathbf{s}^l = \mathbf{x}$. Since the solution set to the parametric linear program is both upper and lower hemicontinuous [9], $\mathcal{OPF}$ is continuous as well.

[1] If there are other constraints on the system limits $\boldsymbol{\xi}$ implied by a physical system, they can be added to the list.

[2] This condition will be satisfied if

$$\Omega_\mathrm{para} \quad := \quad \prod \Omega_{\overline{\mathbf{s}}^g_i} \times \prod \Omega_{\underline{\mathbf{s}}^g_i} \times \prod \Omega_{\overline{\mathbf{p}}_i} \times \prod \Omega_{\underline{\mathbf{p}}_i}$$

where $\Omega_{\overline{\mathbf{s}}^g_i}$, $\Omega_{\underline{\mathbf{s}}^g_i}$, $\Omega_{\overline{\mathbf{p}}_i}$ and $\Omega_{\underline{\mathbf{p}}_i}$ are the set of limits $\overline{\mathbf{s}}^g_i, \underline{\mathbf{s}}^g_i, \overline{\mathbf{p}}_i, \underline{\mathbf{p}}_i$ respectively. This assumes that these limits are independent of each other.

[3] Since the feasible domain of (1) is compact, the optimal solution always exists.

[4] This condition will be satisfied if $\Omega_{\mathbf{s}^l} := \prod \Omega_{\mathbf{s}^l_i}$, where $\Omega_{\mathbf{s}^l_i}$ are the set of $\mathbf{s}^l_i$. This assumes that these loads are independent of each other.

Later analysis is made easier if the set of binding (active) constraints is independent. Here, binding constraints refer to the set of equality constraints (1b), (1c) *and* those inequality constraints (1d), (1e) for which either the upper or lower-bounds are active. Grouping the coefficients of these constraints into a single matrix $\mathbf{Z}$ we refer to them as being independent if $\mathbf{Z}$ is full-rank. Let $\widetilde{\Omega}_{\mathbf{s}^l}(\boldsymbol{\xi})$ be the set

$$\{\mathbf{s}^l \in \Omega_{\mathbf{s}^l}(\boldsymbol{\xi}) \mid (1) \text{ has exactly } N_{\mathrm{G}}-1 \text{ binding inequalities}\}$$

Then we have the following proposition.

Proposition 2: Let $\widetilde{\Omega}_{\mathrm{para}} \subseteq \Omega_{\mathrm{para}}$ be the set such that $\forall \boldsymbol{\xi} \in \widetilde{\Omega}_{\mathrm{para}}$, the set $\widetilde{\Omega}_{\mathbf{s}^l}(\boldsymbol{\xi})$ is dense in $\Omega_{\mathbf{s}^l}(\boldsymbol{\xi})$. Then $\widetilde{\Omega}_{\mathrm{para}}$ is dense in Ω_{para}.

Proof: See Appendix B. ■

Finally, Assumption 3 restricts the value of limits in (1) in order to further simplify our analysis.

Assumption 3: The parameter $\boldsymbol{\xi}$ for the limits of generations and branch power flows is assumed to be in $\widetilde{\Omega}_{\mathrm{para}}$, as defined in Proposition 2.

If Assumption 3 does not hold, Proposition 2 implies that we can always perturb $\boldsymbol{\xi}$ such that the assumption holds. The proof of Proposition 2 can directly extend to the following two corollaries, which we make use of in Appendix C. Proofs are omitted due to the limited space.

Corollary 1: In Proposition 2, $\Omega_{\mathbf{s}^l} \setminus \widetilde{\Omega}_{\mathbf{s}^l}$ can be covered by the union of finitely many subspaces.

Corollary 2: For any $\mathbf{s}^l \in \widetilde{\Omega}_{\mathbf{s}^l}$, the $N_{\mathrm{G}}-1$ tight inequalities in (1), along with $N+1$ equality constraints, are independent.

III. Differential Privacy

In this section, we use differential privacy to evaluate the security status of the power system data disclosed to the public. In general, a differentially private dataset can protect the privacy for each individual user by adding noise to the statistics of the dataset such that the change in a single record cannot be effectively detected [1], [2], [3]. Suppose $\mathcal{D}^n$ is the data space for n users, and for data $\mathbf{d} \in \mathcal{D}^n$, a mechanism $\mathcal{M}: \mathcal{D}^n \to \mathbb{R}^r$ is a randomized function of $\mathbf{d}$, which releases some information of $\mathbf{d}$ with the addition of some randomized noise. The ϱ-differential privacy is defined as follows [1].

Definition 2: The mechanism $\mathcal{M}$ can preserve the ϱ-differential privacy if and only if $\forall \mathbf{d}', \mathbf{d}'' \in \mathcal{D}^n$ such that $\|\mathbf{d}'-\mathbf{d}''\|_0 \leq 1$, and $\forall \mathcal{W} \subseteq \mathbb{R}^r$, we have

$$\mathbb{P}\{\mathcal{M}(\mathbf{d}') \in \mathcal{W}\} \leq \exp(\varrho) \cdot \mathbb{P}\{\mathcal{M}(\mathbf{d}'') \in \mathcal{W}\}.$$

A *query* is a function $\tilde{\mathcal{M}}: \mathcal{D}^n \to \mathbb{R}^r$. Examples include 'count' functions, e.g. return the number of records in the database where property y holds ($r=1$). Other examples include statistical queries such as computing mean and variance. One realization of the *mechanism* $\mathcal{M}$ can be interpreted as follows. Suppose the *query* to the database $\mathbf{d} \in \mathcal{D}^n$ is given by some function $\tilde{\mathcal{M}}$. The *true* answer is thus given by $\tilde{\mathcal{M}}(\mathbf{d})$. Then the mechanism $\mathcal{M}$ returns the value $\mathcal{M}(\mathbf{d}) := \tilde{\mathcal{M}}(\mathbf{d}) + \mathbf{Y}$ for an appropriately chosen noise $\mathbf{Y}$. A mechanism $\mathcal{M}$ that satisfies the properties of Definition 2 ensures that the addition or removal of a single entry to the database does not change (much) the outcome of the query.

A popular mechanism relies on the symmetric Laplace distribution $\mathscr{L}(\cdot)$. For a random variable $X \sim \mathscr{L}(b)$ the probability density function is given by

$$f_X(x|b) = \frac{1}{2b}\exp\left(\frac{-|x|}{b}\right),$$

and X has variance $\sigma^2 = 2b^2$. Intuitively, as b increases, the distribution flattens and spreads symmetrically about the origin. The Laplace mechanism is defined by $\mathcal{M}(\mathbf{d}) := \tilde{\mathcal{M}}(\mathbf{d}) + \mathbf{Y}$ where $Y_i \sim \mathscr{L}(\Delta_1/\varrho)$ for $i=1,\ldots,r$ and Δ_1 is the L_1-sensitivity of the query $\tilde{\mathcal{M}}$:

$$\Delta_1 = \underset{\|\mathbf{d}'-\mathbf{d}''\|_1=1}{\text{maximize}} \quad \|\tilde{\mathcal{M}}(\mathbf{d}') - \tilde{\mathcal{M}}(\mathbf{d}'')\|_1. \tag{4}$$

The following theorem explains the importance of the Laplace mechanism [1]:

Theorem 1: For $\tilde{\mathcal{M}}: \mathcal{D}^n \to \mathbb{R}^r$, the Laplace mechanism defined by $\mathscr{L}(\Delta_1/\varrho)$ provides ϱ-differential privacy.

From the theorem and the definition of the Laplace distribution, it can be seen that for a fixed privacy level (specified by ϱ), as the sensitivity increases the mechanism responds by adding larger amounts of noise. Fortunately, many queries of interest have low sensitivity; e.g., counting queries and sum-separable functions have $\Delta_1 = 1$,

A. OPF Data Privacy

Ideally both the generations $\mathbf{s}^g$ and loads $\mathbf{s}^l$ are available for the research community to build realistic power system models from. However, load data may contain sensitive information, and hence it is desirable to preserve the privacy of $\mathbf{s}^l$.

Suppose $\mathcal{M}(\mathbf{s}^g, \mathbf{s}^l)$ is a (randomized) function of $(\mathbf{s}^g, \mathbf{s}^l)$, and acts as the mechanism of the data. It is reasonable to assume that $\mathbf{s}^g$ is always chosen as the unique optimal solution to the OPF problem, i.e., $\mathbf{s}^g = \mathcal{OPF}(\mathbf{s}^l)$. Then we can write $\mathcal{M}(\mathbf{s}^g, \mathbf{s}^l)$ as $\mathcal{M}(\mathcal{OPF}(\mathbf{s}^l), \mathbf{s}^l)$. The privacy problem is to design a mechanism that hides individual load changes when the database $\mathbf{d}$ containing the vectors $\mathbf{s}^g, \mathbf{s}^l$ is queried. We let Δ denote the changes to an individual load, i.e., $\mathbf{s}^l_i \leftarrow \mathbf{s}^l_i \pm \Delta$ for some $\Delta > 0$. To address this problem, we introduce a modified version of differential privacy:

Definition 3: For $\Delta, \varrho > 0$, the mechanism $\mathcal{M}$ preserves (Δ, ϱ)-differential privacy if and only if $\forall (\mathbf{s}^l)', (\mathbf{s}^l)''$ such that $\|(\mathbf{s}^l)' - (\mathbf{s}^l)''\|_0 \leq 1$ and $\|(\mathbf{s}^l)' - (\mathbf{s}^l)''\|_1 \leq \Delta$, and $\forall \mathcal{W} \subseteq \mathbb{R}^r$, we have

$$\begin{aligned} &\mathbb{P}\{\mathcal{M}(\mathcal{OPF}((\mathbf{s}^l)'), (\mathbf{s}^l)') \in \mathcal{W}\} \\ &\leq \exp(\varrho) \cdot \mathbb{P}\{\mathcal{M}(\mathcal{OPF}((\mathbf{s}^l)''), (\mathbf{s}^l)'') \in \mathcal{W}\}. \end{aligned}$$

Theorem 1 can be readily extended to our (Δ, ϱ)-differential privacy.

Lemma 1: Suppose $\tilde{\mathcal{M}}(\mathbf{s}^g, \mathbf{s}^l)$ is a deterministic function of $(\mathbf{s}^g, \mathbf{s}^l)$ satisfying that $\forall (\mathbf{s}^l)', (\mathbf{s}^l)''$ such that $\|(\mathbf{s}^l)' - (\mathbf{s}^l)''\|_0 \leq 1$ and $\|(\mathbf{s}^l)' - (\mathbf{s}^l)''\|_1 \leq \Delta$, we have

$$\|\tilde{\mathcal{M}}(\mathcal{OPF}((\mathbf{s}^l)'), (\mathbf{s}^l)') - \tilde{\mathcal{M}}(\mathcal{OPF}((\mathbf{s}^l)''), (\mathbf{s}^l)'')\|_1 \leq \Delta_1.$$

Then the mechanism $\mathcal{M} = \tilde{\mathcal{M}} + \mathbf{Y}$, where all Y_i are drawn from independent Laplace distributions $\mathscr{L}(\Delta_1/\varrho)$, preserves $(\Delta, \varrho)-$differential privacy.

In the next section, we will introduce the *aggregation mechanism* as an example for estimating the L_1-sensitivity.

IV. Case Study: Aggregation Mechanism

A. Aggregated injections

In [10], we propose a method to release load and generation data in a way that attempts to strike a balance between the privacy of data owners and the need of the research community for realistic samples. The method consists of two steps. First, instead of $\mathbf{s}^g$ and $\mathbf{s}^l$, the data owner releases their aggregations over discrete regions of the network. Second, a disaggregation algorithm is used to estimate the loads and generations based on the released aggregated data. Typically, generation data are publicly available. Load data however may contain sensitive information, and thus should be protected. In this section, we study how differential privacy is preserved under the aggregation mechanism. See [5] for another approach.

Suppose the buses in $\mathcal{V}$ are partitioned into r regions $\mathcal{R}_1, \mathcal{R}_2 \ldots, \mathcal{R}_r$, where $\mathcal{R}_i \subseteq \mathcal{V}$ is the set of bus IDs in region i. Let the aggregated injections for region i be

$$\tilde{\mathcal{M}}_i^g = \sum_{j \in \mathcal{R}_i} {\mathbf{s}^g}_j, \quad \tilde{\mathcal{M}}_i^l = \sum_{j+N_{\mathrm{G}} \in \mathcal{R}_i} {\mathbf{s}^l}_j,$$

The utility companies disclose a noisy version of the aggregated injections, denoted as $\mathcal{M}_i^g = \tilde{\mathcal{M}}_i^g + Y_i^g$ and $\mathcal{M}_i^l = \tilde{\mathcal{M}}_i^l + Y_i^l$. Here, Y_i^g and Y_i^l are independent random variables and are intentionally added to ensure privacy. Let

$$\mathcal{M}(\mathbf{s}^g, \mathbf{s}^l) = [\mathcal{M}_1^g, \ldots, \mathcal{M}_r^g, \mathcal{M}_1^l, \ldots, \mathcal{M}_r^l]^{\mathsf{T}} \tag{5}$$

be the aggregation mechanism on the power injections. To study the level of differential privacy that (5) can achieve, we will first define a notion of power network monotonicity in the next subsection.

B. System Monotonicity

System monotonicity characterizes how the optimal generation reacts to a change in load. It provides information about the L_1-sensitivity.

Definition 4: A power system is said to be *monotone if* $\forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \Omega_{\mathbf{s}^l}$ *such that* $\boldsymbol{\alpha} \geq \boldsymbol{\beta}$ *and* $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_0 = 1$, *we have* $\mathcal{OPF}(\boldsymbol{\alpha}) \geq \mathcal{OPF}(\boldsymbol{\beta})$.

Recall that the inequalities in Definition 4 are element-wise, i.e., a system is monotone if *all* generations will increase or remain unchanged when any single load increases. This is often too stringent a requirement. We are interested in approximately monotone systems, formalized in the following definition.

Definition 5: For $\delta > 0, \varepsilon \geq 0$, a power system is said to be (δ, ε)*-monotone if* $\forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \Omega_{\mathbf{s}^l}$ *such that* $\boldsymbol{\beta} + \delta \cdot \mathbf{1} \geq \boldsymbol{\alpha} \geq \boldsymbol{\beta}$ *and* $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_0 = 1$, *we have* $\sum_{i=1}^{N_{\mathrm{G}}} [\mathcal{OPF}_i(\boldsymbol{\alpha}) - \mathcal{OPF}_i(\boldsymbol{\beta})]^- \geq -\varepsilon$.

By definition, a monotone system is always $(\delta, 0)$-monotone for any positive δ. In the next sections, we will first study the derivative of $\mathcal{OPF}$ and then relate $\mathcal{OPF}$ derivative to monotonicity.

C. Privacy of Aggregation Mechanism

Lemma 1 and Definitions 4 and 5 on monotonicity immediately imply the following two properties of the aggregation mechanism (5).

Theorem 2: Suppose the power system is monotone. The aggregation mechanism (5) where Y_i^g and Y_i^l are drawn i.i.d. from $\mathscr{L}(2\Delta/\varrho)$ preserves (Δ, ϱ)-differential privacy.

Proof: In view of Lemma 1, it is sufficient to show that for any $(\mathbf{s}^l)', (\mathbf{s}^l)''$ with $\|(\mathbf{s}^l)' - (\mathbf{s}^l)''\|_0 \leq 1$ and $\|(\mathbf{s}^l)' - (\mathbf{s}^l)''\|_1 \leq \Delta$, we have

$$\|\tilde{\mathcal{M}}(\mathcal{OPF}((\mathbf{s}^l)'), (\mathbf{s}^l)') - \tilde{\mathcal{M}}(\mathcal{OPF}((\mathbf{s}^l)''), (\mathbf{s}^l)'')\|_1 \leq 2\Delta.$$

Suppose $(\mathbf{s}^l)' \geq (\mathbf{s}^l)''$ (taken element-wise). Then by Definition 4, we have $\mathcal{OPF}((\mathbf{s}^l)') \geq \mathcal{OPF}((\mathbf{s}^l)'')$ and thereby $(\mathbf{s}^g)' \geq (\mathbf{s}^g)''$. Since $\mathcal{OPF}$ is constructed w.r.t a DC model, the system is lossless and hence $\sum_{i=1}^{N_{\mathrm{G}}} \mathbf{s}_i^g = \sum_{i=1}^{N_{\mathrm{L}}} \mathbf{s}_i^l$. As a result,

$$\|(\mathbf{s}^g)' - (\mathbf{s}^g)''\|_1 = \|(\mathbf{s}^l)' - (\mathbf{s}^l)''\|_1 \leq \Delta$$

and

$$\begin{aligned} &\|\tilde{\mathcal{M}}(\mathcal{OPF}((\mathbf{s}^l)'), (\mathbf{s}^l)') - \tilde{\mathcal{M}}(\mathcal{OPF}((\mathbf{s}^l)''), (\mathbf{s}^l)'')\|_1 \\ \leq &\|(\mathbf{s}^g)' - (\mathbf{s}^g)''\|_1 + \|(\mathbf{s}^l)' - (\mathbf{s}^l)''\|_1 \leq 2\Delta. \end{aligned}$$

■

Theorem 3: Suppose the power system is (Δ, ε)-monotone. The aggregation mechanism (5) where Y_i^g and Y_i^l are drawn i.i.d. from $\mathscr{L}(2(\Delta + \varepsilon)/\varrho)$ preserves (Δ, ϱ)-differential privacy.

Proof: By Definition 5, we have

$$\|(\mathbf{s}^g)' - (\mathbf{s}^g)''\|_1 \leq \|(\mathbf{s}^l)' - (\mathbf{s}^l)''\|_1 + 2\varepsilon \leq \Delta + 2\varepsilon.$$

Thus,

$$\begin{aligned} &\|\tilde{\mathcal{M}}(\mathcal{OPF}((\mathbf{s}^l)'), (\mathbf{s}^l)') - \tilde{\mathcal{M}}(\mathcal{OPF}((\mathbf{s}^l)''), (\mathbf{s}^l)'')\|_1 \\ \leq &\|(\mathbf{s}^g)' - (\mathbf{s}^g)''\|_1 + \|(\mathbf{s}^l)' - (\mathbf{s}^l)''\|_1 \leq 2\Delta + 2\varepsilon. \end{aligned}$$

Then the conclusion is implied by Lemma 1. ■

Remark 2: Theorems 2 and 3 show that the level of differential privacy is independent of how the aggregation regions are divided and how the data are aggregated. In particular, the amount of noise required relies on neither the number r of regions nor the number of buses in each region.

D. Determining Monotonicity

Monotonicity in Definition 4 may not hold for general networks. In this subsection we characterize topologies that are monotone. In particular, we show that radial networks are monotone.

A power system is monotone if the derivative[5] $\partial_{\mathbf{s}^l} \mathcal{OPF}(\mathbf{s}^l)$ of the corresponding $\mathcal{OPF}$ operator is nonnegative. Let $\mathcal{S}_{\mathrm{G}}(\mathbf{s}^l)$ and $\mathcal{S}_{\mathrm{B}}(\mathbf{s}^l)$ denote the set of generators

[5]We adopt the following notation: $\partial_{\mathbf{s}^l} \mathcal{OPF}(\mathbf{s}^l) := \frac{\partial \mathcal{OPF}(\mathbf{s}^l)}{\partial \mathbf{s}^l} = \frac{\partial \mathbf{s}^g}{\partial \mathbf{s}^l}$.

and branches that are binding, respectively, for a given $\mathbf{s}^l$, i.e.

$$\mathcal{S}_{\mathrm{G}}(\mathbf{s}^l) := \{i \in \mathcal{V}_{\mathrm{G}} : \mathbf{s}_i^g \in \{\underline{\mathbf{s}}_i^g, \overline{\mathbf{s}}_i^g\},$$
$$\mathcal{S}_{\mathrm{B}}(\mathbf{s}^l) := \{e \in \mathcal{E} : \mathbf{p}_e \in \{\underline{\mathbf{p}}_e, \overline{\mathbf{p}}_e\}\}.$$

When there is no danger of confusion, we will write $\mathcal{S}_{\mathrm{G}}$ and $\mathcal{S}_{\mathrm{B}}$ for simplicity.

Lemma 2: Under Assumptions 2 and 3, for $\mathbf{s}^l \in \widetilde{\Omega}_{\mathbf{s}^l}$, the derivative $\partial_{\mathbf{s}^l}\mathcal{OPF}(\mathbf{s}^l)$ always exists, and the sets $\mathcal{S}_{\mathrm{G}}$ and $\mathcal{S}_{\mathrm{B}}$ do not change in a neighborhood of $\mathbf{s}^l$.

The lemma can be proved using results in [11], [12]. The proof is omitted due to space limitation, but will be included in the journal version of this paper.

Returning to the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, we divide $\mathcal{E}$ into two disjoint sets:

$$\mathcal{E}^{\mathrm{I}} := \{e \in \mathcal{E} \mid \mathcal{G}(\mathcal{V}, \mathcal{E}\backslash\{e\}) \text{ is not connected}\}$$
$$\mathcal{E}^{\mathrm{II}} := \mathcal{E}\backslash\mathcal{E}^{\mathrm{I}}.$$

Links in $\mathcal{E}^{\mathrm{I}}$ are called bridges in $\mathcal{G}$. In general, it is possible that $\mathcal{E}^{\mathrm{I}} = \emptyset$, e.g., when $\mathcal{G}$ is a cycle. The next result connects monotonicity to network topology.

Theorem 4: For any $\mathbf{s}^l \in \widetilde{\Omega}_{\mathbf{s}^l}$ such that $\mathcal{S}_{\mathrm{B}}(\mathbf{s}^l) \subseteq \mathcal{E}^{\mathrm{I}}$, we have $\partial_{\mathbf{s}^l}\mathcal{OPF}(\mathbf{s}^l) \geq 0$, i.e., the system is monotone.

Proof: See Appendix C. ■

Thereom 4 implies directly the following corollaries.

Corollary 3: Power networks whose graphs $\mathcal{G}$ are trees are monotone.

Corollary 4: If all the possible branch flow bottlenecks[6] in the power system are in $\mathcal{E}^{\mathrm{I}}$, then the system is monotone.

When the network contains only one cycle, we can estimate the value of (δ, ε) in polynomial time such that the system is (δ, ε)-monotone. The algorithm and its proof will be presented in the journal version.

We end this section by pointing out that there are topologies whose behavior is far from monotone, i.e., the system could be (δ, ε)-monotone where $\varepsilon \gg \delta$. An example of such a network is shown in Section V.

E. Generalization

In general, for an arbitrary mechanism not necessary the aggregation mechanism, the L_1-sensitivity Δ_1 in (4) depends on the properties of both $\tilde{\mathcal{M}}$ and $\mathcal{OPF}$. When $\tilde{\mathcal{M}}$ is the aggregation mechanism, the problem boils down to the monotonicity of $\mathcal{OPF}$, as shown in the previous subsection. However, for general $\tilde{\mathcal{M}}$, the estimation of Δ_1 may require a more careful analysis of the structure of $\tilde{\mathcal{M}}$. The next result provides a rough estimation of the required amount of noise for differential privacy. Its proof is omitted due to space limitation.

Theorem 5: Suppose a power system is (Δ, ε)-monotone, and all elements of the Jacobian matrix $\mathbf{J}_{\tilde{\mathcal{M}}}$ with respective to $\tilde{\mathcal{M}} \in \mathbb{R}^r$ are upper bounded by the same constant U. Then the mechanism $\mathcal{M} = \tilde{\mathcal{M}} + \mathbf{Y}$, where all Y_i are drawn from independent Laplace distribution $\mathscr{L}(2Ur(\Delta+\varepsilon)/\varrho)$, preserves $(\Delta, \varrho)-$differential privacy.

[6]We define a bottleneck to be any edge $e \in \mathcal{E}$ such that $\exists \mathbf{s}^l \in \widetilde{\Omega}_{\mathbf{s}^l}$ where the optimal power flow $\mathbf{p}_e \in \{\underline{\mathbf{p}}_e, \overline{\mathbf{p}}_e\}$.

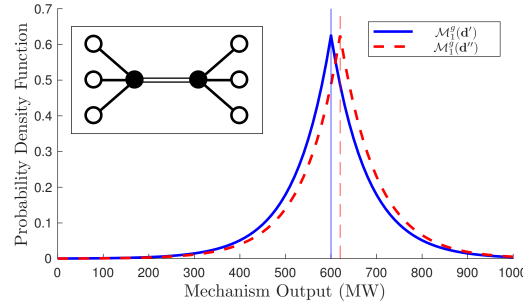

Fig. 1. The embedded diagram shows the topology of a radial power network, where black and white nodes indicate generators and loads, respectively. The double-line edge is the bottleneck of the system, and in our example, its line flow constraint is always binding. The vertical lines indicate the ground-truths of the queries for two datasets whose difference we want to hide. The curves are the probability density functions of the mechanism outputs that contain Laplace noise.

In the aggregation case, $U = 1$ and r is the number of regions. Comparing Theorem 3 and Theorem 5, the required Laplace noise is reduced by a factor of r in Theorem 3 which exploits the simple structure of the aggregation function.

V. Simulation

A. Radial Network

First, we apply the aggregation mechanism (5) to a radial power network (embedded image in Figure 1), i.e., network with a tree topology. Corollary 3 implies that the system is monotone, and by Theorem 2, the noise should be drawn independently from the Laplace distribution $\mathscr{L}(2\Delta/\varrho)$ so as to preserve (Δ, ϱ)-differential privacy. In this simulation, we set $\Delta = 20$ (MW) and $\varrho = 0.5$. The interpretation is that any two datasets whose difference we would like to hide should differ in any one load by at most 20 MW. This example has been constructed so that the double-line edge in Fig. 1 is a bottleneck. According to Appendix C, this bottleneck splits the tree into two subtrees and each subtree contains exactly one generator which is not saturated. Provided the OPF problem remains feasible, any change in the load will directly lead to the same amount of change in the generator which resides in the same subtree as the changing load.

Specifically, if the load on the left increases by Δ, the left generator will increase its generation by Δ while the right generator will remain unchanged. Hence the ground-truths of the aggregation queries (shown by vertical lines in Fig. 1) are separated by 20 MW, the same as Δ. The density functions shown in Fig. 1 are sharper than those of networks with cycles, as shown in Figures 3 and 4.

B. IEEE 9-Bus Network

The IEEE 9-bus testcase is not monotone, as shown in Figure 2. Indeed, increasing the load on either bus 5 (dashed curves) or bus 9 (solid curves) will lead to production decrease in generator 2. However, it is found that for any $\Delta > 0$, the system is $(\Delta, 2.01\Delta)$-monotone.

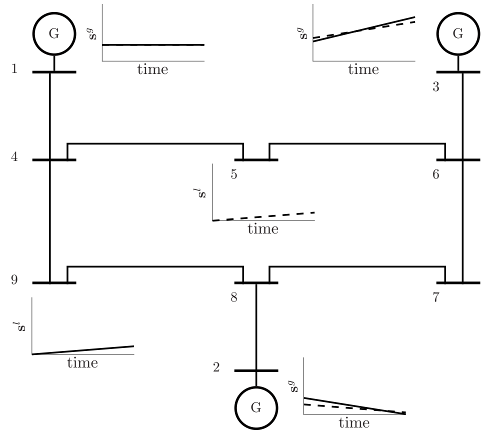

Fig. 2. IEEE 9-bus case. Dashed and solid curves show how the optimal generations change as loads on bus 5 and bus 9 increase. Bus 1 has constant generation since its generation has reached its upper limit.

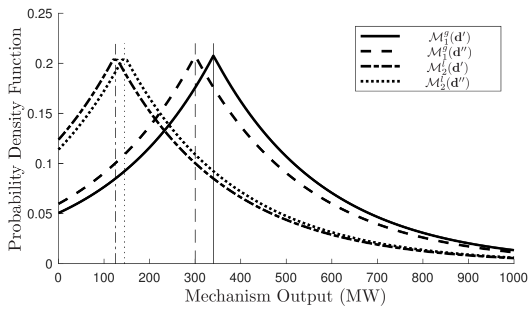

Fig. 3. Differential privacy for IEEE 9-bus case. The vertical lines indicate the ground-truths, and the curves show the probability distribution of the mechanism outputs. Only the aggregated generation in region 1 and the aggregated load in region 2 are presented in the figure.

We again set $\Delta = 20$ (MW), $\varrho = 0.5$ and divide the system into two regions. In our simulations, region 1 contains buses 1, 2, 4, 5, 6, while region 2 contains buses 3, 7, 8, 9. Figure 3 shows the probability density functions for the aggregation mechanism when the load on bus 9 increases by Δ. The difference between $\mathcal{M}_2^l(\mathbf{d}')$ and $\mathcal{M}_2^l(\mathbf{d}'')$ comes directly from the change on bus 9, but the difference between $\mathcal{M}_1^g(\mathbf{d}')$ and $\mathcal{M}_1^g(\mathbf{d}'')$ is mainly due to the fact that generator 3 has to increase its generation so as to compensate for the decrease in generation on bus 2. Hence the distributions for $\mathcal{M}_1^g(\mathbf{d}')$ and $\mathcal{M}_1^g(\mathbf{d}'')$ are further apart compared to the distributions for $\mathcal{M}_2^l(\mathbf{d}')$ and $\mathcal{M}_2^l(\mathbf{d}'')$. As a result, to preserve the same level of differential privacy, the required noise magnitude is greater than what would have been needed if the system were monotone. The distributions in Figure 3 are indeed flatter than those in Figure 1.

C. Bad Topology

There are networks whose behavior can be arbitrarily far from monotone, i.e., they are (δ, ε)-monotone with large ε. For these networks, differential privacy is only possible with the addition of large noise, potentially rendering the output of the mechanism meaningless.

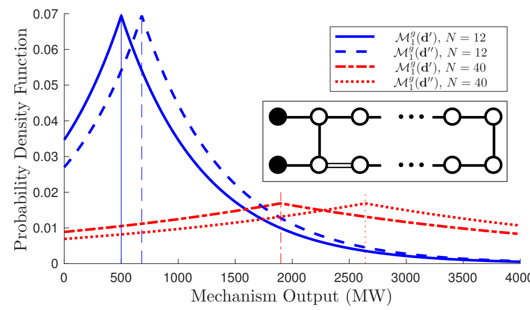

Fig. 4. Differential privacy for a ring network, shown in the embedded diagram. Black nodes represent generators and white nodes represent loads. The figure shows the density functions of the aggregation mechanism for different network sizes.

One such network is shown in Figure 4. This network consists of a cycle with N buses, with generators on two adjacent buses (black nodes). The branch indicated by the double-line edge is the only bottleneck where the line flow constraint is binding. It can be shown that this network is $(\Delta, (N-4)\Delta)$-monotone for some positive Δ. This means that a change in load can be amplified $N-4$ times in some generator, implying a large L_1-sensitivity. Figure 4 shows that to achieve $(20, 0.5)$-differential privacy, a far bigger noise is required than in the monotone case. As N increases, the density function becomes flatter. When $N = 40$ buses, the density function in Figure 4 is close to a uniform distribution, i.e., the mechanism hardly discloses any useful information.

VI. Conclusion

We have proposed a differential privacy model for OPF data in power systems. We have introduced the notion of monotonicity of $\mathcal{OPF}$ operator and used it to determine the amount of noise needed to preserve differential privacy under the aggregation mechanism. We have also shown that, for the aggregation mechanism, the level of differential privacy is independent of the number of aggregation regions and the number of buses in a region. Future work will look at how these results can be applied to the designing of new mechanisms.

References

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography Conference*. Springer, 2006, pp. 265–284.

[2] C. Dwork, "Differential privacy: A survey of results," in *International Conference on Theory and Applications of Models of Computation*. Springer, 2008, pp. 1–19.

[3] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[4] S. Han, U. Topcu, and G. J. Pappas, "Differentially private distributed protocol for electric vehicle charging," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*. IEEE, 2014, pp. 242–249.

[5] F. Fioretto and P. Van Hentenryck, "Constrained-based differential privacy: Releasing optimal power flow benchmarks privately," *To appear in Proc. of the International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, 2018.

[6] A. J. Wood and B. F. Wollenberg, *Power generation, operation, and control*. John Wiley & Sons, 2012.

[7] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*. Athena Scientific Belmont, MA, 1997, vol. 6.

[8] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[9] X.-S. Zhang and D.-G. Liu, "A note on the continuity of solutions of parametric linear programs," *Mathematical Programming*, vol. 47, no. 1-3, pp. 143–153, 1990.

[10] J. Anderson, F. Zhou, and S. H. Low, "Disaggregation for networked power systems," *To appear in Proc. of Power Systems Computation Conference (PSCC)*, 2018.

[11] A. V. Fiacco, "Sensitivity analysis for nonlinear programming using penalty methods," *Mathematical programming*, vol. 10, no. 1, pp. 287–311, 1976.

[12] ——, "Introduction to sensitivity and stability analysis in nonlinear programming," 1983.

[13] G. B. Folland, *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.

Appendix

A. Proof (Sketch) of Proposition 1

We first show that for a fixed network $(\mathbf{B}, \mathbf{C}, \boldsymbol{\xi})$ and $\mathbf{s}^l \in \Omega_{\mathbf{s}^l}$, if the primal optimal solution to (1) is not unique for some $\mathbf{f}$, then there must exist $\boldsymbol{\tau}, \boldsymbol{\mu}_+, \boldsymbol{\mu}_-, \boldsymbol{\lambda}_+, \boldsymbol{\lambda}_-$ such that (2) holds but (3) does not. Geometrically, an LP has multiple optimal solutions if and only if the objective vector is normal to some hyperplane defined by equality constraints and the set of binding inequality constraints. In our case, the objective vector $[\mathbf{f}^\mathsf{T}, \mathbf{0}^\mathsf{T}]^\mathsf{T}$ is a $N_\mathrm{G} + N$ dimensional vector. As there are $N+1$ linearly independent equality constraints in (1b), (1c),[7] it is therefore enough to take $\leq N_\mathrm{G} - 2$ inequality constraint vectors, along with the above $N+1$ to represent $[\mathbf{0}^\mathsf{T}, \mathbf{f}^\mathsf{T}]^\mathsf{T}$. That is to say, there exist Lagrange multipiers that satisfy (2), and there are at most $N_\mathrm{G} - 2$ non-zero coefficients in $\boldsymbol{\mu}_+, \boldsymbol{\mu}_-, \boldsymbol{\lambda}_+, \boldsymbol{\lambda}_-$, i.e.,

$$\|\boldsymbol{\mu}_+\|_0 + \|\boldsymbol{\mu}_-\|_0 + \|\boldsymbol{\lambda}_+\|_0 + \|\boldsymbol{\lambda}_-\|_0 < N_\mathrm{G} - 1. \quad (6)$$

Thereby, (3) implies the uniqueness, and we have

$$\Omega_\mathbf{f} = \{\mathbf{f} \mid \forall \mathbf{s}^l \in \Omega_{\mathbf{s}^l}, \text{the solutions of (2) satisfy (3)}\}. \quad (7)$$

For $\mathcal{S} \subseteq [N]$, $\mathcal{T} \subseteq [N_\mathrm{G}]$ such that $|\mathcal{S}| + |\mathcal{T}| \leq N_\mathrm{G} - 2$, we construct $\mathcal{Q}(\mathcal{S}, \mathcal{T})$ to be the set of $\mathbf{f}$ such that $\exists \boldsymbol{\tau} \in \mathbb{R}^{N+1}, \boldsymbol{\mu} \in \mathbb{R}^E, \boldsymbol{\lambda} \in \mathbb{R}^{N_\mathrm{G}}$ satisfying:

$$\mathbf{0} = \mathbf{M}^\mathsf{T}\boldsymbol{\tau} + \mathbf{C}\mathbf{B}\boldsymbol{\mu} \quad (8a)$$
$$-\mathbf{f} = -[\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_{N_\mathrm{G}}]^\mathsf{T} + \boldsymbol{\lambda} \quad (8b)$$
$$\boldsymbol{\mu}_i \neq 0 \Rightarrow i \in \mathcal{S} \quad (8c)$$
$$\boldsymbol{\lambda}_i \neq 0 \Rightarrow i \in \mathcal{T}. \quad (8d)$$

When $\mathcal{S}$ and $\mathcal{T}$ are fixed, the vector $\mathbf{C}\mathbf{B}\boldsymbol{\mu}$ takes value in an $|\mathcal{S}|$ dimensional subspace. Since $\mathrm{rank}(\mathbf{M}) = N$, the possible values of $\boldsymbol{\tau}$ must fall within an $|\mathcal{S}| + 1$ dimensional subspace. Therefore, (8b) implies that $\mathbf{f}$ must be in an $|\mathcal{S}| + 1 + |\mathcal{T}| \leq N_\mathrm{G} - 1$ dimensional subspace, and hence interior(closure($\mathcal{Q}(\mathcal{S}, \mathcal{T})$) $= \emptyset$. By the Baire Category Theorem [13], we have

$$\bigcup_{\substack{\mathcal{S}\subseteq[N], \mathcal{T}\subseteq[N_\mathrm{G}] \\ |\mathcal{S}|+|\mathcal{T}|\leq N_\mathrm{G}-2}} \mathcal{Q}(\mathcal{S}, \mathcal{T}) \quad (9)$$

is of the first category and thereby nowhere dense in $\mathbb{R}^{N_\mathrm{G}}$. On the other hand, (7) and (8) imply that

$$\left(\bigcup_{\substack{\mathcal{S}\subseteq[N], \mathcal{T}\subseteq[N_\mathrm{G}] \\ |\mathcal{S}|+|\mathcal{T}|\leq N_\mathrm{G}-2}} \mathcal{Q}(\mathcal{S}, \mathcal{T})\right)^\mathsf{c} \subseteq \Omega_\mathbf{f} \quad (10)$$

and hence $\Omega_\mathbf{f}$ is dense in $\mathbb{R}^{N_\mathrm{G}}$.

B. Proof (Sketch) of Proposition 2

Due to the limited space, we only provide a sketch of the proof here. Consider the power equations below:

$$\mathbf{T}\boldsymbol{\theta} := \begin{bmatrix} \mathbf{C}\mathbf{B}\mathbf{C}^\mathsf{T} \\ \mathbf{B}\mathbf{C}^\mathsf{T} \end{bmatrix} \cdot \boldsymbol{\theta} = \begin{bmatrix} \mathbf{s}^g \\ -\mathbf{s}^l \\ \mathbf{p} \end{bmatrix}. \quad (11)$$

Proposition 1 shows that there will always be at least $N_\mathrm{G} - 1$ binding inequality constraints as each non-zero multiplier will force one inequality constraint to be binding. A constraint is binding means some $\mathbf{s}_i^g$ equals either $\overline{\mathbf{s}}_i^g$ or $\underline{\mathbf{s}}_i^g$ (as in the upper N_G rows in (11)), or some $\mathbf{p}_i$ equals either $\overline{\mathbf{p}}_i$ or $\underline{\mathbf{p}}_i$ (as in the lower E rows in (11)). We have $\mathrm{rank}(\mathbf{T}) = N - 1$. We will first use the following procedure to construct a new set $\widetilde{\Omega}'_\mathrm{para}$.

I. $\widetilde{\Omega}'_\mathrm{para} \leftarrow \Omega_\mathrm{para}$

II. For each $\mathcal{S} \subseteq [N_\mathrm{G}] \cup [N+1, N+E]$ such that $|\mathcal{S}| = N_\mathrm{G}$, construct $\mathbf{T}_\mathcal{S}$.

  a) If $\mathrm{rank}(\mathbf{T}_\mathcal{S}) = N_\mathrm{G}$, then continue to the next $\mathcal{S}$.

  b) If $\mathrm{rank}(\mathbf{T}_\mathcal{S}) < N_\mathrm{G}$, then consider

$$\Gamma := \prod_{i\in\mathcal{S}\cap[N_\mathrm{G}]} \{\mathbf{e}_i, \mathbf{e}_{N_\mathrm{G}+i}\} \times \prod_{\substack{j\in[E] \\ j+N\in\mathcal{S}}} \{\mathbf{e}_{2N_\mathrm{G}+j}, \mathbf{e}_{2N_\mathrm{G}+E+j}\}. \quad (12)$$

  Now update $\widetilde{\Omega}'_\mathrm{para}$ as

$$\widetilde{\Omega}'_\mathrm{para} \leftarrow \widetilde{\Omega}'_\mathrm{para} \setminus \bigcup_{\gamma\in\Gamma} \{\boldsymbol{\xi} | \exists \boldsymbol{\theta}, \text{s.t. } \gamma^\mathsf{T}\boldsymbol{\xi} = \mathbf{T}_\mathcal{S}\boldsymbol{\theta}\}. \quad (13)$$

III. Return $\widetilde{\Omega}'_\mathrm{para}$.

In the above procedure, an n-tuple of vectors is also regarded as a matrix of n columns.[8] Since $\gamma \in \Gamma$ is of dimension N_G and $\mathbf{T}_\mathcal{S}\boldsymbol{\theta}$ defines an $N_\mathrm{G} - 1$ dimensional subspace for all $\boldsymbol{\theta} \in \mathbb{R}^N$, each set of $\{\boldsymbol{\xi} | \exists \boldsymbol{\theta}, \text{s.t. } \gamma^\mathsf{T}\boldsymbol{\xi} = \mathbf{T}_\mathcal{S}\boldsymbol{\theta}\}$ in (13) is a subspace with dimension strictly lower than $2N_\mathrm{G} + 2E$. Using the same technique as in the proof of Proposition 1, we have that $\widetilde{\Omega}'_\mathrm{para}$ is dense in Ω_para. It is sufficient to show that $\widetilde{\Omega}'_\mathrm{para} \subseteq \widetilde{\Omega}_\mathrm{para}$.

In fact, $\forall \boldsymbol{\xi} \in \widetilde{\Omega}'_\mathrm{para}$, if for some $\mathbf{s}^l \in \Omega_{\mathbf{s}^l}(\boldsymbol{\xi})$, the optimal solution to (1) has $\geq N_\mathrm{G}$ tight inequality constraints, then we use $\mathcal{S} \subseteq [N_\mathrm{G}] \cup [N+1, N+E], |\mathcal{S}| = N_\mathrm{G}$ again to denote the indices of any N_G tight inequality constraints. As those N_G inequality constraints are tight, there must exist $\boldsymbol{\theta} \in \mathbb{R}^N$ and $\gamma \in \Gamma$, as defined in (12), such that $\gamma^\mathsf{T}\boldsymbol{\xi} = \mathbf{T}_\mathcal{S}\boldsymbol{\theta}$. According to (13), $\mathrm{rank}(\mathbf{T}_\mathcal{S})$ must be exactly N_G. Plugging

[7] Here, independence means that the gradient of the equalities (1b) and (1c) with respect to $[(\mathbf{s}^g)^\mathsf{T}, \boldsymbol{\theta}^\mathsf{T}]^\mathsf{T}$ has full column rank $N+1$.

[8] Hence, each $\gamma \in \Gamma$ can also be regarded as a $(2N_\mathrm{G} + 2E)$-by-N_G matrix.

in the optimal $\boldsymbol{\theta}$, as well as the tight limits indexed by some $\gamma \in \Gamma$, into (11), we have

$$\gamma \cdot \boldsymbol{\xi} = \mathbf{T}_{\mathcal{S}}\boldsymbol{\theta} \tag{14a}$$
$$-\mathbf{s}^l = \mathbf{T}_{[N_{\mathrm{G}}+1, N_{\mathrm{G}}]}\boldsymbol{\theta}. \tag{14b}$$

For each $\gamma \in \Gamma$, as $\mathrm{rank}(\mathbf{T}_{\mathcal{S}}) = N_{\mathrm{G}}$ but $\mathrm{rank}(\mathbf{T}) = N-1$, the set $\{\mathbf{s}^l \mid \exists \boldsymbol{\theta}, (14) \text{ holds}\}$ is a subspace in $\mathbb{R}^{N_{\mathrm{L}}}$ and thereby nowhere dense in $\Omega_{\mathbf{s}^l}$. As the result, we have

$$\widetilde{\Omega}_{\mathbf{s}^l} \supseteq \Omega_{\mathbf{s}^l} \setminus \bigcup_{\gamma \in \Gamma} \{\mathbf{s}^l \mid \exists \boldsymbol{\theta}, (14) \text{ holds for } \gamma\}$$

must be dense in $\Omega_{\mathbf{s}^l}$. Therefore, $\widetilde{\Omega}'_{\mathrm{para}} \subseteq \widetilde{\Omega}_{\mathrm{para}}$ and $\widetilde{\Omega}_{\mathrm{para}}$ is dense in Ω_{para}.

C. Proof of Theorem 4

For fixed $u \in [N_{\mathrm{L}}]$, Lemma 2 shows that there exists $\kappa_u > 0$ such that $\forall \omega_u \in (-\kappa_u, \kappa_u)$, $\hat{\mathbf{s}^l} := \mathbf{s}^l + \omega_u \mathbf{e}_u$ satisfies

$$\mathcal{S}_{\mathrm{G}}(\hat{\mathbf{s}^l}) = \mathcal{S}_{\mathrm{G}}(\mathbf{s}^l)$$
$$\mathcal{S}_{\mathrm{B}}(\hat{\mathbf{s}^l}) = \mathcal{S}_{\mathrm{B}}(\mathbf{s}^l).$$

It is sufficient to show $\mathcal{OPF}(\hat{\mathbf{s}^l}) \geq \mathcal{OPF}(\mathbf{s}^l)$ when $\omega_u \in (0, \kappa_u)$ for any fixed u.[9]

Since $\mathcal{S}_{\mathrm{B}} \subseteq \mathcal{E}^{\mathrm{I}}$, Proposition 2 implies that $|\mathcal{S}_{\mathrm{B}} \cap \mathcal{E}^{\mathrm{I}}| = |\mathcal{S}_{\mathrm{B}}| = N_{\mathrm{G}} - 1$. Thereby $\mathcal{S}_{\mathrm{B}}$ splits $\mathcal{G}$ into N_{G} connected components $\mathcal{G}_1, \ldots, \mathcal{G}_{N_{\mathrm{G}}}$, and each component has vertices $\mathcal{V}_i$ and Edges $\mathcal{E}_i$.

We are first showing that

$$\forall i \in [N_{\mathrm{G}}],\ |\mathcal{V}_i \cap ([N_{\mathrm{G}}] \backslash \mathcal{S}_{\mathrm{G}})| = 1. \tag{16}$$

Since $\cup_{i=1}^{N_{\mathrm{G}}} \mathcal{V}_i = \mathcal{V} \supseteq [N_{\mathrm{G}}] \backslash \mathcal{S}_{\mathrm{G}}$, and

$$\begin{aligned} |[N_{\mathrm{G}}] \backslash \mathcal{S}_{\mathrm{G}}| &= N_{\mathrm{G}} - |\mathcal{S}_{\mathrm{G}}| \\ &= N_{\mathrm{G}} - (N_{\mathrm{G}} - 1 - |\mathcal{S}_{\mathrm{B}}|) \\ &= N_{\mathrm{G}}, \end{aligned}$$

if (16) does not hold, then there must exist $i \in [N_{\mathrm{G}}]$ such that $\mathcal{V}_i \cap ([N_{\mathrm{G}}] \backslash \mathcal{S}_{\mathrm{G}}) = \emptyset$ and thus $\mathcal{V}_i \cap [N_{\mathrm{G}}] \subseteq \mathcal{S}_{\mathrm{G}}$. Now, for component $\mathcal{G}_i$, power flow equations imply that

$$\sum_{j \in \mathcal{V}_i \cap \mathcal{V}_{\mathrm{G}}} \mathbf{s}^g_j - \sum_{j \in \mathcal{V}_i \cap \mathcal{V}_{\mathrm{L}}} \mathbf{s}^l_{j-N_{\mathrm{G}}} = \sum_{\substack{e: e \in \mathcal{E}, \\ \sum\limits_{k \in \mathcal{V}_i} \mathbf{C}_{k,e} = 1}} \mathbf{p}_e - \sum_{\substack{e: e \in \mathcal{E}, \\ \sum\limits_{k \in \mathcal{V}_i} \mathbf{C}_{k,e} = -1}} \mathbf{p}_e. \tag{17}$$

In (17), for $j \in \mathcal{V}_i \cap \mathcal{V}_{\mathrm{G}}$, we have $\mathbf{s}^g_j \in \{0, \overline{\mathbf{s}^g}\}$ as $\mathcal{V}_i \cap [N_{\mathrm{G}}] \subseteq \mathcal{S}_{\mathrm{G}}$. On the other hand, for $e \in \mathcal{E}$ such that $\sum_{k \in \mathcal{V}_i} \mathbf{C}_{k,e} = \pm 1$, e must be the bridge connecting $\mathcal{G}_i$ and some other component, and thereby e is in the cut $\mathcal{S}_{\mathrm{B}}$. By definition, we have $\mathbf{p}_e \in \{\underline{\mathbf{p}}, \overline{\mathbf{p}}\}$. Since all the generators and branch power flows involved in (17) are binding, it contradicts to Corollary 2 and therefore (16) always holds.

Now let $\aleph$ be the mapping such that $\aleph(\mathcal{G}_i)$ is the unique generator in $\mathcal{V}_i \cap ([N_{\mathrm{G}}] \backslash \mathcal{S}_{\mathrm{G}})$ for each $i \in [N_{\mathrm{G}}]$. For any fixed $u \in [N_{\mathrm{L}}]$ and $\omega_u \in (-\kappa_u, \kappa_u)$, we will prove that $\mathcal{OPF}_v(\hat{\mathbf{s}^l}) \geq \mathcal{OPF}_v(\mathbf{s}^l)$ for each $v \in [N_{\mathrm{G}}]$ by discussing the following three possible situations that may arise. Assume $u + N_{\mathrm{G}} \in \mathcal{V}_k$ for $k \in [N_{\mathrm{G}}]$.

- If $v \in \mathcal{S}_{\mathrm{G}}(\mathbf{s}^l)$, then (15a) implies $v \in \mathcal{S}_{\mathrm{G}}(\hat{\mathbf{s}^l})$ as well. Since $\mathcal{OPF}$ is continuous over $\omega_u \in (-\kappa_u, \kappa_u)$ and $\overline{\mathbf{s}^g} > 0$, there must be $\mathcal{OPF}_v(\hat{\mathbf{s}^l}) = \mathcal{OPF}_v(\mathbf{s}^l)$.
- If $v = \aleph(\mathcal{G}_k)$, then similar to (17) we have

$$\begin{aligned} &\sum_{j \in \mathcal{V}_k \cap \mathcal{S}_{\mathrm{G}}} \mathcal{OPF}_j(\mathbf{s}^l) + \mathcal{OPF}_v(\mathbf{s}^l) - \sum_{j \in \mathcal{V}_k \cap \mathcal{V}_{\mathrm{L}}} \mathbf{s}^l_{j-N_{\mathrm{G}}} \\ &= \sum_{\substack{e: e \in \mathcal{E}, \\ \sum\limits_{l \in \mathcal{V}_k} \mathbf{C}_{l,e} = 1}} \mathbf{p}_e - \sum_{\substack{e: e \in \mathcal{E}, \\ \sum\limits_{l \in \mathcal{V}_k} \mathbf{C}_{l,e} = -1}} \mathbf{p}_e \\ &= \sum_{j \in \mathcal{V}_k \cap \mathcal{S}_{\mathrm{G}}} \mathcal{OPF}_j(\hat{\mathbf{s}^l}) + \mathcal{OPF}_v(\hat{\mathbf{s}^l}) - \sum_{j \in \mathcal{V}_k \cap \mathcal{V}_{\mathrm{L}}} \hat{\mathbf{s}^l}_{j-N_{\mathrm{G}}} \end{aligned} \tag{18}$$

  As $\mathbf{s}^l$ and $\hat{\mathbf{s}^l}$ only differ at load u and $\mathcal{OPF}_j(\hat{\mathbf{s}^l}) = \mathcal{OPF}_j(\mathbf{s}^l)$ for all $j \in \mathcal{V}_k \cap \mathcal{S}_{\mathrm{G}}$ as shown above, (18) can be simplified to

$$\mathcal{OPF}_v(\hat{\mathbf{s}^l}) - \mathcal{OPF}_v(\mathbf{s}^l) = \hat{\mathbf{s}^l}_u - \mathbf{s}^l_u = \omega_u > 0,$$

  and therefore $\mathcal{OPF}_v(\hat{\mathbf{s}^l}) > \mathcal{OPF}_v(\mathbf{s}^l)$.
- If $v = \aleph(\mathcal{G}_{k'})$ for some $k' \neq k$, then (18) still holds for $\mathcal{G}_{k'}$ but $\mathbf{s}^l$ and $\hat{\mathbf{s}^l}$ are identical for loads in $\mathcal{G}_{k'}$. Hence we have $\mathcal{OPF}_v(\hat{\mathbf{s}^l}) = \mathcal{OPF}_v(\mathbf{s}^l)$.

Putting this together, we conclude that $\mathcal{OPF}_v(\hat{\mathbf{s}^l}) \geq \mathcal{OPF}_v(\mathbf{s}^l)$ for all $v \in [N_{\mathrm{G}}]$.

[9]By symmetry, we will have $\mathcal{OPF}(\hat{\mathbf{s}^l}) \leq \mathcal{OPF}(\mathbf{s}^l)$ when $\omega_u \in (-\kappa_u, 0)$ for any fixed u.