Disaggregation for Networked Power Systems

James Anderson, Fengyu Zhou, Steven H. Low
California Institute of Technolgy
{james, f.zhou, slow}@caltech.edu

***Abstract*—Electricity data such as supply, demand, prices, and line flows, are sensitive. Utility companies, understandably do not want to, in fact often cannot, make this data publicly available. However, such data is critical in many fundamental areas in power systems research. As a compromise, aggregated data sets are sometimes made available. In such a setting it may be the case that data is aggregated over a geographical region and time. This forces researchers to try and "invert" the data to obtain *disaggregated* data sets. In this paper we rigorously formulate the *disaggregation problem* for networked power systems and present two algorithms that provide solutions to the DC version of the problem. We show that it is possible to invert the data, but that does not imply that ground truth solutions are obtained, thus the utility companies maintain a notion of privacy. The aim of this paper is to highlight the potential benefits to both the research community and utility companies of releasing aggregated data.**

I. Introduction

Motivated by the tension between utility companies not being able to make data available due to non-disclosure agreements and the desire to maintain a competitive edge, and the need in the research community for realistic data for algorithm design, the release of aggregated data can be seen as a good compromise. However, both parties may still be skeptical of such an agreement because i) utility companies do not have guarantees that the ground truth data cannot be reverse-engineered from the aggregated data, thus putting them in a vulnerable position; and ii) researchers do not know if the resulting reverse-engineered (disaggregated) models are realistic.

In this paper we define and formalize the *network disaggregation problem* which seeks to quantify how "good" a model which is reverse engineered from aggregated data is. The complicating factor is that the ground truth data is *never* made available to us. From the utility company's perspective such an approach is beneficial because it means that by handing over desensitized data, researchers can design algorithms based upon realistic models, which the utility company may be able to take advantage of. The contributions of this paper are to formulate at a very general level the disaggregation problem as an optimization problem, and then to provide solutions to the DC version of the problem. The main aim of the paper is to encourage both researchers and utility companies to consider data aggregation as a viable compromise between privacy and cooperation.

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

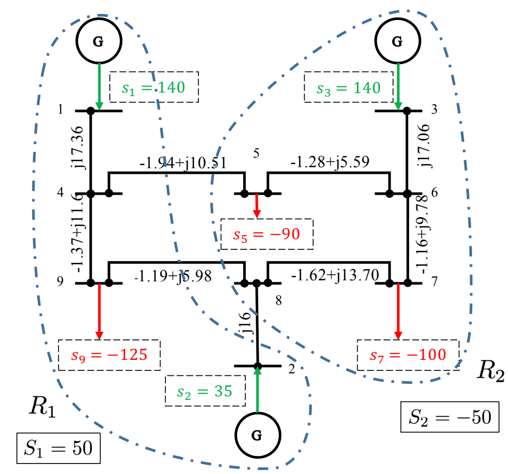

Figure 1. The full network (IEEE 9 bus system) with line admittances, bus generation and loads i.e. ground truth data (shown in dashed rectangles), aggregated data (solid rectangles), and aggregated regions (dot-dashed lines).

Several other approaches have been proposed that aim to provide privacy to grid operators. In [1] affine transformations to DC optimal power flow problems are applied which hide the structure of the underlying network. The same idea was extended to the setting where the transformation preserves some structure and maps the DC OPF problem to an OPF problem on another network [2]; in [3] the AC case was considered. This approach could be considered as a dual to our approach in the sense that it provides full data after a change of coordinates. In comparison aggregation reduces publicized data but does not change its coordinates. Another relevant concept is *differential privacy* [4] which has its roots in database theory and has been used to provide formal guarantees with regards to data privacy. These ideas have been extended to the realm of mechanism design [5] and constrained optimization for electric vehicle charging [6]. The idea of preserving privacy is also important in the disaggregation setting. While the disaggregation problem in this paper is modeled without artificial noise, preserving the differential privacy requires adding noise to the aggregated data of each region. It is found that to keep the same level of differential privacy, one should add the same level of noise regardless of how the regions are divided [7].

We now outline the basic idea of our disaggregation approach. Suppose we are given by the grid operator three types of information. The first type is a set $\mathcal{D}$ of network information, such as the network topology (for this paper we will assume that switch positions are known and fixed) and the admittance matrix Y. The second type is aggregated time series data in the set $\mathcal{A}$ which include the net power generations for specific regions of the network as well as an accurate description of the aggregation regions. Finally, any additional data such as power flows on specific lines, PMU data, or statistical information is contained in the set $\mathcal{X}$. Abstractly, the network disaggregation problem is to determine a mapping $\Phi : \{\mathcal{D}, \mathcal{A}, \mathcal{X}\} \to \mathcal{F}$, where $\mathcal{F}$ contains the vector of loads and generations at each bus that are *most consistent* with the given data. What makes this problem particularly challenging is that the ground truth data $\mathcal{G}$, i.e. the actual load and generation at each bus, are not available thus the problem cannot be reduced to a simple regression problem. An example that illustrates the disaggregation problem is shown in Figure 1. In this case the example has two aggregated regions indicated by the dot-dashed lines.

II. Models

A. Power Network

Consider a power network modelled by a connected undirected graph $G(N^+, E)$ where $N^+ := \{0\} \cup N$, $N := \{1, 2, \ldots, n\}$, and $E \subseteq N^+ \times N^+$. Each node in N^+ represents a bus and each edge in E represents a transmission or distribution line. For each line $(j,k) \in E$ let $y_{jk} := g_{jk} - \mathbf{i} b_{jk} \in \mathbb{C}$ be its series admittance. Let $y_{jj} := g_{jj} - \mathbf{i} b_{jj}$ be the shunt admittance connecting bus j to neutral (where the voltage potential is 0). Here y_{jj} is the sum of shunt admittances y^m_{jk} in the Π-model that are associated with all lines (j,k) connected to bus j, $y_{jj} := \sum_{k: j\sim k} y^m_{jk}$. Usually g_{jk}, b_{jk}, g_{jj} are all nonnegative (inductive) but shunt elements b_{jj} are nonpositive (capacitive). A bus $j \in N^+$ can have a generator, a load, both or neither. Let V_j be the complex voltage at bus $j \in N^+$ and $|V_j|$ denote its magnitude. Bus 0 is the *slack bus*. Its voltage is fixed and we assume without loss of generality that $V_0 = 1\angle 0^\circ$ per unit (pu). Let s_j be the net complex power injection (generation minus load) at bus $j \in N^+$. We use $\mathbf{s}^g := (s^g_j, j = 0, ..., n)$ and $\mathbf{s}^l := (s^l_j, j = 0, ..., n)$ to denote the generation and load vector for buses, respectively. Let $s_j := s^g_j + s^l_j$ be the net power flow for bus j, and $\mathbf{s} := (s_j, j = 0, ..., n) \in \mathbb{R}^{n+1}$ be the vector of net power flows. Note that the vector $\mathbf{s}^l$ contains non-positive values representing loads at each bus. When considering AC power flow, the power injections $\mathbf{s}$ and complex voltages $\mathbf{V} := (V_j, j = 0, \ldots, n)$ satisfy

$$s_j = \sum_{k=0}^{n} Y^H_{jk} V_j V^H_k, \quad \forall j \in N^+, \tag{1}$$

where

$$Y_{jk} = \begin{cases} -y_{jk} & j \sim k \ (j \neq k), \\ 0 & j \nsim k \ (j \neq k), \\ \sum_{i: j\sim i} y_{ji} + y_{jj} & j = k. \end{cases}$$

We will use the DC power flow model in this paper which makes the following assumptions:

1) All lines are lossless: Specifically, the branch resistances and shunt capacitances are negligible and so the network is fully described by the matrix $\mathbf{B}$ of susceptances which can be easily derived from the admittance matrix $\mathbf{Y}$ c.f. [8, p.18].
2) Voltage angle differences are small on each line: $\sin(\theta_j - \theta_k) \approx \theta_j - \theta_k$ for all $(j,k) \in E$.
3) $|V_j|$ is a given constant for all buses $j \in N^+$. Without loss of generality, we assume that $|V_j| = 1$ for all j.
4) Reactive power is ignored.

As a result of these assumptions the DC power flow equations are

$$s_j = \sum_{k: j\sim k} b_{jk} |V_j||V_k| (\theta_j - \theta_k). \tag{2}$$

An optimal power flow problem is then defined as an optimization problem that tries to optimally match supply and demand subject to the power flow equations (2).

B. Optimal Power Flow Model

We first define the disaggregation problem for the general AC (nonlinear) power flow model. We then specialize to the DC power flow case and provide solutions for DC version of the disaggregation problem.

Consider the following AC optimal power flow (OPF) problem:

$$\begin{aligned} \underset{\mathbf{V}, \mathbf{s}^g}{\text{minimize}} \quad & C(\mathbf{V}) \\ \text{subject to} \quad & \underline{v}_j \le |V_j|^2 \le \overline{v}_j, \quad \forall j \in N^+ && (3a) \\ & \underline{s}_j \le s_j \le \overline{s}_j, \quad \forall j \in N^+ && (3b) \\ & s_j \text{ satisfy (1)} \quad \forall j \in N^+. && (3c) \end{aligned}$$

We note that other constraints such as line limits and security constraints can be included without changing the essential structure of problem (3). To the OPF problem above we associate the set

$$\Omega := \{(\mathbf{s}^g, \mathbf{s}^l) \mid \exists\ \mathbf{V}, \text{ s.t. } (3a) - (3c) \text{ are satisfied}\}. \tag{4}$$

We now specialize to the DC OPF problem, following the formulation of [1]. The decision variables are the voltage angles $\boldsymbol{\theta}$ and the power injections $\mathbf{s}^g$. The DC OPF problem takes the following form:

$$\begin{aligned} \underset{\mathbf{s}^g, \boldsymbol{\theta}}{\text{minimize}} \quad & \mathbf{f}(\mathbf{s}^g) && (5a) \\ \text{subject to} \quad & \theta_0 = 0 && (5b) \\ & \mathbf{B}\boldsymbol{\theta} = \mathbf{s}^g + \mathbf{s}^l && (5c) \\ & \underline{\mathbf{s}}^g \le \mathbf{s}^g \le \overline{\mathbf{s}}^g && (5d) \\ & \underline{\mathbf{p}}_{\text{flow}} \le \mathbf{H}\boldsymbol{\theta} \le \overline{\mathbf{p}}_{\text{flow}}, && (5e) \end{aligned}$$

The cost function (5a) can be either a linear or quadric function of the generations. The equality constraint (5b) asserts that bus 0 is the slack bus. The vector of constraints (5c) are the power flow equations (2) that enforce power balance at each bus. The first set of inequality constraints described by (5d) correspond to generator limits. Finally the constraint (5e) corresponds to power flow limits and the matrix $\mathbf{H}$ is defined as $\mathbf{H} := \mathrm{diag}(\mathbf{b}_{\text{br}})\mathbf{A}_{\text{inc}}^{\mathrm{T}}$, where $\mathbf{A}_{\text{inc}}$ is the bus-to-branch incidence matrix and $\mathbf{b}_{\text{br}}$ is a vector containing all the branch susceptances and fixed voltage magnitudes. In fact, $\mathbf{b}_{\text{br}}$ is associated with $\mathbf{B}$ by $\mathbf{B} = \mathbf{A}_{\text{inc}}\mathrm{diag}(\mathbf{b}_{\text{br}})\mathbf{A}_{\text{inc}}^{\mathrm{T}}$. Finally, we define the optimal solution set of (5) by

$$\Sigma^{\star}(\mathbf{s}^l) := \{(\mathbf{s}^g, \boldsymbol{\theta}) \mid \text{solves (5), given } \mathbf{s}^l\},$$

and we overload notation slightly by defining the set

$$\Omega := \{(\mathbf{s}^g, \mathbf{s}^l) \mid \exists\, \boldsymbol{\theta}, \text{ s.t. (5a)} - \text{(5e) are satisfied}\}. \tag{6}$$

III. Formulation

A. Aggregation Data And Assumptions

We assume that the detailed injections at individual buses are not disclosed for privacy reasons. Instead, only the aggregated data are available. Consider that all the buses are divided into m regions $R_1, R_2 \cdots, R_m$, where $R_i \subseteq N^+$ is the set of bus IDs in region i and the set $\{R_1, R_2, \ldots, R_m\}$ is a partition of N^+. We denote the ground truths of $\mathbf{s}^g$ and $\mathbf{s}^l$ as $\boldsymbol{\sigma}^g$ and $\boldsymbol{\sigma}^l$, respectively, and denote the ground truth of $\mathbf{s}$ as $\boldsymbol{\sigma}$. Here, $\boldsymbol{\sigma}^g, \boldsymbol{\sigma}^l, \boldsymbol{\sigma} \in \mathbb{R}^{n+1}$. Then the net aggregated injection for region i is measured as

$$S_i = \sum_{j \in R_i} s_j + \omega_i. \tag{7}$$

Here $\omega_i \in \mathbb{R}$ is the measurement noise of the aggregated injection for region i. Let $\boldsymbol{\omega} \in \mathbb{R}^{n+1}$ be the vector of ω_i. The noise may be modelled as either a random variable drawn from an appropriate distribution or as a bounded disturbance, for example $\|\boldsymbol{\omega}\| < \epsilon$ for some choice of norm. The main task of load disaggregation problem is to recover a solution of $s_j, j \in N^+$ from the observations $S_i, i \in M := \{1, 2, \ldots, m\}$.

In an ideal setting, the disaggregation problem is simply to determine estimates of the ground truth data, $(\hat{s}^g, \hat{s}^l)$ that minimize δ as follows

$$\left\| \begin{bmatrix} \boldsymbol{\sigma}^g \\ \boldsymbol{\sigma}^l \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{s}}^g \\ \hat{\mathbf{s}}^l \end{bmatrix} \right\|_2^2 \leq \delta. \tag{8}$$

However, the obstacle that prevents disaggregation from being run as a simple regression problem is that we *never* have access to $(\boldsymbol{\sigma}^g, \boldsymbol{\sigma}^l)$, thus we cannot evaluate (8).

The key observation that we make, that will allow us to pose (and solve) the aggregation problem is as follows: The ground truth generation data $\boldsymbol{s}^g$ will have been determined by the utility company by solving an *optimal power flow* problem over the network described by the admittance matrix $\mathbf{Y}$. Furthermore, the admittance matrix is given to us. We can thus think of the optimal power flow problem as defining a map from loads to generations. Specifically, given ground truth load data $\boldsymbol{\sigma}^l$ the map is

$$\boldsymbol{\sigma}^g = \mathcal{OPF}(\boldsymbol{\sigma}^l; \mathbf{Y}) + \boldsymbol{\eta}. \tag{9}$$

In the most general setting, $\mathcal{OPF}(\cdot;\cdot)$ is a function that takes the network admittance matrix, the ground truth load data, solves an OPF problem and returns a vector of generations.[1] The $\boldsymbol{\eta}$ term describes measurement noise that may corrupt the observed generations. For brevity we will omit the second argument and write $\mathcal{OPF}(\boldsymbol{s}^l)$ when using the $\mathcal{OPF}$ operator and assume the admittance matrix required is implicit in the formulation.

B. Problem Formulation

To state the most general form of the problem we require the following assumption:

Assumption 1. *The admittance matrix $\mathbf{Y} \in \mathbb{C}^{(n+1)\times(n+1)}$ is known and available.*

Problem 1 (Disaggregation)**.** *Given the admittance matrix $\mathbf{Y} \in \mathbb{C}^{(n+1)\times(n+1)}$, an OPF model* (9)*, and the aggregated data S_i ($i \in M$), find a pair $(\mathbf{s}^g, \mathbf{s}^l)$ which is the most consistent with (7) and (9). Specifically, the problem can be formulated as*

$$\underset{\mathbf{s}^g, \mathbf{s}^l}{\text{minimize}} \quad \mathscr{L}(\mathbf{s}^g, \mathbf{s}^l) \tag{10a}$$

$$\text{subject to} \quad (\mathbf{s}^g, \mathbf{s}^l) \in \Omega \tag{10b}$$

where $\mathscr{L}(\mathbf{s}^g, \mathbf{s}^l)$ is a loss function which keeps both

$$\left(S_i - \sum_{j \in R_i} s_j\right)^2, \ \forall i \in M \quad \text{and} \quad \|\mathbf{s}^g - \mathcal{OPF}(\mathbf{s}^l)\|_2^2$$

small and Ω is defined as (4) *in the AC case and defined as* (6) *in the DC case.*

To avoid dealing with a multi-objective optimization problem, we consider $\mathscr{L}(\mathbf{s}^g, \mathbf{s}^l)$ of the following form:

$$\begin{aligned} \mathscr{L}(\mathbf{s}^g, \mathbf{s}^l) = &\lambda_1 \cdot \sum_{i \in M} \left(S_i - \sum_{j \in R_i} s_j\right)^2 \\ &+ \lambda_2 \cdot \|\mathbf{s}^g - \mathcal{OPF}(\mathbf{s}^l)\|_2^2. \end{aligned} \tag{11}$$

Here, $\lambda_k \geq 0$ $(k = 1, 2)$ are two parameters which balance the weights for the two error terms. We note that the formulation of (10) is very versatile. In many cases, however, we will have to restrict the form of the OPF problem in order to i) make Ω tractable to optimize over and ii) allow us to approximate the $\mathcal{OPF}$ map.

An estimate $(\mathbf{s}^g, \mathbf{s}^l)$ obtained from (10) is better than another estimate $(\hat{\mathbf{s}}^g, \hat{\mathbf{s}}^l)$ if $\mathscr{L}(\mathbf{s}^g, \mathbf{s}^l) < \mathscr{L}(\hat{\mathbf{s}}^g, \hat{\mathbf{s}}^l)$, and they are equally good if $\mathscr{L}(\mathbf{s}^g, \mathbf{s}^l) = \mathscr{L}(\hat{\mathbf{s}}^g, \hat{\mathbf{s}}^l)$. When $\boldsymbol{\omega} = 0$ and $\boldsymbol{\eta} = 0$ it follows from the definition that $\mathscr{L}(\boldsymbol{\sigma}^g, \boldsymbol{\sigma}^l) = 0$.

In the remainder of this section we focus on formulating computationally tractable disaggregation problems. Towards this goal we make the following assumptions:

[1] At this point there is no assumption that the $\mathcal{OPF}$ operator is related to the DC optimal power flow problem (5). The map can represent an AC OPF mapping associated with (3), or it could represent a mapping derived form measured data.

Assumption 2. *The $\mathcal{OPF}$ operator is described by the DC OPF problem* (5) *and is defined by the susceptance matrix* $\mathbf{B} \in \mathbb{R}^{(n+1)\times(n+1)}$.

Assumption 3. *The ground truth data* $(\boldsymbol{\sigma}^g, \boldsymbol{\sigma}^l)$ *is not corrupted by noise. i.e.* $\boldsymbol{\eta} = 0$ *in* (9).

We note that Assumption 2 could be relaxed, and instead an AC OPF problem (or its convex relaxation) is considered. We will pursue this idea in future work. Similarly, Assumption 3 will be lifted in future work. The emphasis of this paper is, however, on formulating the disaggregation problem.

Problem 2 (DC Disaggregation)**.** *Given the susceptance matrix* $\mathbf{B} \in \mathbb{R}^{(n+1)\times(n+1)}$ *the DC OPF problem* (5)*, and the aggregated data, solve*

$$
\begin{aligned}
\underset{\mathbf{s}^g,\mathbf{s}^l,\boldsymbol{\theta}}{\text{minimize}} \quad & \mathscr{L}(\mathbf{s}^g,\mathbf{s}^l) \\
\text{subject to} \quad & \mathbf{B}\boldsymbol{\theta} = \mathbf{s}^g + \mathbf{s}^l && \text{(12a)}\\
& \theta_0 = 0 && \text{(12b)}\\
& \underline{\mathbf{s}}^g \leq \mathbf{s}^g \leq \overline{\mathbf{s}}^g && \text{(12c)}\\
& \underline{\mathbf{s}}^l \leq \mathbf{s}^l \leq \overline{\mathbf{s}}^l && \text{(12d)}\\
& \underline{\mathbf{p}}_{\text{flow}} \leq \mathbf{H}\boldsymbol{\theta} \leq \overline{\mathbf{p}}_{\text{flow}} && \text{(12e)}\\
& (\mathbf{s}^g, \theta) \in \Sigma^\star(\mathbf{s}^l) && \text{(12f)}
\end{aligned}
$$

for $\mathbf{s}^g, \mathbf{s}^l$*, and* $\boldsymbol{\theta}$*. Define the set of optimal disaggregation variables as*

$$\Gamma^\star := \{(\mathbf{s}^g,\mathbf{s}^l) \mid (\mathbf{s}^g,\mathbf{s}^l) \text{ solves (12) for some } \boldsymbol{\theta}\}.$$

Note that due to assumption 3 and (12f), the cost function reduces to $\mathscr{L}(\mathbf{s}^g,\mathbf{s}^l) = \sum_{i\in M}\left(S_i - \sum_{j\in R_i} s_j\right)^2$.

Remark 1. *If, in addition to assumption 3 it is assumed that there is no aggregation noise, i.e.* $\boldsymbol{\omega} = 0$*, then the cost function in the DC Disaggregation problem can always be made to equal zero, and the problem reduces to finding a pair* $(\mathbf{s}^g,\mathbf{s}^l) \in \mathbb{R}^{n+1}\times\mathbb{R}^{n+1}$ *such that:*

$$
\begin{aligned}
S_i &= \sum_{j\in R_i} s_j \quad \textit{for } i \in M\\
\mathbf{s}^g &= \mathcal{OPF}(\mathbf{s}^l), \quad \underline{\mathbf{s}}^l \leq \mathbf{s}^l \leq \overline{\mathbf{s}}^l.
\end{aligned}
$$

We note that the number of equality constraints in (12) is smaller than the number of unknowns (this is true regardless of noise). Thus it is possible to obtain an optimal solution $(\hat{\mathbf{s}}^g,\hat{\mathbf{s}}^l) \in \Gamma^\star$ with $\mathscr{L}(\hat{\mathbf{s}}^g,\hat{\mathbf{s}}^l) = 0$, although such points will likely be different from the ground truth data $(\boldsymbol{\sigma}^g,\boldsymbol{\sigma}^l)$. This is because $\Gamma^\star$ is unlikely to be a singleton. Indeed, from the perspective of the utility company, the larger the set $\Gamma^\star$ is, the more *privacy* they have and the more unlikely it is that we can recover ground truth. It is also worth pointing out that in the noisy case ($\boldsymbol{\eta} \neq 0$) it is possible that $(\boldsymbol{\sigma}^g,\boldsymbol{\sigma}^l) \notin \Gamma^\star$ but is "close to" $\Gamma^\star$ in some sense. Investigating privacy measure will be a focus of future work [7]. Note that in the following sections, we will *not* assume that $\boldsymbol{\omega} = 0$.

IV. Algorithms

We provide two algorithms that attempt to solve the DC Disaggregation problem. The first method takes a game theoretic approach and iterates a *best response* strategy until convergence. The second approach formulates the problem as a bilevel optimization problem with convex upper and lower problems.

A. Game Theory Based Algorithm

One difficulty in (12) is that the function $\mathcal{OPF}(\cdot)$ does not have a closed form expression in general. To overcome this difficulty, we relax (12) to be a two player game [9], [10]. Suppose there are two players Alice and Bob. Alice's pure strategy is given by $\mathbf{s}^g$, while Bob's is given by $\mathbf{s}^l$. For Alice, her set of pure strategies is defined by

$$\{\mathbf{s}^g \mid \exists\theta \text{s.t. (12a)(12b)(12c)(12e) are satisfied, given } \mathbf{s}^l\}$$

and she tries to maximize the payoff function $-\mathbf{f}(\mathbf{s}^g)$, i.e. her payoff is defined by the optimal power flow problem cost function in (5a). For Bob, his pure strategy set is defined by

$$\{\mathbf{s}^l \mid \exists\theta \text{s.t. (12a)(12b)(12d)(12e) are satisfied, given } \mathbf{s}^g\}$$

and he tries to maximize the payoff function $-\mathscr{L}(\mathbf{s}^g,\mathbf{s}^l)$. Both Alice and Bob want to maximize their own payoff functions by playing their best strategies. Since Alice and Bob do not know each other's strategy before they reveal their own strategies, if they happen to pick the strategy that is not in his or her set of pure strategies, then we define their payoff to be $-\infty$.

Any feasible point $(\mathbf{s}^g,\mathbf{s}^l)$ of (12) must be a pair of non-degenerate strategies for Alice and Bob (i.e. under such strategies, both Alice and Bob will have a finite payoff). Furthermore, for any optimal solution $(\mathbf{s}^g,\mathbf{s}^l)$ to (12), as $\mathbf{s}^g = \mathcal{OPF}(\mathbf{s}^l)$ and $\mathbf{s}^l$ minimizes $\sum_{i\in M}\left(S_i - \sum_{j\in R_i} s_j\right)^2$, $(\mathbf{s}^g,\mathbf{s}^l)$ must be a pure strategy Nash equilibrium of the game. It intuitively follows that we should develop an algorithm to find pure strategy Nash equilibrium of this two-player game. The numerical examples in Section V shows that the obtained pure strategy Nash equilibrium point is always an optimal solution to (12) provided that each region S_i contains at least one load bus.

We propose to use the *best response* algorithm [11] to find the Nash equilibrium: Alice and Bob alternatively select the optimal strategy from his or her feasible set according to the other player's previous strategy and his or her payoff function. If such procedure converges, then the limit point is one of the pure strategy Nash equilibrium points since the best response functions for both Alice and Bob are continuous within the set of non-degenerate strategies. In the future, we will investigate the convergence of the *best response* algorithm, as well as other promising algorithms in game theory [12].

Remark 2. *We have shown that the optimal solution set of (12) is the subset of the set of Nash equilibria for the game. However, a Nash equilibrium of the game is not necessarily an optimal solution to (12). This is the case when a region contains generators only, in which case Bob will not be able* *to minimize the aggregation error for that region by only manipulating the values of loads. Although we do not have the theoretical guarantee, in our simulations, where there are always more loads than generators in any region, the Nash equilibrium in most cases is also an optimal solution to (12).*

B. Bi-Level Programming Algorithm

Bi-level optimization problems [13], [14] occur frequently in planning and resource allocation problems. This type of optimization problem contains a constraint that is itself an optimization problem. Typically they are used to model the fact that decision making and optimization in hierarchical organizations cannot proceed from a single point of view. For example, an organisation may consist of several departments each with their own specialization. However, each department cannot act independently, and coordination is required in order to meet the demand of the overall organization. Another interpretation of a bi-level program is as a static, non-cooperative, two-person game, specifically a *Stackelberg* game. In this setting, the decision variables are partitioned between the two players; one is the *leader*, the second the *follower*. Both leader and follower seek to optimize their individual payoff functions. It is assumed that there is perfect information between the players, that is each player knows the other's payoff function and feasible sets. The game is *static* because each player only has one move. The leader makes the first move and attempts to minimize the net cost, and in so doing must anticipate what the follower will do. The follower observes the leader's decision and reacts in a manner that is optimal for her.

The bi-level optimization problem takes the following form:

$$(\mathbf{x},\mathbf{y}) = \arg\min_{\mathbf{x}\in X} \quad \mathbf{F}(\mathbf{x},\mathbf{y}) \tag{13a}$$

$$\text{subject to} \quad \mathbf{G}(\mathbf{x},\mathbf{y}) \leq 0 \tag{13b}$$

$$\mathbf{y} \in \arg\min_{\mathbf{y}\in Y} \quad \mathbf{f}(\mathbf{x},\mathbf{y}) \tag{13c}$$

$$\text{subject to} \quad \mathbf{g}(\mathbf{x},\mathbf{y}) \leq 0 \tag{13d}$$

where $\mathbf{x} \in X \subset \mathbb{R}^n$ is the leader's decision vector and $\mathbf{y} \in Y \subset \mathbb{R}^m$ is the follower's decision vector. We assume the functions $\mathbf{F}, \mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $\mathbf{G} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$, and $\mathbf{g} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$ are continuous and twice differentiable. We follow the formulation from the monograph [13].

The leader first decides $\mathbf{x}$, i.e. attempts to minimize $\mathbf{F}(\mathbf{x}, \mathbf{y}(\mathbf{x}))$. Here we have used the notation $\mathbf{y}(\mathbf{x})$ to indicate that the follower's decision is a function of the leader's choice of x. The sequential nature of the problem implies that $\mathbf{y}$ is a function of $\mathbf{x}$. To lighten notation we will not explicitly write this out, c.f. (13). Frequently in the literature the optimization problem (13a)–(13b) is referred to as the *upper* problem and (13c)–(13d) as the *lower* problem. We now list a few points that highlight the difficulties and peculiarities of a bi-level optimization problem:

- The order of play matters, this is in contrast to the non-cooperative game setting where both players act simultaneously. If we switch the upper and lower problems around, i.e. interchange (13a)–(13b) and (13c)–(13d), the solution will change.
- Even when $\mathbf{F}, \mathbf{f}, \mathbf{G}$ and $\mathbf{g}$ are continuous and X and Y are compact, a solution to (13) may not exist.
- Bi-level optimization is strongly NP-hard [15], and in fact (13) with linear functions $\mathbf{F}$ and $\mathbf{f}$ and affine functions $\mathbf{G}$ and $\mathbf{g}$ is too [16].

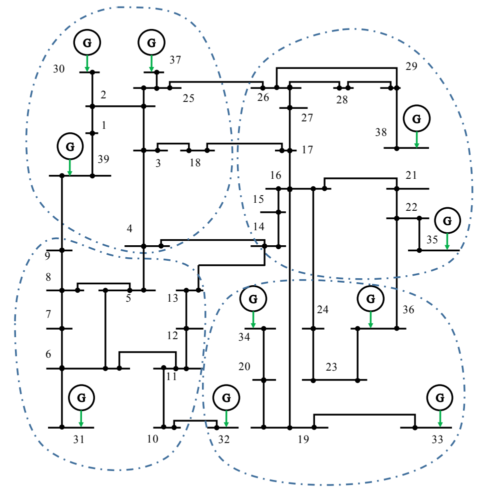

Figure 2. The IEEE 39-bus network with 4 aggregation regions.

There are several ways of specifying the disaggregation problem as a bi-level optimization program. Below we provide one such setup. Consider the following problem:

$$\begin{aligned} \min_{\mathbf{s}^l} \quad & \sum_{i\in M} \Big(\sum_{j\in R_i} s_j - S_i \Big)^2 \\ \text{subject to} \quad & \underline{s}_i^l \leq s_i^l \leq \bar{s}_i^l, \quad \forall i \\ & (\mathbf{s}^g, \theta) \in \arg\min \quad \mathbf{f}(\mathbf{s}^g, \boldsymbol{\theta}, \mathbf{s}^l) \\ & \qquad \text{subject to} \quad \mathbf{g}(\mathbf{s}^g, \boldsymbol{\theta}, \mathbf{s}^l) \leq 0 \end{aligned}$$

where the upper cost function defines the aggregation constraints and the loads are the decision variables. In the lower problem, $\mathbf{f}$ defines an OPF objective (linear or quadratic cost) as specified in (5a) , and $\mathbf{g}$ defines the OPF constraints (12a)–(12c) and (12e).

We will not go into any detail here, but it is important to note that parsers such as YALMIP [17] have in-built functionality to handle these problems. Specifically the problem is reduced to a set of bilinear constraints corresponding to the KKT conditions and locally solved using a nonlienar programming solver. For an overview on bi-level programming see the monographs [13], [14] and the survey paper [18].

V. Simulations

In this section we apply both methods to the IEEE 9-bus and IEEE 39-bus test networks. Both network descriptions are taken from the MATPOWER toolbox [19] for MATLAB.

In the IEEE 9-bus network, we divided the buses into 2 regions R_1 and R_2, as shown in Fig. 1. For the 39-bus case we divide it into 4 regions as shown in Figure 2. We consider two scenarios. In the first case we consider the noise free version of the disaggregation problem, corresponding to $\boldsymbol{\omega} = 0$. The ground truth data was created by running the DC OPF problem (5) with a linear cost function on the generators. Specifically for the 9-bus case we had $\mathbf{f}(\mathbf{s}^g) = \mathbf{c}^{\mathrm{T}}\mathbf{s}^g$ where

$$\mathbf{c} = \begin{bmatrix} 0.80 & 0.42 & 0.73 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}}.$$

The ground truth data for each bus is shown in the second row of Table I, where positive values correspond to generations and negative values to loads. The aggregation data is then simply the sum of the generations and loads in each region (shown in row 3 of Table I). In this case both the game theory and bi-level algorithm provided solutions that perfectly satisfied the aggregation constraints and achieved $\mathscr{L}(\mathbf{s}^g, \mathbf{s}^l) = 0$. Both methods however provided different loads and generations at the buses as can be seen in Table I. The same is true for the 39-bus test case. Here we omit the specific aggregation data but plot the individual loads and generations against ground truth in Figure 6.

In the second scenario we allow for the case where there is aggregation noise. We model $\boldsymbol{\omega}$ as a random vector drawn from independent and identically distributed zero mean Gaussian random variables $\mathcal{N}(\mathbf{0}, \gamma^2\mathbf{I})$ with variance γ^2. The signal-to-noise ratio SNR $:= \sum_{i=1}^{m} S_i^2/m\gamma^2$ defined as a function of the variance is used to parametrize a number of simulations. In Table II we show the disaggregated injections provided by both methods for SNR $= 20$ for the 9-bus test case. Note that the noise-free case in Table I is simply the case where SNR $= \infty$. In both cases the OPF constraints are satisfied but the loss functions take non-zero values. In Figure 3 the value of the loss function (obtained from the game theoretic formulation) as a function of SNR is plotted. Each point corresponds to the mean taken over 100 simulations and the bars mark the 90% confidence interval. For the case of the 39-bus system, Figure 4 plots the loss function against SNR. As expected, in both cases the mean value decreases as does the variance.

To illustrate the scalability of our method, we show in Figure 5 the convergence rate and computational time taken with respect to the IEEE 39 and 118-bus test case and the Synthetic 200-bus Illinois test case [20]. The termination condition is when the change in $(\mathbf{s}^g, \mathbf{s}^l)$ between adjacent iterations is less than 1%.

VI. Conclusion

We have defined the network disaggregation problem for power systems and provided two algorithms for solving the DC version of the problem. It is hoped that formulating the disaggregation problem in a general manner will encourage other researchers to study more complex variants of the problem, which in turn will allow utility companies and grid operators to distribute aggregated data with the confidence that ground truth solutions cannot be easily reverse engineered. There are several immediate research directions to pursue: can we consider more realistic OPF models, can we quantify the size of the set $\Gamma^\star$ in such a way so as to provide formal privacy guarantees, how can the problem be formulated when no knowledge of the OPF cost function is available?

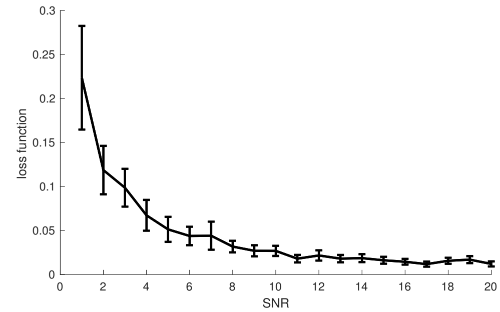

Figure 3. Cost function under different SNR values for IEEE 9-bus system.

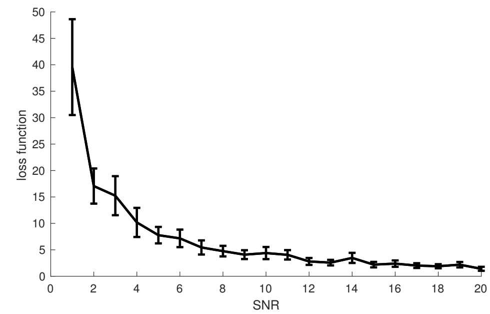

Figure 4. Cost function under different SNR values for IEEE 39-bus system.

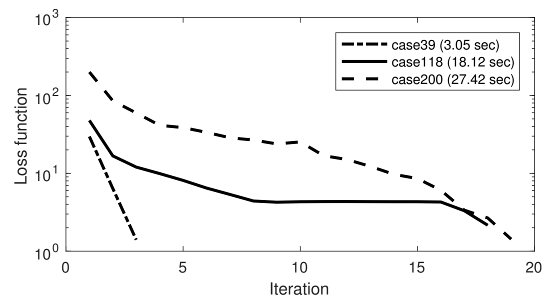

Figure 5. The loss function $\mathscr{L}$ converges over iterations for IEEE 39-bus case, IEEE 118-bus case and Synthetic 200-bus Illinois case [20].

References

[1] A. R. Borden, D. K. Molzahn, P. Ramanathan, and B. C. Lesieutre, "Confidentiality-preserving optimal power flow for cloud computing," in *50th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2012, pp. 1300–1307.

[2] A. R. Borden, D. K. Molzahn, B. C. Lesieutre, and P. Ramanathan, "Power system structure and confidentiality preserving transformation of optimal power flow problem," in *51st Annual Allerton Conference on Communication, Control, and Computing*, 2013, pp. 1021–1028.

[3] D. Wu, B. C. Lesieutre, and P. Ramanathan, "Feasibility of power system structure preserving linear transformations for the ac optimal power flow problem," in *52nd Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2014, pp. 715–722.

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

TABLE I
THE DISAGGREGATED RESULTS FOR BOTH METHODS (NOISE-FREE).

	Region R_1				Region R_2				
Bus ID	1	2	4	9	3	5	6	7	8
Ground Truth	140	35	0	-125	140	-90	0	-100	0
Aggregation: Ground Truth	$S_1 = 50$				$S_2 = -50$				
Output: Game Theory	140	0	-50.43	-39.57	62.67	-31.33	-29.47	-26.14	-25.72
Aggregation: Game Theory	50				-50				
Output: Bi-level	109.59	0	-31.30	-28.28	0	-12.71	-12.33	-12.37	-12.59
Aggregation: Bi-level	50				-50				

TABLE II
THE DISAGGREGATED RESULTS FOR BOTH METHODS (WITH SNR = 20).

	Region R_1				Region R_2				
Bus ID	1	2	4	9	3	5	6	7	8
Ground Truth	140	35	0	-125	140	-90	0	-100	0
Noisy Aggregation: Ground Truth	$S_1 = 41.67$				$S_2 = -40.74$				
Output: Game Theory	140	0	-59.67	-39.12	59.28	-26.89	-27.29	-23.65	-22.66
Aggregation: Game Theory	41.21				-41.21				
Output: Bi-level	140	28.7	-108.99	-100.92	140	-24.06	-26.75	-24.31	-23.68
Aggregation: Bi-level	41.21				-41.21				

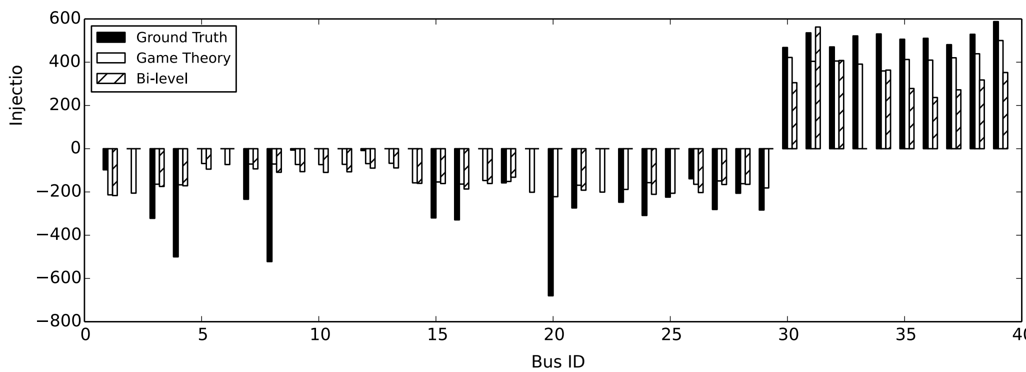

Figure 6. Comparison between the ground truth and the disaggregated results obtained from the game theory and bi-level algorithms.

[5] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *48th Annual IEEE Symposium on Foundations of Computer Science, (FOCS)*. IEEE, 2007, pp. 94–103.

[6] S. Han, U. Topcu, and G. J. Pappas, "Differentially private distributed constrained optimization," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 50–64, 2017.

[7] F. Zhou, J. Anderson, and S. Low, "Data disaggregation and privacy of power flow data," in *In preparation*, 2018.

[8] R. D. Zimmerman, "MATPOWER 4.0b2 user's manual," *Power Syst Eng Res Cent*, pp. 1–105, 2010.

[9] M. J. Osborne, *An introduction to game theory*, 2003, vol. 3, no. 3.

[10] D. Fudenberg and J. Tirole, *Game Theory*, 1st ed. The MIT Press, 1991, vol. 1.

[11] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press Cambridge, 2007, vol. 1.

[12] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 208–220, 2009.

[13] J. F. Bard, *Practical bilevel optimization: algorithms and applications*. Springer Science & Business Media, 2013, vol. 30.

[14] S. Dempe, *Foundations of bilevel programming*. Springer Science & Business Media, 2002.

[15] J. F. Bard, "Some properties of the bilevel programming problem," *Journal of optimization theory and applications*, vol. 68, no. 2, pp. 371–378, 1991.

[16] O. Ben-Ayed and C. E. Blair, "Computational difficulties of bilevel linear programming," *Operations Research*, vol. 38, no. 3, pp. 556–560, 1990.

[17] J. Lofberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*. IEEE, 2004, pp. 284–289.

[18] B. Colson, P. Marcotte, and G. Savard, "An overview of bilevel optimization," *Annals of operations research*, vol. 153, no. 1, pp. 235–256, 2007.

[19] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on power systems*, vol. 26, no. 1, pp. 12–19, 2011.

[20] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on power systems*, vol. 32, no. 4, pp. 3258–3265, 2017.