

On the Use of Matroid Theory for Distributed Cyber–Physical-Constrained Generator Scheduling in Smart Grid

Zakia Asad, *Student Member, IEEE*, Mohammad Assad Rehman Chaudhry, *Member, IEEE*, and Deepa Kundur, *Senior Member, IEEE*

Abstract—Power systems are on the cusp of a rapid technological, economic, and environmental evolution. Classical problems considered by the power community must naturally adapt to new requirements. Physical and computational constraints within a new landscape must be revisited. In this paper, we focus on the generator scheduling problem, which is also known as the unit commitment problem, in which we incorporate the new constraints of transmission line capacity limits and policy to give a more comprehensive view for planning in a smart grid. Moreover, we consider implications of our formulation to solution complexity. We introduce the concept of matroid theory to model unit commitment as a combinatorial problem and propose distributed solutions to facilitate optimal and correct generator scheduling. We account for communications and computational complexity and demonstrate how although the constrained generator scheduling problem is NP-hard, simpler versions of the problem lead to polynomial-time solutions.

Index Terms—Computational complexity, generator scheduling, matroid theory, NP-hardness, policy constraints, time-varying cost of generation, transmission-line constraints.

I. INTRODUCTION

A GROWING body of research on cyber–physical systems (CPSs) has focused on understanding the interactions between the control and communications infrastructure and the underlying physical system [1], [2]. A significant component particularly in the context of smart grid CPSs has focused on assessing the physical impacts of cyber integration and evolution ranging from the presence of advanced control to cyber attacks [3], [4].

Our work in this paper represents a dual to this latter problem as we study, in part, the implications of physical evolution on computing complexity. We explore this in the context of smart grid systems, which represent an ideal case study as they

Manuscript received February 24, 2014; accepted April 27, 2014. Date of publication June 25, 2014; date of current version December 19, 2014. This work was supported in part by the Schlumberger Foundation and in part by the Irish Research Council.

Z. Asad and D. Kundur are with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: z.asad@mail.utoronto.ca; dkundur@comm.utoronto.ca).

M. A. R. Chaudhry is with IBM Research, Dublin, Ireland, and also with Hamilton Institute, Kildare, Ireland (e-mail: asadchau@ie.ibm.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIE.2014.2331020

exhibit advances in both computational and communications (cyber) technology as well as power system (physical) components through the incorporation of renewable sources and other forms of distributed generation. We focus on the well-known optimization problem of generator scheduling to investigate how physical evolution leads to new constraints on the problem while cyber evolution facilitates distributed solutions. Our overall framework enables us to better understand how certain restrictive forms of physical constraints lend themselves naturally to polynomial-time solutions while their more general counterparts lead to NP-hard optimizations.

Our investigation involves reformulating the classical scheduling problem, but we emphasize that this is not intended as a proposal to replace existing generator scheduling formulations. Instead, our problem focus enables us to explore how complexity is affected to guide the long-term future evolution of these smart grid optimization tasks. Moreover, a better understanding of how the structure of cyber and physical constraints on the problem affect computational and communications complexity can enhance the prioritization of cyber–physical evolution.

A. Generator Scheduling

Scheduling generation plants to meet load demand at the least operating cost has historically been a cardinal problem in power systems. A number of tools including priority lists, dynamic programming, integer programming, branch and bound, simulate annealing, Lagrangian relaxation, and heuristic and genetic algorithms (see [5], [6] and references therein) have been applied to solve this known NP-hard problem [7]. In its classical setting, the generator scheduling problem involves the selection of generation units to supply forecasted load that minimizes cost over a required planning horizon subject to system constraints while leaving a specified margin of spinning reserve. The constraints typically include individual startup delays, generation costs, and physical operational constraints.

As we shift to a “smarter” cyber–physical grid in an deregulated market, we witness a fluctuating technical, political, and financial landscape in which decisions are constrained by additional issues including transmission-line bottlenecks and policy. For instance, although it may take one year to build a wind farm, it can take five years to build the necessary transmission lines needed to carry its power to cities [8]. Moreover, the smart grid vision incorporates consumer- and regulator-driven

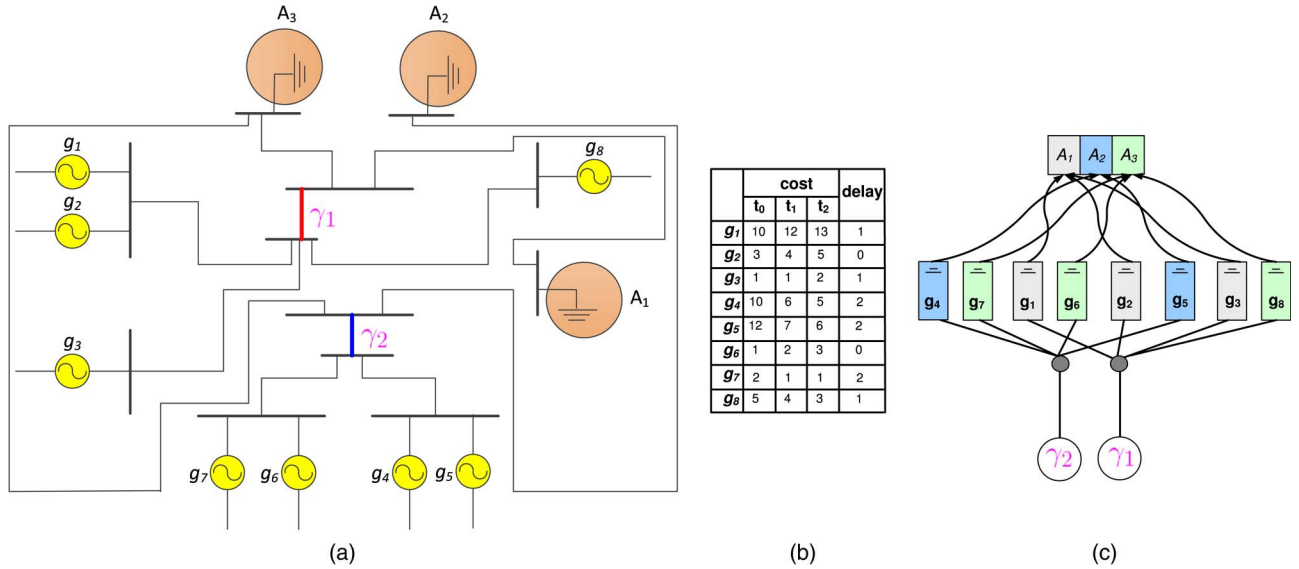


Fig. 1. (a) A 12-bus power system with eight generating points g_1, g_2, \dots, g_8 and their load areas to be served A_1, A_2, A_3 . (b) Startup delays and generation cost associated with each generating point for each time unit t_0, t_1, t_2 and a planning horizon of three time units. (c) An equivalent representation of the constraints imposed by the 12-bus power system.

policies such as priority of renewable sources providing novel system constraints.

In addition, evolving cyber infrastructure promotes wide-area monitoring and high-performance computing (HPC) applications for grid optimization necessitating that distributed computation be considered to avoid the high degree of coordination, data collection, and single point-of-failure weakness of centralized systems. Traditional NP-hard problems even at the planning phase are being revisited to study distributed strategies for computation. Moreover, smart grid as it is often regarding the evolution of the distribution system, we point out that many of the new challenges and opportunities affect the workings of the transmission system. For example, policy dictates constraints on the choices of energy sources for various areas in the distribution system, which affects the schedule of generators and their traffic in the transmission system, resulting in even stronger bond between distribution and transmission systems of the power networks. Thus, we assert that an important step in smart grid development involves revisiting classical optimization problems to study strategies to reduce their complexity in the face of these changing cyber-physical constraints.

In this vein, the objective of this paper is to explore the interaction of the problem structure with complexity in the context of emerging smart grid constraints. For this reason, we focus on the fundamental aspects of the generator scheduling problem to more effectively identify the tradeoffs and the relationship to complexity. In particular, we consider a generator scheduling problem in which we take into account a subset of constraints from its classical counterpart while exploring the integration of new requirements. We develop a modified formulation that lends itself to a better understanding of issues of complexity and tractability. Within this formulation, we explore solution complexity to ask: Can we solve the general problem in polynomial time? If not, what makes the structure of the problem NP-hard? For what special cases can we find efficient solutions? We emphasize that we are not proposing to replace existing generator scheduling algorithms. Instead,

we aim to assess insight into cyber-physical constraints and complexity to aid future opportunities for evolution.

We specifically address the following physical system requirements: 1) *delay constraints* including startup lags classically considered, along with novel 2) *transmission line constraints* involving line overloads that arise from usage by multiple generating points, and 3) *policy constraints* that deal with prioritization of generation source classes for specific load areas. We call this the *constrained generation scheduling* (CGS) problem. We associate a time-varying cost of operation to each generating source to reflect the changing cost of renewable power.

An instance of the CGS problem consists of a number of generating points, a set of transmission lines, and a set of load areas. Each generating point can represent a collection of generators, each with an associated delay. Each transmission line has a known capacity limit. Each load area has a forecasted demand and a set of geographical regions with policy-driven generation preferences. An example follows.

B. Example Case Study: 12-Bus Power System

Consider a 12-bus power system, which we shall be revisiting throughout this paper, with generating points g_1, g_2, \dots, g_8 , load areas to be served A_1, A_2, A_3 , and a set of two bottleneck transmission lines γ_1 (in red) and γ_2 (in blue) shown in Fig. 1(a). The sum from any group of two or more generating point outputs exceeds the transmission line capacities of γ_1 and γ_2 and, therefore, prohibits simultaneous use by more than one generating point. A careful study reveals the following facts.

If selected for dispatch, generating points $G_{\gamma_1} = \{g_1, g_2, g_3, g_8\}$ need to use one common transmission line γ_1 and, thus, cannot be simultaneously scheduled. Similarly, $G_{\gamma_2} = \{g_4, g_5, g_6, g_7\}$ cannot be simultaneously scheduled for dispatch on γ_2 . Thus, the transmission line constraints impose that no two generating points from G_{γ_1} and G_{γ_2} be scheduled to use γ_1 and γ_2 , respectively, at the same time. Moreover, due

to policy matters, each load area has preference of generating points. In this example, A_1 chooses $G_{A_1} = \{g_1, g_2, g_3\}$, A_2 opts for $G_{A_2} = \{g_4, g_5\}$, and area A_3 must fulfill demand with $G_{A_3} = \{g_6, g_7, g_8\}$. Thus, within the scope of the planning horizon, the transmission line and policy constraints dictate selection of only one generating point from each of G_{A_1} , G_{A_2} , and G_{A_3} for load areas A_1 , A_2 , and A_3 , respectively.

In this example, we assume that any generating point within G_{A_i} can accommodate the forecasted load for A_i ; if that is not the case, then more than one generating point from the set can be selected. A three-tier view of the constraints is shown in Fig. 1(c), where generating points are connected by lines to show the area they can serve and the shared transmission line they need to couple to. Moreover, for each generating point, the associated startup delays and the generation cost for a planning horizon of three time units t_0, t_1, t_2 is shown in Fig. 1(b).

C. Motivation

To motivate our work, we start by presenting a toy example for generator scheduling with minimal constraints. In this example, we are given a set of three generators g_1 , g_2 , and g_3 . To make the problem simpler and easier, we assume that each generator has different, but fixed over time, cost of generation and generation output. The costs of generation for generators g_1 , g_2 , and g_3 are 10, 40, and 20, respectively. The generation outputs for generators g_1 , g_2 , and g_3 are 6, 20, and 4, respectively. The generator scheduling just needs to satisfy basic power flow/transmission line and demand constraints specified by inequality and equality, respectively, in the mathematical program below. The variables x_1 , x_2 , and x_3 represent the binary (ON/OFF) decision-associated selection of generators g_1 , g_2 , and g_3 , respectively. Thus,

$$\begin{aligned} & \text{minimize} && 10x_1 + 40x_2 + 20x_3 \\ & \text{subject to} && 6x_1 + 20x_2 + 4x_3 \leq 15 \\ & && 6x_1 + 20x_2 + 4x_3 = 10 \\ & && x_i \in \{0, 1\} \quad 1 \leq i \leq 3. \end{aligned}$$

Based on the computational complexity (how fast the solution is produced) of the solution, there are two possible approaches to solve the mathematical program previously given as follows.

- 1) Perform relaxation operation, i.e., replace the constraints $x_i \in \{0, 1\}$ $1 \leq i \leq 3$, with $x_i \in \mathbb{R}$ $1 \leq i \leq 3$. After the relaxation operation, this mathematical program reduces to a linear program (LP). Solve the LP and then perform the rounding operation to convert the solution from real numbers (possibly irrational numbers) back to the binary (ON/OFF) domain. Any LP has a polynomial computational complexity, which results in the solution to be quickly produced. However, on the other side, the solution found using LP and rounding might not even be a feasible solution. For the given example, the optimal solution to the relaxed LP (i.e., $x_i \in \mathbb{R}$) is $x_1 = 1.6666\dots$, $x_2 = 0$, and $x_3 = 0$. First, note that $x_1 = 1.6666\dots$ is an irrational number and that it is not possible for any gen-

erator to tune the power output to any irrational number; rather, generator outputs are discrete. Furthermore, both rounding up ($\lceil x_1 \rceil = 2$) and rounding down ($\lfloor x_1 \rfloor = 1$) operations make the solution infeasible due to violation of the equality constraints in the original mathematical program. Hence, in general, linear programming cannot be used to solve generator scheduling problems even with basic minimal constraints.

- 2) Use integer LP (ILP)/mixed-integer LP (MILP) to solve the mathematical program, which can provide the feasible solution. However, ILP/MILP not only is an *NP-hard problem* itself [9], [10] but also has *no known polynomial-time approximation solution* since the Boolean satisfiability problem can be solved using ILP [11]. Furthermore, an NP-hard problem *with no known polynomial-time approximation solution* is believed to have exponential computational complexity [9], and the solution is not practically tractable for large instances [12]. Hence, such methods are not practically usable for future power systems where the scheduling instances are going to be very large due to enormous number of distributed generation sources, bidirectional flows, and microgrids. In short, given the complexity class of the ILP/MILP, trying to solve all instances of the generator scheduling problem using ILP/MILP is similar to *killing a bird with a big cannon*, i.e., using a much expensive method than needed. Hence, it is important to classify the problem into different complexity classes and then, based on the classification use, the most efficient solution in terms of computational complexity and resource solution.

Furthermore, the solutions for ILP/MILP are usually centralized, whereas the mammoth size of the future power systems calls for a distributed solution [13], [14]. Even the distributed solutions for ILP/MILP are in fact parallel algorithms loosely assuming synchronous models [15], [16], which are suitable only for colocated HPC clusters, for example, IBM's CPLEX on an HPC cluster. However, these solutions cannot be used for the distributed entities that not only are far apart geographically but also do not have a reliable communication mechanism as required by HPC clusters. Note that information and communication technology (ICT) support in the evolving smart grid is mostly based on asynchronous TCP and unreliable IP infrastructure; hence, there is a need for distributed solutions in solving the generator scheduling asynchronously while taking into account the complexity class of the problem instance. Furthermore, any efficient distributed solution utilizing an TCP/IP infrastructure needs to take into account the inherent rate-control limitations of TCP, which can be a result of huge number of messages exchanged (packets sent) over the network.

In this paper, we address these challenges. We present a general framework that can be used to capture a wide variety of constraints faced by evolving smart grids. Moreover, we classify the problem at hand based on computational complexity to identify easy and tough instances of the problem. The theoretical analysis in terms of computational and message complexity

presented in this paper is of particular interest in the context of designing industrial-scale systems. The analysis presented can be helpful in deciding the performance requirements of system components such as CPU/controller power, communication requirements on the interconnection TCP/IP network, etc. We use matroid theory to develop an efficient/fast solution with polynomial-time computational complexity, least utilization of communication network, and CPU/controller usage.

D. Contribution

The evolution of smart grid invites a new era of optimization problems and resource scheduling (including generator scheduling), where joint actions of all users are coordinated, to allocate system resources in a distributed fashion under a set of novel constraints. Previously, from the perspective of active distribution networks (ADNs), scheduling of distributed energy resources (DERs) with respect to the network topology has been considered for the optimal coordination of energy resources [17]. In [18], the production cost minimization and network constraint management for active management of DERs are presented. Distributed generation (DG) has been explored in terms of compact integrated energy systems that use 25% less semiconductors in [19]. Control techniques for the integration of DG resources into the electrical power network have been studied in [20]. Furthermore, in [21], the dynamic allocation of the resources in the dispatch of DG using an evolutionary game-theoretical approach for optimal and feasible solutions in a microgrid structure is studied. Moreover, the placement of generators in DG for loss reduction in a primary distribution network is studied in [22]. In [23], a distribution optimal power flow model to integrate local distribution system feeders into a smart grid is presented. In [24], a jump-and-shift method is developed to handle multiobjective optimization to minimize the operating costs by reducing the emission levels while meeting the load demand.

One of the primary objectives of this paper is to explore how the emerging constraints can significantly affect the complexity of the problem. Therefore, we deviate from classical formulation to consider a new approach that allows us to use matroid theory to get a better insight into the problem. We believe that studying a problem that deviates from the classical formulation shall help shed light on the interaction between constraints and complexity.

Particularly, in this paper, we consider generator scheduling in the presence of several physical constraints imposed by the evolving smart grid. Although intelligent generation scheduling (involving unit commitment [25] and economic dispatch [26]) have been widely studied, but classically, generator scheduling has not accounted for policy and transmission line constraints together, often leaving its inclusion to the next stage of economic dispatch. However, the growing trend in power systems is one in which these limitations must be accounted for early on during the scheduling phase to provide a more comprehensive view [27]. Therefore, research has also been done to solve the generator scheduling problem for real-life large-scale power systems (e.g., see [28]) and also for generator scheduling in the presence of security constraints [29]

and transmission line capacity constraints [30], but the drift toward a “smarter” cyber–physical grid calls for the inclusion of policy-based constraints. Previously, several algorithms have been presented for generation scheduling where the operational policy in terms of the number of control action [31], [32] has been considered. Similarly, in [33] and [34], the maximization of social welfare and management of environmental impact issues are explored. A generalized formulation for artificial-intelligence-based intelligent energy management to minimize the operation cost and the environmental impact of a microgrid has been explored in [35]. However, the vision of smart grid calls for consumer-driven policy initiatives in addition to the social welfare, which is not addressed before, e.g., the option of explicitly selecting from the energy sources, which is addressed in this paper.

In short, scheduling generation to match demand while simultaneously meeting constraints involves extensive combinatorics due to the discrete nature of the problem; a generator is either selected or not. In general, choosing the least cost combination of generating points requires exhaustive search.

An alternative approach could relax the integer constraints imposed by the problem and solve it via linear programming; however, such a solution may lack optimality, and in some cases, rounding might result in an infeasible solution. Thus, in this work, we propose an approach for combinatorial optimization, suitable for distributed computing, that aims to leverage the natural symmetry and structure of the CGS problem constraints while providing insight into issues of computational complexity. We focus on the application of *matroid theory* [36] to the smart grid scheduling problem. The structure of matroids enables one to effectively rule out large groups of generator sets for scheduling with a simple test in contrast to sifting through all possible combinations.

Our contributions are threefold.

- 1) We develop a framework for CGS in which the problem is broken down into building blocks, effective in relating scheduling complexity to physical system constraints; specifically, we introduce the *I-restricted constrained generation scheduling (I-RCGS)* and the *T-restricted constrained generation scheduling (T-RCGS)* problems.
- 2) We present distributed polynomial-time solutions to the *I-RCGS* and *T-RCGS* problems. The *I-RCGS* problem assumes zero delay and a planning horizon of one time unit, thus dealing only with transmission line and policy constraints, which can be represented as matroids. The *T-RCGS* problem extends the *I-RCGS* problem to a planning horizon of T time units and includes delay constraints. Our distributed solution proposes a novel way for assignment of IDs to the generating points to facilitate the distributed solution.
- 3) We show how for less-restrictive physical constraints the CGS problem is NP-hard to approximate within a ratio of $n^{1-\epsilon}$ for any constant $\epsilon > 0$.

II. PROBLEM FORMULATION

We consider scheduling of a set of n generating points $G = \{g_1, \dots, g_n\}$. Each generating point is associated with

a variable $u_{g_i,t}$, where $u_{g_i,t} \neq 0$ means that generating unit g_i is on/up at time t and generating power $W_{g_i,t}$ such that $W_{g_i}^{\min} \leq W_{g_i,t} \leq W_{g_i}^{\max}$. There are q buses $B = \{b_1, \dots, b_q\}$, let \wedge_{b_i} be the index of the generating units at bus b_i and $D_{b_i,t}$ be the forecasted demand at bus b_i in time period t . We assume ℓ transmission lines $\Gamma = \{\gamma_1, \dots, \gamma_\ell\}$ and transmission capacity F_{γ_i} for each transmission line γ_i . Let $\tau_{\{\gamma_i, b_j\}}$ be the line flow distribution factor at the transmission line γ_i by bus b_j . We further assume k load areas A_1, \dots, A_k with forecasted load demand D_{A_i} for area A_i and a planning horizon T . Note that transmission lines connect different G 's with different A 's. For the case study in Section I-B, the corresponding set of generating points is $G = \{g_1, g_2, \dots, g_8\}$, and the set of (bottleneck) transmission lines is $\Gamma = \{\gamma_1, \gamma_2\}$, load areas are A_1, A_2 , and A_3 .

Each generating point is associated with a time-dependent generation cost $c: G \times \{0, 1, \dots, T-1\} \rightarrow \mathbb{R} \geq 0$ that assigns nonnegative costs to each generating point g_i over all time units $t_j \in \{0, 1, \dots, T-1\}$ such that $c(g_i, t_j)$ refers to the cost of generation cost of g_i . Since the generation cost of a generating point can choose any value from a set of real numbers, it incorporates nonlinearity and nonconvexity as a result of valve effects. In addition, each generating point g_i is associated with a generation delay $d(g_i) \in \{0, 1, \dots\}$ to reflect wait times in startup. If g_i is *allocated* a desired transmission line at t_j , it implies that the actual time when g_i couples to the transmission line is $d(g_i) + t_j$.

An *allocation at time* t_i denoted a_{t_i} is a set of generating points that are allocated their desired transmission line at time t_i . A *schedule* is defined to be the set of allocations at times $0, 1, \dots, T-1$ for a planning horizon of T . A *feasible schedule* is defined to be a schedule that satisfies the following three constraints:

- 1) **Transmission line constraints** captured by the dc power flow model

$$\sum_{b_j} \tau_{\{\gamma_i, b_j\}} \left(\sum_{k \in \wedge_{b_j}} u_{g_k,t} W_{g_k,t} - D_{b_j} \right) \leq |F_{\gamma_i}|, \forall \gamma_i \in \Gamma, t \leq T.$$

- 2) **Delay constraints** whereby in a feasible allocation and for a planning horizon T , a generating point g_i can only be allocated its required transmission line at time t_j if $d(g_i) + t_j \leq T-1$. Delay constraint captures satisfies the *minimum downtime* of a generating point, i.e., $d(g_i)$ is the minimum downtime.
- 3) **Policy constraints** regarding meeting the demand constraints of the load areas while respecting policy initiatives of each area. More specifically, each area A_i is associated with a set $s_i \subseteq G$ of generating points opted by it to meet its demand. For each s_i , we define a variable $f(s_i) \in \{0, 1\}$, where $f(s_i) = 1$ if area A_i has been provided the service, and $f(s_i) = 0$ otherwise. Since a goal is to provide service to as many service areas as possible, formally, the policy constraint is as follows: $\max_i \sum_{s_i} f(s_i)$.

The *cost of an allocation* a_{t_i} is defined to be the sum of the generation costs of all the generating points in a_{t_i} given

by: $\text{cost}(a_{t_i}) = \sum_{g_j: g_j \in a_{t_i}} c(g_j, t_i)$. Subsequently, the *cost of a schedule* Q is given by of the sum of costs over all its allocations given by: $\text{cost}(Q) = \sum_{a_{t_i}: a_{t_i} \in Q} \text{cost}(a_{t_i})$.

Definition 1 (CGS Problem): For an instance of the CGS problem with transmission line, delay, and policy constraints as previously defined, find a least cost feasible schedule; that is, no other feasible schedule has a lower cost.

Informally, we can say that the objective of the CGS problem is to provide service to the maximum possible number of service areas while respecting their policy initiatives without violating the transmission line and delay constraints such that the solution has the least cost among all constrained possibilities.

Note that we choose to ignore the ramp constraints since it can be incorporated in generator scheduling treating it as local constraints at the bus level [37].

III. RCGS

Here, we deal with a restricted version of the CGS problem, called the restricted constrained generation scheduling (RCGS) problem, where $W_i = D_j, \forall g_i \in s_j$ and $W_{g_i} < F_{\gamma_j} < 2W_{g_i}, \forall g_i \in \wedge_{b_k}$ with $\tau_{\{\gamma_j, b_k\}} \neq 0$ (i.e., bus b_k injects flow to the transmission line γ_j).

Later, in Section IV, we show that without this restriction, the problem becomes NP-hard. We start by presenting the solution to the *I-RCGS* problem, which exhibits the following additional restrictions:

- $T = 1$, which naturally implies
- $d(g_i) = 0 \forall i$ and
- $c(g_i, t_j) = c(g_i) \forall i, j$.

It is obvious that the least cost *schedule* for the *I-RCGS* problem is the same as least cost *allocation* at time $t_i = 0$ since the planning horizon is just one time unit. Thus, we may use the solution to the *I-RCGS* problem as a building block for the solution of the *T-RCGS* problem with planning horizon T .

A. Relation to Matroids

To apply matroid theory, we first capture the constraints in Section II as matroids.

Definition 2 (Matroid): A *matroid* $\mathcal{M}(X, \mathcal{L})$ is an ordered pair formed by a *ground set* X and a collection \mathcal{L} of subsets of X , which satisfy the following three conditions:

- $\emptyset \in \mathcal{L}$;
- if $Y \in \mathcal{L}$ and $Y' \subseteq Y$, then $Y' \in \mathcal{L}$;
- if $Y_1 \in \mathcal{L}, Y_2 \in \mathcal{L}, |Y_1| > |Y_2|$, then there exists $x \in Y_1 \setminus Y_2$ such that $Y_2 \cup \{x\} \in \mathcal{L}$.

Each $Y \in \mathcal{L}$ is referred to as an *independent set*.

1) Policy Constraints: We first define a set $PC(G) = \{s_1, \dots, s_k\}$ consisting of distinct (nonoverlapping) policy sets $s_i \subseteq G$ defined for each load area A_i . For the service areas A_1, A_2 , and A_3 in the case study of Section I-B, the corresponding policy constraint set is: $PC(G) = \{\{g_1, g_2, g_3\}, \{g_4, g_5\}, \{g_6, g_7, g_8\}\}$. To capture the policy constraints, we start by working in k -dimensional vector space, with k basis vectors given by b_1, b_2, \dots, b_k . We further define a matrix A with k rows and n columns. Each generating point

$g_i \in G$ is associated with a unique column vector z_i in matrix A . Each z_i is a vector in the k -dimensional vector space. To each policy set $s_i \in SC(G)$, we assign a unique base b_i . Moreover, all the generating points in s_i are assigned unique column vectors that are positive (sequential) integer multiples of b_i such that they are unique but linearly dependent over b_i . Specifically, if $g_j, g_k, \dots, g_{y+x} \in s_i$, then we assign vector $z_j = b_i$ to g_j , $z_k = 2 \cdot b_i$ to g_k , and similarly to the remaining generating points using consecutive integer multiples of b_i .

We define $\mathcal{L}_1 = \{L_1\}$ where $L_1 \subseteq G$, such that all the column vectors z_i associated with $g_i \in L_1$ are linearly independent for all i ; that is, \mathcal{L}_1 is the family of all subsets of G whose associated column vectors are linearly independent.

Lemma 3: $\mathbf{SM} \triangleq (G, \mathcal{L}_1)$ is a vector matroid over ground set G , with a collection of independent sets given by \mathcal{L}_1 .

Proof: Follows directly from the definition of vector matroid [38]. ■

For our case study in Section I-B, matrix A is given as follows:

$$A = \begin{Bmatrix} 1 & 2 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 \end{Bmatrix}.$$

In this example, the ground set for \mathbf{SM} is $G = \{g_1, \dots, g_8\}$. Furthermore, \mathcal{L}_1 is the set of all subsets of G whose associated column vectors are linearly independent; an example of one such set is $\{g_1, g_4, g_7\}$.

2) Transmission Line Constraints: We define a family of subsets of G , $TLC(G) = \{x_1, \dots, x_R | x_i \subseteq G \forall i\}$, where set x_i contains all generating points that require line γ_i but have conflicts such that their total output exceeds the line capacity F_{γ_i} . Let \mathcal{L}_2 be the family of all subsets of G that are feasible with respect to transmission line constraints, that is, they do not have transmission line conflicts. Formally, $\mathcal{L}_2 = \{L_2 | L_2 \subseteq L'_2\}$, where $L'_2 \subseteq 2^G$ such that $L'_2 = \{e_1, \dots, e_k | e_1 \in x_1, \dots, e_k \in x_k\}$ and $k \leq \ell$.

Lemma 4: $\mathbf{TLM} \triangleq (G, \mathcal{L}_2)$ is a partition matroid over ground set G , with the collection of independent sets given by \mathcal{L}_2 .

Proof: The proof is based on two observations. The first is that each L_2 is a partial transversal of the transmission line constraint set $TLC(G) = \{x_1, \dots, x_\ell\}$; hence, \mathcal{L}_2 is the set of all the partial transversals of the transmission line constraint set $TLC(G)$. The second observation is that the transmission line constraint set $TLC(G)$ defines a partition of G . Hence, by definition of a partition matroid [38], \mathbf{TLM} is a partition matroid over ground set G . ■

For the case study in Section I-B, the ground set for \mathbf{TLM} is G . \mathcal{L}_2 is the set of all partial transversals of $TLC(G)$; one such component is given by $\{g_1, g_3\}$.

B. Distributed Algorithm for the 1-RCGS Problem

It has been shown in the previous section that policy constraints can be captured by a vector matroid $\mathbf{SM} = (G, \mathcal{L}_1)$, and transmission line constraints can be captured by a partition

matroid $\mathbf{TLM} = (G, \mathcal{L}_2)$. Moreover, both constraints need to be satisfied in finding a feasible solution for the I -RCGS problem.

Specifically, the I -RCGS problem can be solved by finding the maximum cardinality intersection of \mathcal{L}_1 and \mathcal{L}_2 (to maximize areas served) with least cost.

Lemma 5: The optimal solution to I -RCGS is given by the least cost maximal cardinality common independent set of matroids \mathbf{SM} and \mathbf{TLM} .

The proof of Lemma 5 is given in Appendix A.

We first give a brief overview of the proposed *distributed* algorithm. Each generating point executes a copy of Algorithm I -RCGS based on Edmonds' weighted matroid intersection algorithm [38]–[40].

We assume a synchronous model in which communication is executed in rounds. Moreover, the r th round, i.e., round $_r$, is further subdivided into three phases denoted *Phase 1*, *Phase 2*, and *Phase 3*. For the sake of simplicity, we assume that each generating point can communicate with any other generating point (via a possible multihop or mesh network). Moreover, each generating point g_i has a unique vector ID id_i that consists of two parts. The first part id_i^1 corresponds to the vector z_i assigned to it (used for capturing policy constraints by using a vector matroid representation as mentioned in Section III-A). The second part id_i^2 indicates the transmission line it must couple with; for instance, the second part of id_i for the generating point g_i is j if $g_i \in x_j$. For our case study, the ID for g_1 is $id_1 = 1001$, where $id_1^1 = z_1 = 100$, and $id_1^2 = 1$ since $g_1 \in x_1$. We assume that each generating point knows the IDs of all the other generating points. The selection of generating points in the optimal solution needs at most ℓ rounds; this is because the exact number of rounds needed is given by the maximum possible areas served, which is bounded from above by ℓ .

The formal Algorithm I -RCGS for generating point g_i is shown in Fig. 2, whereas Subroutine $Chk\mathcal{L}$ is given in Fig. 3.

Theorem 6: The distributed Algorithm I -RCGS gives an optimal solution to the I -RCGS problem.

The sketch of proof of Theorem 6 is given in Appendix B.

C. Complexity Analysis

The complexity of Algorithm I -RCGS is measured in terms of the number of messages exchanged M and the local computation time at each generating point T_{local} .

The number of messages sent across the network in *Phase 2* is incurred in two steps. 1) The first is in computing the shortest path from all generating points in \mathbf{G}_1 to all generating points in \mathbf{G}_2 , which, in the worst case, can turn out to be all pairs shortest paths. The complexity in terms of the number of messages for the all pairs shortest paths algorithm given by Haldar [41] is $2n^2$ messages. 2) In the worst case, if $|\mathbf{G}_1| = |\mathbf{G}_2| = n$, the number of messages sent is n^2 . In *Phase 3*, the source node of the shortest selected path sends n messages one to each generating point. Therefore, the total message complexity for the three phases of a round is $O(n^2)$. There can be, at most, ℓ rounds. Therefore, the total number of messages exchanged is $M = O(\ell n^2)$.

Algorithm 1 – RCGS (G): Code for $g_i, 1 \leq i \leq n$

- 1 When init():
- 2 $G_{on}(g) = 0 \forall g$
- 3 $c(g_i) = c(g_i)$
- 4 For each round
- 5 **Phase 1:**
- 6 $\overleftarrow{E}(g)=0, \overrightarrow{E}(g)=0, \mathbf{G}_1(g)=0, \mathbf{G}_2(g)=0 \forall g$
- 7 Compute routing table
- 8 If $G_{on}(g_i) = 1$;
- 9 For each g_j such that $G_{on}(g_j) = 0$
- 10 $G' = G_{on}$
- 11 $G'(g_i) = 0$
- 12 $\overrightarrow{E}(g_j) = Chk\mathcal{L}(g_j, G', 1)$
- 13 $\overleftarrow{E}(g_j) = Chk\mathcal{L}(g_j, G', 2)$
- 14 Else
- 15 For each g_j such that $G_{on}(g_j) = 1$
- 16 $G' = G_{on}$
- 17 $G'(g_j) = 0$
- 18 $\overrightarrow{E}(g_j) = Chk\mathcal{L}(g_i, G', 1)$
- 19 $\overleftarrow{E}(g_j) = Chk\mathcal{L}(g_i, G', 2)$
- 20 For each g_j such that $G_{on}[g_j] = 0$
- 21 $\mathbf{G}_1(g_j) = Chk\mathcal{L}(g_j, G_{on}, 1)$
- 22 $\mathbf{G}_2(g_j) = Chk\mathcal{L}(g_j, G_{on}, 2)$
- 23 **Phase 2:**
- 24 If $\mathbf{G}_1(g_i) = 1$
- 25 For all $\mathbf{G}_2(g_j) = 1$
- 26 Find a least cost path from g_i to g_j using routing tables \overleftarrow{E} and \overrightarrow{E}
- 27 Let g_m be the generating points connected through least cost path $\mathfrak{G}_{g_i \rightsquigarrow g_m}$ with cost c_m^i and hop count h_m^i
- 28 Send($\langle c_m^i, h_m^i \rangle$) to every generating point g_j with $\mathbf{G}_1(g_j) = 1$
- 29 Upon Receive($\langle c_m^j, h_m^j \rangle$) from every processor g_j with $\mathbf{G}_1(g_j) = 1$
- 30 $c_m^* = \min_{k|G_1(g_k)=1} (c_m^k)$. If $c_m^k = c_m^*$ for more than one generating point then choose the generating point with least hop count h_m^* . Let
- 31 $\mathfrak{G}_{g_i' \rightsquigarrow g_m'}$ be the least cost path among all paths $\mathfrak{G}_{g_k \rightsquigarrow g_j}$ such that $\mathbf{G}_1(g_k) = 1$ and $\mathbf{G}_2(g_j) = 1$
- 32 **Phase 3:**
- 33 If $c_m^* = \text{infinity}$
- 34 Then G_{on} is a maximum-cardinality common independent set. Send($\langle \text{stop} \rangle$) to all generating point
- 35 Else-If $g_i' = g_i$
- 36 Let the path $\mathfrak{G}_{g_i' \rightsquigarrow g_m'}$ = $\{y_0, z_1, y_1, \dots, z_l, y_l\}$.
- 37 $G'_{on}(g_k) = 1$ such that $g_k \in \{g | G_{on}(g) = 1\} \setminus \{z_1, \dots, z_l\} \cup \{y_0, \dots, y_l\}$
- 38 Send($\langle G'_{on} \rangle$) to all generating point
- 39 When Receive($\langle G'_{on} \rangle$):
- 40 Update G_{on}
- 41 if $G_{on}[g_i] = 1$
- 42 $c(g_i) = -c(g_i)$
- 43 Else
- 44 $c(g_i) = c(g_i)$
- 45 goto step 4
- 46 When Receive($\langle \text{stop} \rangle$):
- 47 If $G_{on}[g_i] = 1$, then start working.

Fig. 2. Algorithm 1-RCGS.

The time complexity is incurred in the following steps of each round. The time to fill the routing tables using subroutine $Chk\mathcal{L}$ is $O(n^2)$, as each generating point checks its neighborhood at most with all other generating points; precisely checking for one entry in vector \overrightarrow{E} requires $O(n)$ steps. Similarly, computing entries of vectors \overleftarrow{E} , \mathbf{G}_1 and \mathbf{G}_2 also requires $O(n^2)$ computations each. Steps 24 and 27 require the identification of generating points with minimum distance

Algorithm $b = Chk\mathcal{L}(g_k, G', num)$:

- 1 $b=1$;
- 2 If $num=1$
- 3 For g_j with $G'[g_j] = 1$
- 4 if id_k^1 and id_j^1 are linearly dependent
- 5 Return $b=0$;
- 6 If $num=2$
- 7 For g_j with $G'[g_j] = 1$
- 8 if $id_k^2=id_j^2$
- 9 Return $b=0$;

Fig. 3. Algorithm $b = Chk\mathcal{L}$.

and hops, which necessitates at most $O(n^2)$ computations. Therefore, in each round, the total local time complexity at each generating point is at most $O(n^2)$, and the total complexity in all rounds is $T_{\text{local}} = O(\ell n^2)$.

D. T-RCGS Problem

In Section III, we presented a solution for the *I-RCGS* problem. Here, we extend this solution to the *T-RCGS* problem. We start by extending matroids **SM** and **TLM** to capture time-varying generation costs $c(g_i, t_j)$ and generating point delays $d(g_i)$. This extension is based on the concept of time-expanded matroids [42].

1) Extended Ground Set: We start by extending the ground set G to the extended ground set EG , i.e.,

$$EG \triangleq \{(g_i, t_j) | g_i \in G, t_j \in 0, \dots, T-1, d(g_i) + t_j \leq T-1\}.$$

EG consists of all generating points and their possible allocation times represented as a set of pairs (g_i, t_j) that satisfy the delay constraints. For the 12-bus case study in Section I-B, the extended ground set is: $EG = \{(g_1, 0), \dots, (g_8, 0), (g_1, 1), (g_2, 1), (g_3, 1), (g_6, 1), (g_8, 1), (g_2, 2), (g_6, 2)\}$.

Policy Constraints: We extend **SM**, the matroid capturing the policy constraints as follows. First, with each $(g_i, t_m) \in EG$, we associate a column vector z_i . This vector is the same column vector z_i as that associated with g_i discussed in Section III-A.1. We define $\mathfrak{EL}_1 = \{\{EL_1\} | EL_1 \subseteq EG\}$ such that all the $(g_i, t_m) \in EL_1$ have linearly independent associated column vectors for fixed time. In other words, \mathfrak{EL}_1 is the family of all subsets of EG whose associated column vectors are linearly independent.

We define: $\mathbf{ESM} \triangleq (EG, \mathfrak{EL}_1)$.

For the 12-bus case study, an example of a set EL_1 belonging to \mathfrak{EL}_1 is $EL_1 = \{(g_1, 0), (g_4, 0), (g_7, 0)\}$, where the column vectors associated with $(g_1, 0)$, $(g_4, 0)$, and $(g_7, 0)$ are 100, 010, and 002, respectively.

Transmission Line Constraints: We extend **TLM**, the matroid capturing the transmission line constraints as follows. Define $\mathfrak{EL}_2 = \{EL_2 | EL_2 \subseteq \bigcup_{j=0}^{T-1} EL_{2,t_j}, EL_{2,t_j} \subseteq EG\}$ such that for each $(g_i, t_m) \in EL_{2,t_j}$ $d(g_i) + t_m = t_j$, and for all $(g_i, t_m), (g_r, t_p) \in EL_{2,t_j}$ g_i and g_r satisfy the line constraints. Thus, each $\{EL_{2,t_j}\}$ is the family of subsets of EG , which can use their required transmission line at time t_j , without violating the line constraints.

We define: $\mathbf{ETLM} \triangleq (EG, \mathfrak{EL}_2)$.

Algorithm T – RCGS (\cdot): Code for g_i , $1 \leq i \leq n$

- 1 Generate

$$EG_{g_i} \triangleq \{(g, t_j) | g = g_i, t_j \in 0, \dots, T-1, d(g_i) + t_j \leq T-1\}$$
- 2 For each generating point $(g_i, t_j) \in EG_{g_i}$, assign a vector $id_{(i,j)}$ as follows:
 - 3 $id_{(i,j)}^1 = z_i$
 - 4 $id_{(i,j)}^2 = k$ such that $g_i \in x_k$,
 - 5 $id_{(i,j)}^3 = j + d(g_i)$
- 6 Broadcast $id_{(i,j)}$ for all $(g_i, t_j) \in EG_{g_i}$ to all the other generating points
- 7 For each $(g_i, t_j) \in EG_{g_i}$, call Algorithm *RGM* with the following modification in subroutine *ChkQ* given in Fig. 3.
- 8 Step 8 will be modified to “If $id_k^2 = id_j^2$ AND $id_k^3 = id_j^3$ ”

Fig. 4. Algorithm T-RCGS.

For the 12-bus case study, an example of a set EL_2 belonging to $\mathcal{E}\mathcal{L}_2$ is $EL_2 = \{(g_3, 0), (g_6, 0), (g_4, 0)\}$. Note that $(g_3, 0)$, $(g_6, 0)$, and $(g_4, 0)$ all require the same transmission line, that is, γ_1 ; however, they will actually couple to the transmission line at time slots 1, 2, and 0, respectively, and, therefore, do not violate the transmission line constraint.

Lemma 7: Both **ESM** and **ETLM** are matroids. Furthermore, the optimal solution to the *T-RCGS* problem is given by the least cost maximum cardinality common independent set of matroids **ESM** and **ETLM**.

Proof: The fact that both **ESM** and **ETLM** are matroids follows directly from the concept of time-expanded matroids [42]. Then, using the similar arguments as in the proof of Lemma 5, it can be shown that the intersection of **ESM** and **ETLM** satisfies the policy, transmission line, and generation delay constraints, and the optimal solution to the *T-RCGS* problem shall be the least cost maximum cardinality common independent set of **ESM** and **ETLM**. ■

A sketch solution for the *T-RCGS* problem, i.e., Algorithm *T-RCGS*, is shown in Fig. 4. In Algorithm *T-RCGS*, a generating point g_i runs multiple copies of Algorithm *I-RCGS*, one for each $(g_i, t_j) \in EG_{g_i}$, where EG_{g_i} represents copies of g_i at different time slots. For the 12-bus power system example, the execution of Algorithm *T-RCGS* would schedule generating points g_3 , g_4 and g_6 all at $t = 0$.

Theorem 8: The Algorithm *T-RCGS*(EG) provides an optimal solution to the *T-RCGS* problem.

Proof: The correctness of algorithms follows from Lemma 7 and correctness of the weighted matroid intersection algorithm by Edmonds [39]. ■

Lemma 9: For Algorithm *T-RCGS*, the number of messages exchanged is $M = O(T\ell n^2)$, and the local time complexity is $T_{\text{local}} = O(T\ell n^2)$.

Proof: The complexity of Algorithm *T-RCGS* follows directly from the complexity of Algorithm *I-RCGS*. The only difference is the increased number of generating points (each generating point is copied in the worst case for each time slot), which is precisely T times more than the actual set of generating points. Therefore, message and local time complexity is multiplied by a factor T as compared with Algorithm *I-RCGS*. ■

IV. CGS PROBLEM

Here, we show that the *CGS* problem is NP-hard by a reduction from the optimization version of the independent set (IS) problem, i.e., the *maximum independent set*. Before proceeding to the proof of NP-hardness, we give the conflict graph representation of the transmission line constraints.

Definition 10 (Conflict Graph): A conflict graph $G(V, E)$ has a vertex set defined as $V = \{g_1, \dots, g_n\}$ and an edge set defined as $E = \{(g_i, g_j) | W_i + W_j > L_k, \text{ where } g_i, g_j \in x_k\}$.

An instance *IS* for the maximum independent set problem is defined by a graph $H(V', E')$, with vertex set V' and edge set E' . The objective is to find a maximum cardinality subset of V' such that the vertices in this subset are not directly connected to each other. We define the corresponding instance of the *CGS* problem as follows.

- $G = V'$, and $g_i = v'_i \forall i$, i.e., each vertex $v_i \in V'$ corresponds to a generating point g_i and a load area A_i in the instance of the *CGS* problem.
- The policy constraint set is: $PC(G) = \{s_1, \dots, s_{|V'|}\}$, where $s_1 = \{g_1\}, \dots, s_{|V'|} = \{g_{|V'|}\}$.
- The conflict graph $G(V, E)$, capturing the transmission line constraints is defined as: $V = V'$, and $E = \{(g_i, g_j) | (v'_i, v'_j) \in E'\}$. Thus, any two generators (g_i, g_j) cause overload for the competing transmission line if (v'_i, v'_j) is an edge in E' .
- For the delay constraints:
 - the planning horizon is defined to be $T - 1 = 0$ or $T = 1$; and
 - $d(g_i) = 0 \forall i$
- $c(g_i, t_j) = 1 \forall i, j$.

Theorem 11: The *CGS* problem is not only NP-hard but also NP-hard to approximate within a ratio of $n^{1-\epsilon}$ for any constant $\epsilon > 0$.

Sketch of proof: Due to the one-to-one correspondence of graphs $H'(V', E')$ and $G(V, E)$, a maximum independent set in $H'(V', E')$ is the same as a maximum independent set in $G(V, E)$. Formally, an independent set in $G(V, E)$ corresponds to the generating points that can simultaneously use the transmission line because they do not result in overload. As each vertex in $H'(V', E')$ corresponds to a load area in *CGS*, therefore maximizing the cardinality of independent set is equivalent to maximizing the number of load areas being served, and the cardinality of the the maximum independent set is equal to $\sum_{s_i \in PC(G)} f(s_i)$. Finding the maximum independent set is not only an NP-hard problem but also NP-hard to approximate within a ratio of $n^{1-\epsilon}$ for any constant $\epsilon > 0$ [43]. Therefore, the *CGS* problem is NP-hard to approximate within a ratio of $n^{1-\epsilon}$ for any constant $\epsilon > 0$. ■

V. PERFORMANCE EVALUATION

We start by highlighting the specific properties that make the proposed solution technique a preferred choice. There are many ways to interpret the preference of solutions, for instance, the notion of feasibility, optimality, time complexity, and restriction on the objective function if any. We discussed

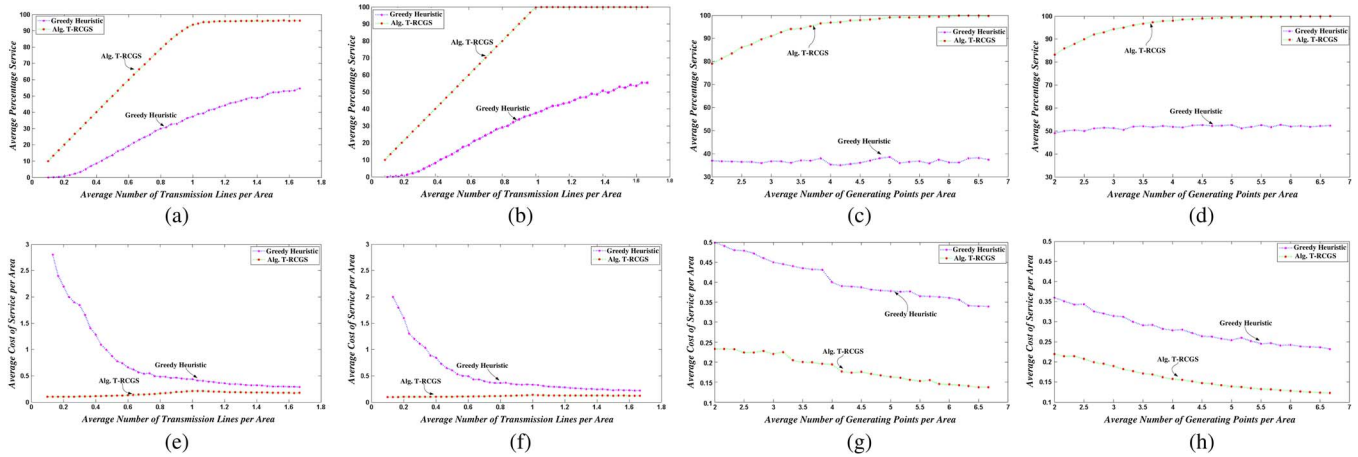


Fig. 5. Average percentage service of power network as a function of average number of transmission lines per area for (a) 2 generating points per area (on average) and (b) 4 generating points per area (on average). Average percentage service of power network as a function of average number of generating points per area for (c) 30 transmission lines per area and (d) 45 transmission lines per area. Average cost of generation per area for power network as a function of average number of transmission lines per area for (e) 4 generators per area (on average) and (f) 8 generating points per area (on average). Average cost of generation per area for power network as a function of average number of generating points per area for (g) 30 transmission lines per area and (h) 45 transmission lines per area.

these important characteristics in Section I-C, which implies that choosing either LP or ILP for solving the *RCGS* problem, proven in this paper to be polynomial-time solvable, results in undesirable solution quality in terms of one or more parameters compared with the presented matroid-theory-based solution. We therefore propose to compare our solution technique to a greedy solution that has desirable solution parameters similar to the proposed matroid-theory-based solution in the subsequent section. However, we point out that the other techniques do have merits for classical problems but show weakness when we have to expand the classical problem to include the emerging constraints.

To evaluate the performance of the proposed scheme on practical power systems, we performed a simulation study using an IEEE 300-bus test system [44]. The IEEE 300-bus test system was developed by the IEEE Test Systems Task Force in 1993; hence, it does not incorporate the smart grid scenario with a large number of generating points. To incorporate the smart grid scenario with a large number of generating points, we extended the number of generating points in this system, which is a known practice (see, for example, [45]). Specifically, we divided the given network into 30 distribution areas. Each area is representing a central load, which is the combined sum of all the individual loads in that area. Note that different load demands, operation of circuit breakers, manual operation of the isolators, maintenance and failure patterns of different devices (transformers, generators, transmission lines, buses, relays, etc.), and availability of renewable sources (such as power production from wind turbines, solar cells, etc.) give rise to a large number of possible realizations of this system. For each experiment, we consider a realization of this system selected from all possible realizations in an unbiased way. Each point in the graphs presented is an average taken over 100 experiments.

To supply an area, at least one generator point must be able to supply its load within the scheduling horizon, and the generating point should also be allocated a transmission line

with sufficient capacity to be able to supply the power to the area. All the generating points that share the same transmission line have conflict with each other; this defines *Transmission Line Constraints*. For the sake of simplicity of explanation, we assume the scheduling horizon of one time unit and generation delays to be zero, so *Delay Constraints* impose the schedule duration of one time unit (i.e., only one allocation at time 0). Regarding *Policy Constraints*, we say an area has been served if at least one generating point is able to supply its load before the elapse of the scheduling horizon, and the generating point is also allocated a transmission line to be able to supply the power to the area.

The objective is to serve maximum number of areas with the minimum total cost of generation, without violating the Transmission Line and Delay Constraints. Throughout the rest of this section, we shall use the word *cost* to specify *cost of generation*. We use average percentage of the areas served and average normalized cost of generation as performance metrics.

For a given power network, the *percentage service* is defined to be the percentage of the areas served without violating the Transmission Line and Delay Constraints. For a given power network, the *cost of the service per area* is defined to be the sum of the generation costs of all the selected generating points divided by the number of areas served. For a given power network, the *number of transmission lines per area* is defined to be the total number of transmission lines in the power network divided by the total number of the areas. Note that as the transmission lines are placed randomly, it might happen that there might be more transmission lines available to serve one area than another area. Hence, the *number of transmission lines per area* shows, on average, how many transmission lines connect an area to the generating point. For a given power network, the *number of generating points per area* is defined to be the total number of generating points divided by the total number of areas.

The first set of experiments shows percentage service and cost of the service per area as a function of the number of

transmission lines per area. We consider two different settings. Fig. 5(a) and (b) shows the average percentage service versus the number of transmission lines per area using both the Algorithm *T-RCGS* and the greedy heuristic with 4 and 8 generating points per area, respectively. Fig. 5(e) and (f) shows the average cost of the service per area for the results in the part Fig. 5(a) and (b), respectively. The results show that with a greater number of generating points per area, Algorithm *T-RCGS* provides a greater percentage service, whereas the greedy heuristic does not show improvement even when the number of generating points per area is increased. The results also point out the large gap between optimal solution (using Algorithm *T-RCGS*) and the greedy solution in terms of both percentage service and cost. An interesting point to note is that when the number of generating points per area is less than a certain threshold, even Algorithm *T-RCGS*, i.e., the optimal solution, cannot provide 100% service. This follows from the fact that generating points and areas are randomly connected to the transmission lines, which might result in Transmission Line Conflicts. Note that Algorithm *T-RCGS* can provide up to 100% service, whereas the greedy heuristic cannot provide more than 60% service.

The second set of experiments shows the relationship of the percentage service and the average number of generating points per area. We consider two different settings for the number of transmission lines per area. Fig. 5(c) and (d) shows the comparison of the average percentage service versus the average number of generating points per area, for power networks with 30 and 45 number of transmission lines per area, respectively. Fig. 5(g) and (h) shows the average cost of the service per area corresponding to the results in Fig. 5(c) and (d), respectively. The results show that Algorithm *T-RCGS* is able to provide 100% service even with a small number of resources (number of transmission lines per area), which highlights the importance of the proposed solution to achieve the best possible service in practical settings where the resources are usually scarce. Furthermore, having more resources (number of transmission lines per area and number of generating points per area) helps in reducing the overall cost of the generation (average cost of the service per area). The results also point out the large gap between using an optimal solution (using Algorithm *T-RCGS*) and the greedy heuristic in terms of both percentage service and cost. Another important observation is that the greedy solution provides poor service (at most 60%) even when optimally 100% coverage could have been achieved.

VI. FINAL REMARKS AND FUTURE WORK

In this paper, we have studied how cyber–physical evolution of power systems can affect the complexity of optimization processes, giving rise to opportunities to consider new formulations of classically NP-hard problems. In particular, we present a general framework for modeling the CGS problem, where the power grid exhibits consumer-driven policy constraints, transmission line constraints, and delay constraints for a given planning horizon to meet the forecasted demand. We have presented polynomial-time distributed solutions to the *I-RCGS* and *T-RCGS* problems and shown the NP-hardness of

the CGS problem. We have demonstrated the gains possible through reformulation of power system problems such that they can be solved in polynomial time in contrast to greedy algorithms. Such gains no doubt will be valuable in an ever-growing competitive marketplace.

Future work will focus on building upon the foundation developed here to include more comprehensive system constraints in the presence of data uncertainty, reliability, and security issues. Furthermore, the impact of storage systems and renewable power generators while incorporating a detailed analysis of the specific parameters associated with the nature of the generating source is the other thrust. Investigation of these important aspects in a comprehensive fashion shall also be the subject of future work. Moreover, we would like to study the developed algorithm under real-time and online scheduling while taking into account the effects of the communication system performances.

APPENDIX A PROOF OF LEMMA 5

Proof: Consider a set G_f consisting of f generating points, such that $G_f \in \mathcal{L}_1 \cap \mathcal{L}_2$. We show that G_f does not violate the transmission line constraints and provides service to exactly f load areas $s_i, \forall i$. First, note that since $G_f \in \mathcal{L}_2$, the f generating points do not have transmission line conflicts. Second, note that as $G_f \in \mathcal{L}_1$, the associated column vectors of these f generating points are linearly independent, implying that they belong to f distinct policy sets. Thus, G_f provides service to at least f different load areas since each policy set represents the preferences of a distinct load area.

Therefore, a set $G_f \in \mathcal{L}_1 \cap \mathcal{L}_2$, such that G_f has maximum cardinality among all sets belonging to $\mathcal{L}_1 \cap \mathcal{L}_2$, shall provide service to the maximum possible number of load areas within constraints. Furthermore, if no other set of maximum cardinality belonging to $\mathcal{L}_1 \cap \mathcal{L}_2$ has cost lesser than G_f , then G_f provides an optimal solution to the *I-RCGS* problem. It has been shown that both TLM and SM are matroids; thus, a set belonging to $\mathcal{L}_1 \cap \mathcal{L}_2$ and having least cost and maximum cardinality among all the sets in $\mathcal{L}_1 \cap \mathcal{L}_2$ can be found in polynomial time using the weighted matroid intersection algorithm [39], [40]. ■

APPENDIX B PROOF OF THEOREM 6

Sketch of proof: The correctness of distributed Algorithm *I-RCGS* follows from Lemma 5 and correctness of the weighted matroid intersection algorithm by Edmonds [39]. Specifically, the matroid intersection algorithm in [39] consist of four steps with corresponding counterparts in Algorithm *I-RCGS*.

- 1) In [39], Edmonds constructs a directed graph where edges represent valid replacements of a set of generating points selected in G_{on} from the set of generating points not yet selected. This corresponds to steps 10, 11, 16, and 17 in Algorithm *I-RCGS*; in our algorithm, the entries of the routing table \vec{E} and \vec{E} indicate the valid replacement.
- 2) Similar to [39], Algorithm *I-RCGS* finds vector \mathbf{G}_1 such that each generating point in \mathbf{G}_1 corresponds to the

valid addition of generating points to the already-selected generating points (G_{on}) with respect to \mathcal{L}_1 of the SM matroid. Similarly, we find vector G_2 that corresponds to the generating points that are a valid addition with respect to \mathcal{L}_2 of the TLM matroid. Steps 19 and 20 do the same.

- 3) Analogous to [39], Algorithm I -RCGS finds a least cost path corresponding to *Phase 2* of our algorithm.
- 4) Step 32 of Algorithm I -RCGS computes an updated vector G_{on} by necessary replacements and additions similar to [39].

Therefore, the correctness of Algorithm I -RCGS follows directly from the correctness of Edmonds' matroid intersection algorithm.

REFERENCES

- [1] P. Derler, E. Lee, and A. Vincentelli, "Modeling cyber–physical systems," *Proc. IEEE*, vol. 100, no. 1, pp. 13–28, Jan. 2012.
- [2] J. Wei, D. Kundur, T. Zourntos, and K. Butler-Purry, "A flocking-based dynamical systems paradigm for smart power system analysis," in *Proc. IEEE PES GM*, 2012, pp. 1–8.
- [3] D. Kundur, X. Feng, S. Liu, T. Zourntos, and K. Butler-Purry, "Towards a framework for cyber attack impact analysis of the electric smart grid," in *Proc. IEEE SmartGridComm*, Oct. 2010, pp. 244–249.
- [4] J. Wei and D. Kundur, "A flocking-based model for dos-resilient communication routing in smart grid," in *Proc. IEEE GLOBECOM*, 2012, pp. 3519–3524.
- [5] S. Orero and M. Irving, "A genetic algorithm for generator scheduling in power systems," *Int. J. Elect. Power Energy Syst.*, vol. 18, no. 1, pp. 19–26, Jan. 1996.
- [6] X. Guan, P. Luh, and B. Prasanna, "Power system scheduling with fuzzy reserve requirements," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 864–869, May 1996.
- [7] C. L. Tseng, "On power system generation unit commitment problems," Ph.D. dissertation, Univ. California, Berkeley, CA, USA, 1996.
- [8] State Energy Conservation Office. [Online]. Available: <http://www.seco.cpa.state.tx.us>
- [9] R. Karp, *Reducibility among Combinatorial Problems*. New York, NY, USA: Springer-Verlag, 1972.
- [10] R. Kannan and C. Monma, "On the computational complexity of integer programming problems," in *Optimization and Operations Research*, vol. 157, *Lecture Notes in Economics and Mathematical Systems*, R. Henn, B. Korte, and W. Oettli, Eds. Berlin, Germany: Springer-Verlag, 1978, pp. 161–172.
- [11] S. S. Skiena, *The Algorithm Design Manual*, 2nd ed. New York, NY, USA: Springer-Verlag, 2008.
- [12] C. M. Papadimitriou, *Computational Complexity*. Reading, MA, USA: Addison-Wesley, 1994.
- [13] S. Bu, F. R. Yu, P. X. Liu, and P. Zhang, "Distributed scheduling in smart grid communications with dynamic power demands and intermittent renewable energy resources," in *Proc. IEEE ICC Workshops*, 2011, pp. 1–5.
- [14] C. Gong, X. Wang, W. Xu, and A. Tajer, "Distributed real-time energy scheduling in smart grid: Stochastic model and fast optimization," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1476–1489, Sep. 2013.
- [15] R. E. Bixby, W. Cook, A. Cox, and E. K. Lee, "Parallel mixed integer programming," Rice Univ. Center Research Parallel Comput. Res. Monogr., Houston, TX, USA, CRPC-TR95554, 1995.
- [16] J. T. Linderoth and M. W. P. Savelsbergh, "A computational study of search strategies for mixed integer programming," *INFORMS J. Comput.*, vol. 11, no. 2, pp. 173–187, Feb. 1997.
- [17] F. Pilo, G. Pisano, and G. Soma, "Optimal coordination of energy resources with a two-stage online active management," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4526–4537, Oct. 2011.
- [18] A. Timbus, M. Larsson, and C. Yuen, "Active management of distributed energy resources using standardized communications and modern information technologies," *IEEE Trans. Ind. Electron.*, vol. 56, no. 10, pp. 4029–4037, Oct. 2009.
- [19] P. C. Loh, L. Zhang, and F. Gao, "Compact integrated energy systems for distributed generation," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1492–1502, Apr. 2013.
- [20] E. Poursmaeil, C. Miguel-Espinar, M. Massot-Campos, D. Montesinos-Miracle, and O. Gomis-Bellmunt, "A control technique for integration of dg units to the electrical networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 7, pp. 2881–2893, Jul. 2013.
- [21] A. Pantoja and N. Quijano, "A population dynamics approach for the dispatch of distributed generators," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4559–4567, Oct. 2011.
- [22] D. Q. Hung and N. Mithulananthan, "Multiple distributed generator placement in primary distribution networks for loss reduction," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1700–1708, Apr. 2013.
- [23] S. Paudyal, C. Canizares, and K. Bhattacharya, "Optimal operation of distribution feeders in smart grids," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4495–4503, Oct. 2011.
- [24] S. Chen and H. Gooi, "Jump and shift method for multi-objective optimization," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4538–4548, Oct. 2011.
- [25] A. Bhardwaj, N. Tung, and V. Kamboj, "Unit commitment in power system: A review," *Int. J. Elect. Power Eng.*, vol. 6, no. 1, pp. 51–57, 2012.
- [26] H. Happ, "Optimal power dispatch: A comprehensive survey," *IEEE Trans. Power App. Syst.*, vol. PAS-96, no. 3, pp. 841–854, May 1977.
- [27] D. Kirshen and G. Strbac, *Fundamentals of Power System Economics*. Hoboken, NJ, USA: Wiley, 2004.
- [28] J. López, J. Meza, I. Guillén, and R. Gómez, "A heuristic algorithm to solve the unit commitment problem for real-life large-scale power systems," *Int. J. Elect. Power Energy Syst.*, vol. 49, pp. 287–295, Jul. 2013.
- [29] S. Ng and J. Zhong, "Security-constrained dispatch with controllable loads for integrating stochastic wind energy," in *Proc. IEEE PES ISGT*, Oct. 2012, pp. 1–8.
- [30] K. Zare and S. Hashemi, "A solution to transmission-constrained unit commitment using hunting search algorithm," in *Proc. IEEEIC*, May 2012, pp. 941–946.
- [31] R. Bacher, "Power system models, objectives and constraints in optimal power flow calculations," in *Optimization in Planning and Operation of Electric Power Systems*. Berlin, Germany: Springer-Verlag, 1993, pp. 217–263.
- [32] F. Capitanescu, W. Rosehart, and L. Wehenkel, "Optimal power flow computations with constraints limiting the number of control actions," in *Proc. IEEE PowerTech*, 2009, pp. 1–8.
- [33] E. N. Azadani, S. H. Hosseini, P. H. Divshali, and B. Vahidi, "Stability constrained optimal power flow in deregulated power systems," *Elect. Power Compon. Syst.*, vol. 39, no. 8, pp. 713–732, May 2011.
- [34] J. J. Shaw, "A direct method for security-constrained unit commitment," *IEEE Trans. Power Syst.*, vol. 10, no. 3, pp. 1329–1342, Aug. 1995.
- [35] A. Chaouachi, R. M. Kamel, R. Andoulsi, and K. Nagasaka, "Multiobjective intelligent energy management for a microgrid," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1688–1699, Apr. 2013.
- [36] J. G. Oxley, *Matroid Theory*. Oxford, U.K.: Oxford Univ. Press, 2006.
- [37] C. Tseng *et al.*, "A transmission-constrained unit commitment method," in *Proc. Int. Conf. Syst. Sci.*, 1998, vol. 3, pp. 71–80.
- [38] A. Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*. Berlin, Germany: Springer Verlag, 2003.
- [39] J. Edmonds, "Matroid intersection," *Ann. Discr. Math.*, vol. 4, pp. 39–49, 1979.
- [40] L. Lawler, "Matroid intersection algorithms," *Math. Program.*, vol. 9, no. 1, pp. 31–56, 1975.
- [41] S. Haldar, "An 'all pairs paths' distributed algorithm using $2n^2$ messages," in *Graph-Theoretic Concepts in Computer Science*. New York, NY, USA: Springer-Verlag, 1994, pp. 350–363.
- [42] H. Hamacher, "A time expanded matroid algorithm for finding optimal dynamic matroid intersections," *Math. Methods Oper. Res.*, vol. 29, no. 5, pp. 203–215, Nov. 1985.
- [43] D. Zuckerman, "Linear degree extractors and the inapproximability of max clique and chromatic number," in *Proc. ACM STOC*, 2006, pp. 681–690.
- [44] IEEE Power Systems Test Case Archive. [Online]. Available: <http://www.ee.washington.edu/research/pstca/index.html>
- [45] A. Diniz, "Test cases for unit commitment and hydrothermal scheduling problems," in *Proc. IEEE PES GM*, 2010, pp. 1–8.

Zakia Asad (S'13), photograph and biography not available at the time of publication.

Mohammad Assad Rehman Chaudhry (S'97–M'10), photograph and biography not available at the time of publication.

Deepa Kundur (S'93–M'99–SM'03), photograph and biography not available at the time of publication.