

Scheduling Flexible Nonpreemptive Loads in Smart-Grid Networks

Nathan Dahlin¹, *Student Member, IEEE*, and Rahul Jain¹, *Senior Member, IEEE*

Abstract—A market consisting of a generator with thermal and renewable generation capability, a set of *nonpreemptive* loads (i.e., loads which cannot be interrupted once started), and an independent system operator (ISO) is considered. Loads are characterized by durations, power demand rates, and utility for receiving service, as well as disutility functions giving preferences for time slots in which service is preferred. Given this information, along with the generator's thermal generation cost function and forecast renewable generation, the social planner solves a mixed-integer program to determine a load activation schedule which maximizes social welfare. Assuming price-taking behavior, we develop a competitive equilibrium concept based on a relaxed version of the social planner's problem, which includes prices for consumption and incentives for flexibility, and allows for probabilistic allocation of power to loads. Considering each load as representative of a population of identical loads with scaled characteristics, we demonstrate that the relaxed social planner's problem gives an exact solution to the original mixed integer problem in the large population limit, and give a market mechanism for implementing the competitive equilibrium. Finally, we evaluate via case study the benefit of incorporating load flexibility information into power consumption and generation scheduling in terms of proportion of loads served and overall social welfare.

Index Terms—Electric vehicles, power system planning, power systems economics.

I. INTRODUCTION

OVER THE roughly century-long history of the electrical power grid, the situation facing both grid managers and end users has remained largely the same: electricity available on demand. In the case of the latter, operation of lightbulbs, television sets, and other appliances has been just the flip of a switch away, while for the former, the set of available controls and

Manuscript received August 10, 2021; accepted November 29, 2021. Date of publication January 6, 2022; date of current version May 26, 2022. This work was supported by NSF Awards under Grant ECCS-1611574 and Grant ECCS-1810447. Recommended by Associate Editor Alejandro D. Dominguez-Garcia. (Corresponding author: Nathan Dahlin.)

Nathan Dahlin was with the Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089 USA. He is now with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: dahlin@usc.edu).

Rahul Jain is with the Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: rahul.jain@usc.edu).

Digital Object Identifier 10.1109/TCNS.2022.3141017

actions was oversupply, i.e., which generators to activate—How much to generate and when [21]? Managers have consistently succeeded in providing an adequate supply to meet the demand of end users from second to second largely due to the fact that over the past century, demand forecasting has reached day-ahead accuracy within 5% [14].

Recently, circumstances have changed on both the supply and demand sides of the grid. Increased adoption of renewables means that the available power supply is becoming less controllable. Thus, even in the presence of relatively predictable aggregate load, forecasting errors in *excess load* can be significant. Meanwhile, the rise of networked appliances in homes and buildings is now facilitating synchronization and coordination of consumption to the extent that the demand-side flexibility stands to become one of the most important assets available to grid operators [20]. Water heaters and electric vehicles (EVs) typify loads characterized by such flexibility. A recent report from the Brattle Group estimates that load flexibility could be expanded to satisfy nearly 20% of the US peak demand, and avoid nearly \$18 billion in annual generation capacity, energy, transmission, and ancillary service costs [13].

Currently, aggregate flexibility is leveraged through *demand response* programs. Typically, these programs are used to reduce peaks in demand, either by indirect load control via real-time pricing or direct control, where utilities have the ability to turn devices ON or OFF. Moving forward, much of the additional benefit is expected to come from expanding the use of demand response to applications such as load shifting and building, e.g., to track a time-varying supply of renewable energy, and services such as frequency regulation and voltage control [13].

This work considers a population of *nonpreemptive* loads, i.e., loads which must be served continuously for a predetermined amount of time without interruption once service has started. Examples of such loads are household appliances like dishwashers and EV charging with tight deadlines [10]. Users report their level of discomfort for being served at each time slot of a finite time horizon. The social planner tasked with serving these loads has access to a thermal generator with convex generation cost, as well as a renewable generator with zero marginal cost. Given the users' preferences, thermal generator's cost function, and knowledge of the renewable generator's output, the scheduler determines an efficient schedule for cost minimization. We seek to answer the following questions: How can these flexible loads be scheduled over the available time slots? Once a schedule has been determined, how should users be compensated for their

flexibility? What is the “price of inflexibility” in this setting? In particular, the problem of monetizing flexibility has proven challenging thus far due to a lack of suitable optimization formulations [20].

A. Related Work

While the problem of scheduling processor time for service of both pre-emptive and nonpreemptive tasks has long been studied in computer science literature [7], load scheduling in the context of demand response in energy systems has received considerable attention over the past several years. Model predictive control (MPC) [8], successive binary optimization [25], and greedy algorithms [19] have been applied to the particular objective of tracking a target aggregate load profile at minimum cost, without considering pricing.

Beyond costs associated directly with generation, several works incorporate exogenous energy pricing schemes as parameters of their respective formulations. Fixed uniform rates [24], tiered rates based upon given affine per unit pricing functions [9], and peak/off-peak pricing schemes [11] modeling programs implemented by utilities have been considered in concert with earliest deadline first, least laxity first, and MPC scheduling policies.

Of the works that derive flexibility pricing schemes endogenously alongside optimal schedules, relatively few consider nonpreemptive loads. This is primarily due to the binary start time decision variables necessary when introducing nonpreemptive loads, which precludes direct use of traditional convex optimization or marginal pricing-based schemes [20]. Without explicit inclusion of nonpreemptive consumption profiles, the scheduling and pricing of a continuum of deadline differentiated loads is studied in [2], wherein the longer a consumer is willing to defer, the lower their energy price.

Targeting nonpreemptive loads with hard constraints on acceptable service windows, both [4] and [18] adopt a bilevel programming approach in which an energy provider optimizes a pricing schedule in the upper level, and consumers react by scheduling their loads in order to minimize costs and discomfort. Genetic algorithm [18] and evolutionary and particle-swarm [4] based scheduling heuristics are used to handle the nonpreemptive-related integer constraints.

More recently, Qin *et al.* [20] detail a power exchange platform allowing for flexible suppliers, as well as consumers with nonpreemptive consumption patterns. A fluid relaxation on demand profile shapes along with a projection method for deriving a feasible schedule is given, and the resulting solution is shown to be asymptotically optimal in the infinite load limit. Marginal pricing, given a schedule of the flexible loads, is shown to be inadmissible with respect to incoming offers.

Closest to this work, in [7], a setting similar to the one presented here in continuous time is examined. Prices for load consumption and inflexibility are derived as dual variables to the scheduler’s convex optimization problem, and a competitive equilibrium with respect to reported loads reported consumption level and duration is studied. Under time discretization,

approximately optimal scheduling and pricing heuristics are developed. In contrast to this work, user disutility is not modeled, and flexibility incentives do not arise from the optimization formulation. Further, the optimality of the associated heuristics cannot be proven.

B. Statement of Contributions

In this article, we propose a tractable optimization formulation for scheduling and pricing nonpreemptive load service. Inclusion of such loads necessitates specification of constraints and variables capturing their noninterruptible nature, and in general results in NP-hard optimization problems [20]. In summary, the key steps and contributions of this article are as follows.

- 1) We propose a novel specification of load flexibility and decentralized optimization formulation for the scheduling of nonpreemptive loads.
- 2) We formulate a corresponding centralized welfare maximization problem, and prove the existence of a competitive equilibrium in the relaxed version of this setting with finitely many loads, i.e., we show that there exist prices for per unit energy consumption and inflexibility such that the thermal generator produces efficient levels at each time step, and a social planner schedules loads such that demand equals supply while respecting the loads’ flexibility preferences.
- 3) We prove that the competitive equilibrium determined for the finite population setting is also a competitive equilibrium in the original mixed binary problem, when each load is interpreted as representing an infinite population of loads with appropriately scaled demand. Thus, the prices derived via convex relaxation are suitable for use in the binary-constrained setting. Such a result is currently absent in the related literature.
- 4) We specify a market mechanism for implementing the competitive equilibrium.
- 5) We present a case study demonstrating the utility of our formulation, based on real-world EV charging data drawn from the Adaptive Charging Network (ACN) project [15].

II. PROBLEM FORMULATION

The market consists of M nonpreemptive loads (or consumers) and a single thermal generator. Additionally, an ISO (independent system operator) ensures safe grid operation. Let $\mathcal{T} = [1, \dots, T]$ denote the discrete time horizon over which loads are scheduled. For simplicity, we assume a single-bus network model. We assume throughout that all entities are *price taking*, i.e., their actions do not affect market prices.

A. Market Entity Problems

1) Loads: Each load i is characterized by a tuple $(\tau_i, l_i, \bar{U}_i, u_i^{dS}, u_i^{dE})$, where τ_i gives the duration in time slots, l_i gives the consumption level, and u_i^{dS} and u_i^{dE} give the disutility functions of consumer i due to service starting prior to or after a desired service window, respectively. Consumer i demands l_i MW of electricity for τ_i consecutive time slots, and derives

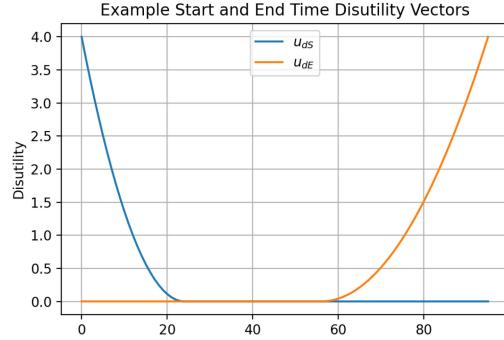


Fig. 1. Example disutility vectors.

utility \bar{U}_i as their load is fulfilled. Fig. 1 plots an example pair of disutility functions vectors. If consumer i is allowed to be served in time slot t , it suffers disutility $u_{it}^{dS} + u_{it}^{dE} \geq 0$. For example, an EV commuter may prefer that his or her vehicle be charged for a 5-h period at 5 kW/h between 9 AM and 5 PM—while he or she works—rather than arrive early or stay late in order to charge the vehicle. Thus, the consumer’s overall utility is a function of the flexibility that it allows for in the scheduling of its load.

Denote $x_i^C := (x_{i1}^C, \dots, x_{iT}^C) \in \{0, 1\}^T$. We will similarly define vector- and matrix-valued quantities throughout. Given flexibility incentives $p_i^S \in \mathbb{R}_+^T$ and $p_i^E \in \mathbb{R}_+^T$, consumer i solves the following optimization problem:

$$(\text{CONS}_i) \quad \min_{\substack{x_i^C \in \{0, 1\}^T, \\ y_i^C \geq 0, z_i^C \geq 0}} \sum_t p_{it}^{\text{con}} x_{it}^C - \bar{U}_i \sum_{t=1}^{T-\tau_i+1} x_{it}^C + \sum_t ((1 - y_{it}^C)(u_{it}^{dS} - p_{it}^S) + (1 - z_{it}^C)(u_{it}^{dE} - p_{it}^E)) \quad (1)$$

$$\text{s.t.} \quad \sum_{s=1}^t \sum_{r=\max\{1, s-\tau_i+1\}}^s x_{ir}^C \leq \tau_i(1 - y_{it}^C) \quad \forall t \quad (2)$$

$$\sum_{s=t}^T \sum_{r=\max\{1, s-\tau_i+1\}}^s x_{ir}^C \leq \tau_i(1 - z_{it}^C) \quad \forall t. \quad (3)$$

In the context of EV charging, if $x_{it}^C = 1$, then consumer i chooses to begin charging his or her vehicle at time slot t and pays activation price p_{it}^{con} . The inner sums on the left-hand side in constraints (2) and (3) give the charging/idle status of load i at each time slot s . That is, the sums will be equal to 1 if consumer i ’s vehicle started charging in any time slot $\{\max\{1, s-\tau_i+1\}, \dots, s\}$ and 0 otherwise. The term $(1 - y_{it}^C) > 0$ when consumer i ’s EV has started charging prior to or at time slot t . In such a case, consumer i incurs disutility $(1 - y_{it}^C)u_{it}^{dS} \geq 0$ for having started by time t , but is compensated at early start rate p_{it}^S . Similarly, $(1 - z_{it}^C) > 0$ indicates that consumer i ’s EV will be charging at or after time slot t , with u_{it}^{dE} and late charge ending rate p_{it}^E analogous to u_{it}^{dS} and p_{it}^S .

2) Thermal Generator: The generator is characterized by its thermal generation cost function $c(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, which is

assumed to be strictly convex, increasing and twice differentiable on \mathbb{R}_+ . In addition to the generator’s thermal plant, we assume that it also owns a renewable generator which produces energy at zero marginal cost. The output of the renewable generator, $g : \mathcal{T} \rightarrow (0, \infty)$, is assumed to be known to all market participants at time $t = 0$. Given prices $p_t^{\text{gen}} \in \mathbb{R}_+^T$, the generator chooses generation levels $q_t^G \in \mathbb{R}_+^T$ to solve the following profit maximization problem:

$$(\text{GEN}) \quad \max_{q_t^G \geq 0} \sum_t (p_t^{\text{gen}}(q_t^G + g_t) - c(q_t^G)).$$

3) ISO: Finally, the ISO collects all load profiles and determines the set of admissible load and generation schedules by solving

$$(\text{ISO}) \quad \min_{\substack{q^I \geq 0, x^I \\ y^C \geq 0, z^C \geq 0}} \sum_t p_t^{\text{gen}} \left(q_t^I + g_t - \sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}} x_{is}^I \right) \\ \text{s.t.} \quad \sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}} x_{is}^I - g_t \leq q_t^I \quad \forall t, x^I \in \{0, 1\}^{M \times T}$$

where $q^I \in \mathbb{R}_+^T$. Note that the ISO incurs positive cost at any time slot t , where thermal generation q_t^I exceeds residual demand (aggregate demand less renewable generation) and, thus, will find those schedules which balance thermal generation and residual demand optimal.

B. Centralized Social Planner’s Problem

In order to study the welfare properties of the competitive equilibrium given later, we introduce a social planning problem. The social planner is concerned with maximizing the combined welfare of all market participants, while ensuring safe operation of the power grid. Specifically, the social planner collects the profiles of each load i , and schedules them so that each is served without interruption for the entire duration. In practice, either the ISO or equivalent market participant, or a government organization often assumes responsibility for these tasks [6]. Here, we introduce the social planner as a distinct entity for clarity as we investigate properties of our market formulation. Let $\hat{x}_{it} \in \{0, 1\}$ denote the social planner’s decision as to whether load i will begin service in time slot t , where $\hat{x}_{it} = 1$ denotes that load i will start at time slot t . A schedule is then defined as $\hat{x} \in \{0, 1\}^{M \times T}$. The social planner selects a schedule, auxiliary load status variables \hat{y} and \hat{z} , and corresponding generation levels $\hat{q} := (\hat{q}_1, \dots, \hat{q}_T)$ in order to solve the following problem:

$$(\text{SPP}) \quad \min_{\substack{\hat{q} \geq 0, \hat{x} \in \{0, 1\}^{M \times T} \\ \hat{y} \geq 0, \hat{z} \geq 0}} \sum_t c(\hat{q}_t) + \sum_i \sum_t u_{it}^{dS} (1 - \hat{y}_{it}) + \sum_i \sum_t u_{it}^{dE} (1 - \hat{z}_{it}) - \sum_i \bar{U}_i \sum_{t=1}^{T-\tau_i+1} \hat{x}_{it} \quad (4)$$

$$\text{s.t.} \quad \sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}} \hat{x}_{is} - g_t \leq \hat{q}_t \quad \forall t \quad (5)$$

$$\sum_{s=1}^t \sum_{r=\max\{1, s-\tau_i+1\}}^s \hat{x}_{ir} \leq \tau_i(1 - \hat{y}_{it}) \quad \forall i, t \quad (6)$$

$$\sum_{s=t}^T \sum_{r=\max\{1, s-\tau_i+1\}}^s \hat{x}_{ir} \leq \tau_i(1 - \hat{z}_{it}) \quad \forall i, t. \quad (7)$$

III. CONVEX RELAXATION AND PRICING

In order to develop prices for electricity consumption and load inflexibility, as well as a competitive equilibrium concept, we relax the binary constraints on \hat{x} , and consider the following problem:

$$\begin{aligned} (\text{SPP-R}) \quad & \min_{\hat{q}, \hat{x}, \hat{y}, \hat{z} \geq 0} \sum_t c(\hat{q}_t) + \sum_i \sum_t u_{it}^{dS} (1 - \hat{y}_{it}) \\ & + \sum_i \sum_t u_{it}^{dE} (1 - \hat{z}_{it}) - \sum_i \bar{U}_i \sum_{t=1}^{T-\tau_i+1} \hat{x}_{it} \end{aligned} \quad (8)$$

$$\text{s.t. } \hat{\lambda}_t : \sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}}^t \hat{x}_{is} - g_t \leq \hat{q}_t \quad \forall t \quad (9)$$

$$\hat{\nu}_{it}^S : \sum_{s=1}^t \sum_{r=\max\{1, s-\tau_i+1\}}^s \hat{x}_{ir} \leq \tau_i(1 - \hat{y}_{it}) \quad \forall i, t \quad (10)$$

$$\hat{\nu}_{it}^E : \sum_{s=t}^T \sum_{r=\max\{1, s-\tau_i+1\}}^s \hat{x}_{ir} \leq \tau_i(1 - \hat{z}_{it}) \quad \forall i, t \quad (11)$$

where $\hat{\lambda}_t$, $\hat{\nu}_{it}^S$, and $\hat{\nu}_{it}^E$ denote the dual variables corresponding to constraints (9)–(11). It can be shown that constraints (10) and (11) ensure that all entries of \hat{x} , \hat{y} , and \hat{z} are less than 1, and

$$\sum_{t=1}^{T-\tau_i+1} \hat{x}_{it} \leq 1 \quad \forall i, t. \quad (12)$$

Under the relaxation, since in addition to (12), each schedule decision variable \hat{x}_{it} satisfies $0 \leq \hat{x}_{it} \leq 1$, \hat{x} may be interpreted as a matrix specifying the *probability* that a given load of type i will be scheduled at time slot t for all i and t . That is, for each i , the planner will choose $\hat{x}_{i \cdot} \in \mathbb{R}^T$ equal to e_t , the t th standard basis vector, with probability \hat{x}_{it} . Therefore, \hat{x} in (SPP-R) gives a *probabilistic* schedule for the loads and if, for a given i , (12) holds with equality, then load i is certain to be activated at some time slot $t \in \mathcal{T}$. Otherwise, the load only has a chance of ever being activated. Fixing a matrix of probabilities \hat{x} , $(1 - \hat{y}_{it})$ and $(1 - \hat{z}_{it})$ give probabilities that load i has been activated up to time t , and will be active from time slot t onward, respectively, for all i and t . Thus, the (SPP-R) objective may be viewed as the expectation of overall social welfare, and the constraints as in expectation. This interpretation is key to the competitive equilibrium definition and properties we detail in later sections.

Note that due to the non-negativity of u_{it}^{dS} and u_{it}^{dE} for all i, t , for any fixed \hat{x} , it is always optimal to choose each entry of matrices $\{1 - \hat{y}\}$ and $\{1 - \hat{z}\}$ as small as possible, where $\mathbf{1}$ denotes the matrix of size $M \times T$ with each entry being equal to 1. Therefore, constraints (10) and (11) may be replaced with

equalities, and matrices \hat{y} and \hat{z} are completely determined given a particular \hat{x} .

Having relaxed the binary constraints on matrices \hat{x} , \hat{y} , and \hat{z} , we may employ Lagrangian analysis in order to arrive at a solution to (SPP-R). Let

$$p_{it}^{\hat{\lambda}} = l_i \sum_{s=t}^{\min\{T, t+\tau_i-1\}} \hat{\lambda}_s \quad (13)$$

$$\begin{aligned} p_{it}^{\hat{\nu}} = & \sum_{s=t}^T \hat{\nu}_{is}^S \min\{s - t + 1, \tau_i\} \\ & + \sum_{s=1}^{\min\{T, t+\tau_i-1\}} \hat{\nu}_{is}^E \min\{T - t + 1, \tau_i, \tau_i - (s - t)\}. \end{aligned} \quad (14)$$

See [5] for the derivation of $p_{it}^{\hat{\lambda}}$ and $p_{it}^{\hat{\nu}}$. In addition to feasibility, the KKT optimality conditions for (SPP-R) are

$$c'(\hat{q}_t^*) - \hat{\lambda}_t^* \geq 0, \hat{q}_t^* (c'(\hat{q}_t^*) - \hat{\lambda}_t^*) = 0 \quad \forall t \quad (15)$$

for all $i, t \leq T - \tau_i + 1$

$$p_{it}^{\hat{\lambda}^*} + p_{it}^{\hat{\nu}^*} - \bar{U}_i \geq 0, \hat{x}_{it}^* (p_{it}^{\hat{\lambda}^*} + p_{it}^{\hat{\nu}^*} - \bar{U}_i) = 0 \quad (16)$$

for all $i, t > T - \tau_i + 1$

$$p_{it}^{\hat{\lambda}^*} + p_{it}^{\hat{\nu}^*} \geq 0, \hat{x}_{it}^* (p_{it}^{\hat{\lambda}^*} + p_{it}^{\hat{\nu}^*}) = 0 \quad (17)$$

and for all i and t

$$\tau_i \hat{\nu}_{it}^{S*} - u_{it}^{dS} \geq 0, \hat{y}_{it}^* (\tau_i \hat{\nu}_{it}^{S*} - u_{it}^{dS}) = 0 \quad \forall i, t \quad (18)$$

$$\tau_i \hat{\nu}_{it}^{E*} - u_{it}^{dE} \geq 0, \hat{z}_{it}^* (\tau_i \hat{\nu}_{it}^{E*} - u_{it}^{dE}) = 0 \quad \forall i, t \quad (19)$$

$$\hat{\lambda}_t^* \left(\sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}}^t \hat{x}_{is}^* - g_t - \hat{q}_t^* \right) = 0 \quad \forall t \quad (20)$$

$$\hat{\lambda}_t^* \geq 0 \quad \forall i, t. \quad (21)$$

Note that due to condition (16), load i may only be activated with nonzero probability during time slots when the sum $p_{it}^{\hat{\lambda}^*} + p_{it}^{\hat{\nu}^*}$ is equal to the constant marginal utility term \bar{U}_i . Additionally, condition (15) implies that for time slots in which it is optimal to produce a positive quantity of electricity, we have that $\hat{\lambda}_t^* = c'(\hat{q}_t^*) > 0$, the marginal cost of production. In turn, condition (20) implies that generation in such time slots will be equal to demand less forecast renewable generation.

In view of our interpretation of the (SPP-R) objective as the expected value of social welfare and the constraints as being met in expectation, the solution to (SPP-R) yields a set of admissible load activation schedules which may be randomly selected by the social planner in a single shot of the original, binary-constrained problem. We will further examine this correspondence in later sections. Fixing such an activation schedule and taking into account renewable generation output g_t , the optimal generation schedule follows from constraints (5) and (9), and condition (19).

A. Convex Relaxed Entity Problems

Inspection of the (SPP-R) Lagrangian suggests the following decomposition of the relaxed social planner's problem into relaxed versions of the individual entity problems presented above. See [5] for the optimality conditions corresponding to each of these problems.

1) Loads: Starting with the consumer problems, we have

$$(CON\text{-}R_i) \min_{\substack{x_i^C, y_i^C \\ z_i^C \geq 0}} \sum_t p_{it}^{\text{con}} x_{it}^C - \bar{U}_i \sum_{t=1}^{T-\tau_i+1} x_{it}^C + \sum_t ((u_{it}^{dS} - p_{it}^S)(1 - y_{it}^C) + (u_{it}^{dE} - p_{it}^E)(1 - z_{it}^C)) \quad (22)$$

$$\text{s.t. } \theta_{it}^S : \sum_{s=1}^t \sum_{r=\max\{1, s-\tau_i+1\}}^s x_{ir}^C = \tau_i(1 - y_{it}^C) \quad \forall t \quad (23)$$

$$\theta_{it}^E : \sum_{s=t}^T \sum_{r=\max\{1, s-\tau_i+1\}}^s x_{ir}^C = \tau_i(1 - z_{it}^C) \quad \forall t. \quad (24)$$

Again, under the relaxation of the binary constraints on x_i^C , we may interpret the consumer's problem as selecting probabilities of activation for each time slot t , in the interest of maximizing their expected net utility (here written in minimization form).

2) Generator: The generator's problem remains the same

$$(GEN\text{-}R) \max_{q^G \geq 0} \sum_t (p_t^{\text{gen}}(q_t^G + g_t) - c(q_t^G)). \quad (25)$$

3) ISO: Finally, the relaxed ISO problem is given by

$$(ISO\text{-}R) \min_{\substack{q^I \geq 0 \\ x^I \geq 0}} \sum_t p_t^{\text{gen}} \left(q_t^I + g_t - \sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}}^t x_{is}^I \right) \text{s.t. } \alpha_t : \sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}}^t x_{is}^I - g_t \leq q_t^I \quad \forall t.$$

IV. COMPETITIVE EQUILIBRIUM AND THEOREMS OF WELFARE ECONOMICS

The competitive or Walrasian equilibrium is a standard reference point in economic analysis for assessing market outcomes. A competitive equilibrium is specified by an allocation of goods and prices, with the defining characteristic that taking the equilibrium prices as given, every market participant finds it optimal to select the corresponding equilibrium allocation [16]. At equilibrium prices, the quantity of goods demanded by consumers is equal to that produced by suppliers., i.e., the market clears. Therefore, equilibrium prices provide a coordinating signal for *decentralized* market operation.

Assuming that a competitive equilibrium exists for a particular market setting, it is natural to compare the equilibrium allocation to allocations which directly maximize the aggregate welfare of all market participants. The latter allocations are called *efficient*, and here correspond to solutions to (SPP-R).

We now give the competitive equilibrium definition for our setting, and explore existence, as well as welfare properties of

such an equilibrium. As related above, competitive equilibria are typically specified in two-sided settings involving consumers and producers maximizing their individual well-being and profits, respectively. Similar to the analysis found in [22] and [26], we augment the standard definition to include a nonprofit entity, i.e., the ISO.

Definition 1: (Competitive Equilibrium). A tuple $(\bar{q}^*, \bar{x}^*, \bar{y}^*, \bar{z}^*, \bar{p}^{\text{con}*}, \bar{p}^S^*, \bar{p}^E^*, \bar{p}^{\text{gen}*})$ with $\bar{p}^{\text{gen}*} \geq 0$ is said to be a competitive equilibrium if, given $(\bar{p}_i^{\text{con}*}, \bar{p}_i^S^*, \bar{p}_i^E^*)$, $(\bar{x}_i^*, \bar{y}_i^*, \bar{z}_i^*)$ solves $(CON\text{-}R_i)$ for each i , \bar{q}^* solves $(GEN\text{-}R)$, given $\bar{p}^{\text{gen}*}$, \bar{q}^* solves $(GEN\text{-}R)$, and given $\bar{p}^{\text{gen}*}$, (\bar{q}^*, \bar{x}^*) solves $(ISO\text{-}R)$.

As noted in the previous section, since solutions to $(CON\text{-}R_i)$ will, in general, give values of $x_{it}^C \in [0, 1]$, the quantities $(\bar{x}^*, \bar{y}^*, \bar{z}^*)$ in the competitive equilibrium in Definition 1 have probabilistic interpretations: consumers select probabilities x_{it}^C of being scheduled at each time slot $t \in \mathcal{T}$, in order to maximize their *expected* net utility.

Our first result addresses the existence of the competitive equilibrium defined above.

Theorem 1: There exists a competitive equilibrium, given by an optimal solution $(\hat{q}^*, \hat{x}^*, \hat{y}^*, \hat{z}^*)$ to (SPP-R), and the following prices derived from an optimal dual solution to (SPP-R), for all i and t

$$\bar{p}_{it}^{\text{con}*} = p_{it}^{\hat{\lambda}^*} + p_{it}^{\hat{\nu}^*}, \bar{p}_i^{\text{gen}*} = \hat{\lambda}_i^*, \bar{p}_{it}^S = \tau_i \hat{\nu}_{it}^S, \bar{p}_{it}^E = \tau_i \hat{\nu}_{it}^E. \quad (26)$$

Proof: See [5].

The two fundamental theorems of welfare economics describe the relationship between competitive equilibria and efficient allocations. The first fundamental theorem states that competitive equilibria lead to, or *support* efficient allocations [16]. The second fundamental theorem states that the converse also holds, and in our setting corresponds to Theorem 1. We now state and prove the first fundamental theorem for our setting, given by Theorem 2. While proofs of the efficiency of competitive equilibria often require that the balance of supply and demand be included in the definition of such equilibria, in our development, this equality arises from the given formulation of the ISO's problem (ISO-R). That is, facing equilibrium prices, the ISO will act to balance supply and demand as it optimizes (ISO-R).

Theorem 2: Any competitive equilibrium forms an optimal solution for (SPP-R).

Proof: By definition, the competitive equilibrium $(\bar{q}^*, \bar{x}^*, \bar{y}^*, \bar{z}^*, \bar{p}^{\text{con}*}, \bar{p}^S^*, \bar{p}^E^*, \bar{p}^{\text{gen}*})$ satisfies

$$c'(\bar{q}_t^*) - \bar{p}_t^{\text{gen}*} \geq 0, \bar{q}_t^* (c'(\bar{q}_t^*) - \bar{p}_t^{\text{gen}*}) = 0 \quad \forall t \quad (27)$$

$$\bar{p}_{it}^{\text{con}*} + p_{it}^{\theta*} - \bar{U}_i \geq 0 \quad \forall i, t \leq T - \tau_i + 1 \quad (28)$$

$$\bar{x}_{it}^* (\bar{p}_{it}^{\text{con}*} + p_{it}^{\theta*} - \bar{U}_i) = 0 \quad \forall i, t \leq T - \tau_i + 1 \quad (29)$$

$$\bar{p}_{it}^{\text{con}*} + p_{it}^{\theta*} \geq 0 \quad \forall i, t > T - \tau_i + 1 \quad (30)$$

$$\bar{x}_{it}^* (\bar{p}_{it}^{\text{con}*} + p_{it}^{\theta*}) = 0 \quad \forall i, t > T - \tau_i + 1 \quad (31)$$

$$\bar{p}_{it}^S + \tau_i \theta_{it}^S - u_{it}^{dS} \geq 0 \quad \forall i, t \quad (32)$$

$$\bar{y}_{it}^* (\bar{p}_{it}^S + \tau_i \theta_{it}^S - u_{it}^{dS}) = 0 \quad \forall i, t \quad (33)$$

$$\bar{p}_{it}^{E*} + \tau_i \theta_{it}^{E*} - u_{it}^{dE} \geq 0 \quad \forall i, t \quad (34)$$

$$\bar{z}_{it}^* (\bar{p}_{it}^{E*} + \tau_i \theta_{it}^{E*} - u_{it}^{dE}) = 0 \quad \forall i, t \quad (35)$$

$$\bar{p}_t^{\text{gen}*} - \alpha_t^* \geq 0, \bar{q}_t^* (\bar{p}_t^{\text{gen}*} - \alpha_t^*) = 0 \quad \forall t \quad (36)$$

$$-p_{it}^{\bar{p}^{\text{gen}*}} + p_{it}^{\alpha^*} \geq 0, \bar{x}_t^* \left(-p_{it}^{\bar{p}^{\text{gen}*}} + p_{it}^{\alpha^*} \right) = 0 \quad \forall t \quad (37)$$

$$\alpha_t^* \left(\sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}}^t \bar{x}_{is}^* - g_t - \bar{q}_t^* \right) = 0 \quad \forall t \quad (38)$$

$$\alpha_t^* \geq 0 \quad \forall t \quad (39)$$

for some θ^{S*} , θ^{E*} , and $\alpha^* \geq 0$, as well as the feasibility conditions for each of the individual entity problems. Therefore, observing that for any $\bar{p}^{\text{gen}*} \geq 0$ the form of the objective in (ISO-R) ensures that complementary slackness condition (20) will be satisfied at the competitive equilibrium, selecting $(\hat{q}^*, \hat{x}^*, \hat{y}^*, \hat{z}^*) = (\bar{q}^*, \bar{x}^*, \bar{y}^*, \bar{z}^*)$ as the primal variables, and dual variables $\hat{\lambda}^* = \bar{p}^{\text{gen}*} = \alpha^*$ and $(\hat{\nu}_{it}^{S*}, \hat{\nu}_{it}^{E*}) = (\bar{p}^{S*}/\tau_i + \theta_{it}^{S*}, \bar{p}^{E*}/\tau_i + \theta_{it}^{E*})$ for all i, t , forms optimal primal and dual solutions to (SPP-R).

V. REPLICATED AND LARGE ECONOMIES

In general a competitive equilibrium is not guaranteed to exist when the social planner's problem is a mixed integer programming problem. Still, our competitive equilibrium definition allows for probabilistic allocations, and thus the existence of a competitive equilibrium is related to the existence of a primal and dual solution to the relaxed (SPP-R) problem. In this section, we justify the study of this relaxed problem by demonstrating its equivalence to the original, binary constrained (SPP) when each load i is interpreted as representing an infinite population of identical loads, with scaled demand.

Thus far, our development has crucially relied on the assumption that market participants are price-taking, i.e., presented with market prices; they make decisions in view of their own preferences and constraints, revealing their true demand without consideration of how their choices might influence these prices. But why should they act in this manner? In economic theory, the notion of large economies provides one justification for adoption of this assumption. The essence of the argument is as follows. As the number of market participants increases, any influence that an individual participant might have on market prices diminishes. When that number grows to infinity, that influence vanishes entirely, and the price-taking assumption becomes reasonable [1].

Following this intuitive argument, the question of how to add individuals to the market still remains. A special method, known as *replication*, is to introduce participants with preferences and constraints identical to existing types, in the same proportion as existing ones [12]. In the context of EV charging, this could mean scaling up the number of drivers with the same model vehicle and desired charging schedule, with demand scaled down so as to avoid infeasible aggregate demands as the population of each type grows. In general, it can be shown that as participants are

added in this way, those of the same type will receive the same allocation.

Suppose that each load i is replicated N times, and that the resulting loads have demand, utility, and disutility scaled by N . Indexing the replicas of each type i with the index n , the binary-constrained SPP with N replication is

$$\begin{aligned} (\text{SPP}(N)) \min_{\substack{\hat{x} \in \{0,1\}^{M \times N \times T} \\ \hat{q}, \hat{y}, \hat{z} \geq 0}} & \sum_t c(\hat{q}_t) - \sum_i \sum_n \frac{\bar{U}_i}{N} \sum_{t=1}^{T-\tau_i+1} \hat{x}_{int} \\ & + \sum_i \sum_n \sum_t \frac{u_{it}^{dS}}{N} (1 - \hat{y}_{int}) + \sum_i \sum_n \sum_t \frac{u_{it}^{dE}}{N} (1 - \hat{z}_{int}) \\ \text{s.t. } & \hat{\lambda}_t : \sum_i \sum_n \frac{l_i}{N} \sum_{s=\max\{1, t-\tau_i+1\}}^t \hat{x}_{ins} - g_t \leq \hat{q}_t \quad \forall t \\ & \hat{\nu}_{int}^S : \sum_{s=1}^t \sum_{r=\max\{1, s-\tau_i+1\}}^s \hat{x}_{inr} = \tau_i (1 - \hat{y}_{int}) \quad \forall i, n, t \\ & \hat{\nu}_{int}^E : \sum_{s=t}^T \sum_{r=\max\{1, s-\tau_i+1\}}^s \hat{x}_{inr} = \tau_i (1 - \hat{z}_{int}) \quad \forall i, n, t. \end{aligned} \quad (40)$$

We refer to the problem with N replication which relaxes the binary constraint on \hat{x} as (SPP(N)-R) (instead of (SPP(1)-R), we will still refer to the original relaxed problem as (SPP-R)). When we wish to emphasize the dependence of variables on the replication factor N , we will append (N) , e.g., $\hat{x}_{int}(N)$.

Proposition 3: Let $(\hat{q}^*, \hat{x}^*, \hat{y}^*, \hat{z}^*, \hat{\lambda}^*, \hat{\nu}_{int}^S, \hat{\nu}_{int}^E)$ denote an optimal solution to (SPP-R). Then, for any N , an optimal solution to (SPP(N)-R) can be formed by setting $\hat{x}_{int}^*(N) = \hat{x}_{it}^*$, $\hat{y}_{int}^*(N) = \hat{y}_{it}^*$, $\hat{z}_{int}^*(N) = \hat{z}_{it}^*$, for all i, n, t , $\hat{\lambda}_t^*(N) = \hat{\lambda}_t^*$ for all t , and $\hat{\nu}_{int}^S(N) = \hat{\nu}_{it}^S/N$ and $\hat{\nu}_{int}^E(N) = \hat{\nu}_{it}^E/N$ for all i and t .

Proof: (SPP(N)-R) has the following KKT conditions. For all t

$$c'(\hat{q}_t^*(N)) - \hat{\lambda}_t^*(N) \geq 0 \quad (42)$$

$$\hat{q}_t^*(N) \left(c'(\hat{q}_t^*(N)) - \hat{\lambda}_t^*(N) \right) = 0 \quad (43)$$

for all i, n and $T \leq T - \tau_i + 1$

$$\hat{p}_{it}^{\lambda*}(N)/N + \hat{p}_{int}^{\nu*}(N) - \bar{U}_i/N \geq 0 \quad (44)$$

$$\hat{x}_{int}^*(N) \left(\hat{p}_{it}^{\lambda*}(N)/N + \hat{p}_{int}^{\nu*}(N) - \bar{U}_i/N \right) = 0 \quad (45)$$

for all i, n , and $T > T - \tau_i + 1$

$$\hat{p}_{it}^{\lambda*}(N)/N + \hat{p}_{int}^{\nu*}(N) \geq 0 \quad (46)$$

$$\hat{x}_{int}^*(N) \left(\hat{p}_{it}^{\lambda*}(N)/N + \hat{p}_{int}^{\nu*}(N) \right) = 0 \quad (47)$$

for all i, n , and t

$$\tau_i \hat{\nu}_{int}^S(N) - u_{it}^{dS}/N \geq 0 \quad (48)$$

$$\hat{y}_{int}^*(N) \left(\tau_i \hat{\nu}_{int}^S(N) - u_{it}^{dS}/N \right) = 0 \quad (49)$$

$$\tau_i \hat{\nu}_{int}^E(N) - u_{it}^{dE}/N \geq 0 \quad (50)$$

$$\hat{z}_{\text{int}}^*(N) (\tau_i \hat{\nu}_{\text{int}}^{E*}(N) - u_{it}^{dE}/N) = 0 \quad (51)$$

and for all t (here suppressing the dependence of $\hat{\lambda}^*$, \hat{x}^* , and \hat{q}^* on N)

$$\hat{\lambda}_t^* \left(\sum_i \frac{l_i}{N} \sum_n \sum_{s=\max\{1, t-\tau_i+1\}}^t \hat{x}_{\text{ins}}^* - g_t - \hat{q}_t^* \right) = 0 \quad (52)$$

$$\hat{\lambda}_t^* \geq 0. \quad (53)$$

The proof of the theorem follows from making the selections specified in the theorem statement, substituting into (42)–(53), and comparing with (15)–(21).

Proposition 3 states that an optimal probabilistic schedule $\hat{x}^*(N)$ in the problem with N replication can be derived from an optimal probabilistic schedule \hat{x}^* for (SPP-R) and specifies how to do so. In the limit as $N \rightarrow \infty$, we can use \hat{x} to generate an optimal *deterministic*, binary-constrained schedule if we interpret \hat{x}_{it}^* as the proportion of the population of type i activated at time t . This is stated formally as follows.

Theorem 4: An optimal solution to (SPP(∞)) is given by activating proportion \hat{x}_{it}^* of type i population at time t for each i and t , where \hat{x}^* is an optimal solution to (SPP-R).

Proof: Note that constraint (40) may be rewritten

$$\hat{y}_{\text{int}} = 1 - \frac{1}{\tau_i} \sum_{s=1}^t \sum_{r=\max\{1, s-\tau_i+1\}}^s \hat{x}_{\text{inr}} \quad \forall i, n, t \quad (54)$$

and similarly for (41), so that (SPP(N)) can be written as

$$\begin{aligned} \min_{\substack{\hat{q} \geq 0, \\ \hat{x} \in \{0,1\}}} \quad & \sum_t c(\hat{q}_t) - \sum_i \overline{U}_i \sum_{t=1}^{T-\tau_i+1} \frac{1}{N} \sum_n \hat{x}_{\text{int}} \\ & + \sum_i \sum_t u_{it}^{dS} \left(\frac{1}{\tau_i} \sum_{s=1}^t \sum_{r=\max\{1, s-\tau_i+1\}}^s \frac{1}{N} \sum_n \hat{x}_{\text{inr}} \right) \\ & + \sum_i \sum_t u_{it}^{dE} \left(\frac{1}{\tau_i} \sum_{s=t}^T \sum_{r=\max\{1, s-\tau_i+1\}}^s \frac{1}{N} \sum_n \hat{x}_{\text{inr}} \right) \\ \text{s.t. } \hat{\lambda}_t : \quad & \sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}}^t \frac{1}{N} \sum_n \hat{x}_{\text{ins}} - g_t \leq \hat{q}_t \quad \forall t. \end{aligned} \quad (55)$$

Now, if $\hat{x}_{\text{int}}(N)$ is considered as a Bernoulli random variable with $P(\hat{x}_{\text{int}}(N) = 1) = \hat{x}_{it}^*$ and $\hat{q}_t^*(N)$ is chosen as $\hat{q}_t^*(1) = \hat{q}_t$ for all t , then by the Law of Large Numbers, constraint (55) converges to $\sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}}^t \hat{x}_{is}^* - g_t \leq \hat{q}_t^*$ for all t . Similarly, terms of the form $\frac{1}{N} \sum_t \hat{x}_{\text{int}} \rightarrow \hat{x}_{it}^*$ in the (SPP(N)) objective function as well. Since the optimal objective of the relaxed problem provides a lower bound for the binary-constrained problem, and load service is satisfied in the limit as $N \rightarrow \infty$, the solution produced by randomly activating loads according to \hat{x}^* converges to an optimal binary-constrained solution as $N \rightarrow \infty$.

VI. MARKET MECHANISM FOR LARGE POPULATION ECONOMY

Market mechanism design is an approach in economic theory which, rather than taking economic institutions as fixed and predicting the outcomes generated by such institutions, starts with an outcome identified as desirable and attempts to construct a mechanism by which it may be delivered [17]. In this section, we consider the competitive equilibrium concept discussed in prior sections as the target outcome for our market, and specify a mechanism by which it can be achieved. Mechanism design plays a crucial role for market in which participants may mis-report preferences, costs, or other information when it is in their individual best interest to do so. Therefore, the mechanism presented in this section may be viewed as a starting point for future work in which market participants are allowed to behave strategically.

The competitive equilibrium definition given in the previous section allows for nonbinary activation schedule \bar{x}^* . As mentioned, since $0 \leq \bar{x}_{it}^* \leq 1$, and $\sum_t \bar{x}_{it}^* \leq 1$, each \bar{x}_{it}^* may be interpreted as giving the portion of load i activated at time t under relaxation of the binary constraints on the activation schedule or the probability that an individual load of type i in the infinite replication setting is fully activated at time t .

Let us explore the infinitely replicated setting from the perspective of an individual load n of type i . It can be shown that under N replication and relaxation, the optimization problem for consumer n of type i is given by

$$\begin{aligned} (\text{CONS}_{in}(N)\text{-R}) \quad & \min_{\substack{x_{in}^C, y_{in}^C \\ z_{in}^C \geq 0}} \sum_t p_{\text{int}}^{\text{con}}(N) x_{in}^C - \frac{\overline{U}_i}{N} \sum_{t=1}^{T-\tau_i+1} x_{in}^C \\ & + \sum_t \left(\left(\frac{u_{it}^{dS}}{N} - p_{\text{int}}^S \right) (1 - y_{in}^C) + \left(\frac{u_{it}^{dE}}{N} - p_{\text{int}}^E \right) (1 - z_{in}^C) \right) \\ \text{s.t. } \theta_{\text{int}}^S : \quad & \sum_{s=1}^t \sum_{r=\max\{1, s-\tau_i+1\}}^s x_{inr}^C = \tau_i (1 - y_{in}^C) \quad \forall t \end{aligned} \quad (56)$$

$$\theta_{\text{int}}^E : \sum_{s=t}^T \sum_{r=\max\{1, s-\tau_i+1\}}^s x_{inr}^C = \tau_i (1 - z_{in}^C) \quad \forall t \quad (57)$$

$\hat{p}_{\text{int}}^{\lambda}(N)$ and $\hat{p}_{\text{int}}^{\nu}(N)$ are defined analogously to (13) and (14). Multiplying by N , the (CON_{in}(N)-R) objective function is

$$\begin{aligned} \sum_t N p_{\text{int}}^{\text{con}} x_{in}^C + \sum_t (u_{it}^{dS} - N p_{\text{int}}^S) (1 - y_{in}^C) \\ + \sum_t (u_{it}^{dE} - N p_{\text{int}}^E) (1 - z_{in}^C) - \overline{U}_i \sum_{t=1}^{T-\tau_i+1} x_{in}^C. \end{aligned} \quad (58)$$

As in Theorem 1, set

$$\begin{aligned} p_{\text{int}}^{\text{con}}(N) &= \hat{p}_{\text{int}}^{\lambda}(N)/N + \hat{p}_{\text{int}}^{\nu}(N), \quad p_t^{\text{gen}*}(N) = \hat{\lambda}_t^*(N) \\ p_{\text{int}}^S(N) &= \tau_i \hat{\nu}_{\text{int}}^{S*}(N), \quad p_{\text{int}}^E(N) = \tau_i \hat{\nu}_{\text{int}}^{E*}(N) \end{aligned}$$

and as in Proposition 3, choose

$$\hat{\lambda}_t^*(N) = \hat{\lambda}_t^*(1) = \hat{\lambda}_t^*, \quad \hat{\nu}_{\text{int}}^{S*}(N) = \hat{\nu}_{it}^{S*}/N, \quad \hat{\nu}_{\text{int}}^{E*}(N) = \hat{\nu}_{it}^{E*}/N.$$

Then, letting $N \rightarrow \infty$ gives

$$\lim_{N \rightarrow \infty} N\tau_i \hat{\nu}_{\text{int}}^{S*}(N) = \tau_i \hat{\nu}_{it}^{S*}, \quad \lim_{N \rightarrow \infty} N\tau_i \hat{\nu}_{\text{int}}^{E*}(N) = \tau_i \hat{\nu}_{it}^{E*}.$$

This implies that $\lim_{N \rightarrow \infty} Np_{\text{int}}^{\text{con}}(N) = p_{it}^* + p_{it}^{\hat{\nu}*}$. Therefore, posing the prices described in Proposition 3 in the limit as $N \rightarrow \infty$, the objective functions for all (CON_{in}) converge to

$$\begin{aligned} & \sum_t \left(p_{it}^* + p_{it}^{\hat{\nu}*} \right) x_{\text{int}}^C - \bar{U}_i \sum_{t=1}^{T-\tau_i+1} x_{\text{int}}^C + \\ & \sum_t \left((u_{it}^{dS} - \tau_i \hat{\nu}_{it}^{S*}) (1 - y_{\text{int}}^C) + (u_{it}^{dE} - \tau_i \hat{\nu}_{it}^{E*}) (1 - z_{\text{int}}^C) \right). \end{aligned}$$

Thus, the pricing facing each load of type i is identical, and, in fact, the problem facing each is the same as the single load of type i in the decomposition with relaxation but not replication. Further, each will select the same $x_{in}^{C*} = \hat{x}_i^* \in \mathbb{R}_+^T$, where $x_{\text{int}}^{C*} = \hat{x}_{it}^*$ gives the probability that the load will be scheduled at time t . Therefore, the equal allocation for individuals of the same type mentioned earlier holds in our setting.

The following mechanism (FLEX-SCHED(N)) uses the probability values selected by the continuum of consumers to generate a binary-constrained schedule in the setting with N replication. Note that since the generator's problem does not involve consumer utility and disutility functions, nor consumer-scheduling variables, its problem is not affected by replication (or relaxation). Therefore, $(\text{GEN}(N))$ is the same as (GEN) for all N , including $N = \infty$.

- 1) Each consumer (i, n) submits u_i^{dS} and u_i^{dE} , and the generator submits c to the social planner (i.e., the entity taking on this role, such as the government or ISO).
- 2) The social planner solves (SPP-R), and announces $(\bar{p}^{\text{con}*}, \bar{p}^{S*}, \bar{p}^{E*}, \bar{p}^{\text{gen}*}, \bar{p}^{\text{bal}*})$ as specified in Theorem 1.
- 3) Each consumer i solves $(\text{CON}(\infty)\text{-R}_{in})$, the generator solves $(\text{GEN}(\infty))$, and $(\bar{x}_i^*, \bar{y}_i^*, \bar{z}_i^*)$ for all i , as well as \bar{q}^* are submitted to the social planner.
- 4) The social planner randomly assigns proportion \bar{x}_i^* of loads of type i to start at time t , for each i and t . The generator produces \bar{q}^* over the finite horizon. Combined with the renewable generation output g , this generated power is allocated to the consumers according to \bar{x}_i^* and demands l_i for each i .

In the large population context, *individual rationality* is achieved if the optimal objective values to $(\text{GEN}(N))$, $(\text{ISO}(N))$, and $(\text{CONS}_{in}(N))$ for all i and n with $N = \infty$, i.e., the individual entity problems under infinite replication but without relaxation, under the (FLEX-SCHED(∞)) solutions are nonnegative. *Budget balance* is achieved if

$$\begin{aligned} & \sum_t p_t^{\text{gen}}(q_t^G + g_t) + \sum_t \sum_i \sum_n p_{\text{int}}^S (1 - y_{\text{int}}^C) \\ & + \sum_t \sum_i \sum_n p_{\text{int}}^E (1 - z_{\text{int}}^C) = \sum_i \sum_t \sum_n p_{\text{int}}^{\text{con}} x_{\text{int}}^C. \end{aligned} \quad (59)$$

Given these definitions, the following result regarding (FLEX-SCHED(∞)) holds.

Theorem 5: The mechanism (FLEX-SCHED(∞)) is *ex post* individually rational, budget balanced, and efficient.

Proof: See Ref. [5]. \square

VII. MULTIBUS EXTENSION

Thus far, we have set aside the network model in order to focus on the scheduling and pricing aspects of this problem. As posed above, our formulation above could be interpreted as modeling the cases where a single network point (e.g., EV charging station or smart building) is equipped with renewable generation, such as a solar panel array, as well as a backup generator [23]. More broadly, our single-bus model could represent the operation of a microgrid, with abstracted, aggregated renewable, and backup generation, while neglecting distribution network features in serving the flexible loads.

While we leave generalization to a full distribution network setting to future work, as a first step in this direction, we here consider an extension to a two-bus model, in which the renewable generation and flexible loads are co-located at one bus, and the conventional generator is located at the other.

We designate the bus where the loads and renewable generation are located as bus 1, and the bus where the conventional generator is located as bus 2. Selecting bus 1 as the slack bus, i.e., choosing from solutions with $\hat{\theta}_{1,t} = 0$ for all t gives the following relaxation of the two-bus social planner's problem:

$$\begin{aligned} & \min_{\hat{q}, \hat{x}, \hat{y}, \hat{z}, \hat{\theta} \geq 0} \sum_t c(\hat{q}_t) + \sum_i \sum_t u_{it}^{dS} (1 - \hat{y}_{it}) \\ & + \sum_i \sum_t u_{it}^{dE} (1 - \hat{z}_{it}) - \sum_i \bar{U}_i \sum_{t=1}^{T-\tau_i+1} \hat{x}_{it} \\ & \text{s.t. } \hat{\lambda}_t : g_t - \sum_i l_i \sum_{s=\max\{1, t-\tau_i+1\}}^t \hat{x}_{is} \geq -B\hat{\theta}_{2,t} \quad \forall t \\ & \hat{\mu}_t : \hat{q}_t = B\hat{\theta}_{2,t} \quad \forall t \end{aligned} \quad (60)$$

$$(10) - (11) \quad (61)$$

$$\hat{\gamma}_{12,t} : -B\hat{\theta}_{2,t} \leq f^{\max}, \quad \hat{\gamma}_{21,t} : B\hat{\theta}_{2,t} \leq f^{\max} \quad \forall t. \quad (62)$$

Note that given the slack bus choice, $0 \leq \hat{q}_t = B\hat{\theta}_{2,t}$, so that $-B\hat{\theta}_{2,t} < 0$ for all t , i.e., the first constraint in (62) is always loose and $\hat{\gamma}_{12,t}^* = 0$. See Ref. [5] for this problem's optimality conditions.

Employing the traditional locational marginal pricing scheme yields the nodal prices $\bar{p}_{1,t}^* = \hat{\lambda}_t^*$, $\bar{p}_{2,t}^* = \hat{\mu}_t^*$. It can be shown that at optimality, we have

$$\hat{\mu}_t^* = \hat{\lambda}_t^* - \hat{\gamma}_{21,t}^* \leq \hat{\lambda}_t^* \quad (63)$$

so that the per unit energy price at bus 2, i.e., the rate paid to the conventional generator, is upper bounded by the per unit energy price paid by the collection of flexible loads.

The two-bus network model affects the individual problems only via the generator and ISO objectives, as given below:

$$\text{generator : } \sum_t p_{2,t} q_t^G - c(q_t^G)$$

$$\text{ISO} : \sum_t \left[p_{1,t} \left(\sum_i l_i \sum_{s=\max\{1,t-\tau_i+1\}}^t x_{is}^I - g_t \right) - p_{2,t} q_t^I \right].$$

The consumer problem formulation does not change, as the energy consumption component of p_{it}^{con} now includes a $p_{it}^{\hat{\lambda}^*}$ term derived from the nodal price at bus 1, rather than the dual variables to the single power balance constraint from our initial formulation. See [5] for the aggregated individual entity problem optimality conditions.

Both the definition and proof of existence of a competitive equilibrium do not change in the two-bus model, save for the addition of nodal prices, rather than the single per time slot energy consumption/production price. Nor do the proofs of the relationship between competitive equilibria from the original relaxed setting, and the relaxed, replicated settings.

Given the shift from a single per time slot energy consumption/production price to a pair of prices, one for consumption at node 1, and one for production at node 2, the only remaining question is whether Theorem 5 continues to hold, in particular the budget balance component. Essentially due to inequality (63), instead of budget balance, we now have budget adequacy, i.e., under the proposed scheduling and pricing schemes, our market mechanism is either budget-balanced, or incurs a surplus in the form of congestion revenue.

To see this, the inequality which we would like to show is

$$\begin{aligned} \sum_t p_t^{\text{gen}} q_t^G + \sum_t \sum_i \sum_n (p_{\text{int}}^S (1 - y_{\text{int}}^C) + p_{\text{int}}^E (1 - z_{\text{int}}^C)) \\ \geq \sum_i \sum_t \sum_n p_{\text{int}}^{\text{con}} x_{\text{int}}^C = \sum_i \sum_t (p_{it}^{\hat{\lambda}^*} + p_{it}^{\hat{\rho}^*}) \hat{x}_{it}^*. \end{aligned} \quad (64)$$

Note that at optimality, if renewable generation exceeds the load scheduled for a time slot t , then we have that $\hat{\theta}_{2,t}^* = \hat{q}_t^* = \hat{\lambda}_t^* = \hat{\mu}_t^* = \hat{\gamma}_{21,t}^* = 0$, i.e., the generator is not dispatched at time t . For all other time slots, we have that

$$\hat{q}_t^* = \sum_i l_i \sum_{s=\max\{1,t-\tau_i+1\}}^t \hat{x}_{is}^* - g_t. \quad (65)$$

Together, (63) and (65) show that in problem instances where the line connecting the generator and load nodes becomes congested at any time slot, the summed load payments will in general exceed payments to the generator

$$\sum_t \hat{\mu}_t^* \hat{q}_t^* \leq \hat{\lambda}_t^* \left(\sum_i l_i \sum_{s=\max\{1,t-\tau_i+1\}}^t \hat{x}_{is}^* - g_t \right).$$

In such cases, the system operator will collect nonzero congestion revenue, and the mechanism is either budget-balanced or runs a surplus for the system operator.

VIII. CASE STUDY: EV CHARGING

EV charging constitutes one of the most important and challenging applications of load-scheduling optimization currently facing power grid operators. Today, the transportation sector accounts for approximately 64% of global consumption of oil, a

resource which has been linked to increasing CO₂ emissions, and further is expected to expire in about 50 years. In contrast, transportation sector operations comprise just 1.5% of worldwide electricity usage [6]. Reliance on electricity is more amenable to a shift toward renewable sources of energy such as solar and wind, which in total are expected to make up approximately one-third of all power generation by 2040. Further, according to the International Energy Agency, 740 000 vehicles were produced in total in 2014, and that figure is expected to reach 20 million by 2020 [6]. Charging scheduling is now viewed as one of the key technologies for integration of EVs into existing grids.

In order to demonstrate the utility of our flexible scheduling problem formulation, we simulated its performance on real-world load and renewable generation data in the context of EV charging. The input load parameters $(\tau_i, l_i, u_i^{dS}, u_i^{dE})$ are derived from data included in the ACN-Data dataset, a dynamic dataset of workplace EV charging [15]. In particular, we take as our base set of loads the recorded vehicle arrivals for May 28, 2018. For each vehicle charging session, the ACN dataset includes vehicle connection and disconnection times, as well as a charging completion time. In these simulations, each time index represents a 15-min period. We take τ_i as the difference between charging completion time and the first time period where the EV drew a positive amount of current. We then divide the total kWh delivered to the vehicle by τ_i to arrive at l_i .

Turning to load disutility functions, let t_C denote the vehicle's recorded connection time, and t_D denote its recorded disconnection time. Then, for each load i , we let

$$\begin{aligned} u_{it}^{dS} &= \alpha(t_C - t)^2 \mathbb{1}\{t \leq t_C - 1\} \\ u_{it}^{dE} &= \alpha(t - t_D)^2 \mathbb{1}\{t \geq t_D + 1\} \end{aligned} \quad (66)$$

where α is a scaling parameter. Thus, $u_{it}^{dS} + u_{it}^{dE} = 0$ for $t \in [t_C, t_D]$, and elsewhere increases quadratically away from this desired service window. The sample disutility curves pictured in Fig. 1 are generated according to this quadratic form.

We set $\bar{U}_i = \bar{U}$ for some nonnegative scalar \bar{U} . We take our renewable generation profile g_t from data generated by NREL's SAM tool [3] for downtown Los Angeles, also for May 28, 2018.

We compare the performance of our flexible load-scheduling approach to a schedule which naively begins charging loads as soon as they arrive. We implemented the latter by setting

$$\begin{aligned} u_{it}^{dS'} &= \alpha \max\{t_C^2, (T - t_C)^2\} \mathbb{1}\{t \leq t_C - 1\} \\ u_{it}^{dE'} &= \alpha \max\{t_C^2, (T - t_C)^2\} \mathbb{1}\{t \geq t_C + 1\} \end{aligned}$$

i.e., loads are essentially inflexible, with 0 disutility at time t_C and the maximum disutility in (66) for all other time periods $t \neq t_C$. We consider the social welfare objective value achieved, as well as the percentage of loads served.

We adjusted α , \bar{U} , and a scale on the quadratic cost $c(z_t)$ in order to ensure that both scheduling approaches successfully scheduled all loads. In particular, we let $\alpha = 0.01$, $\bar{U} = 100$, and scaled the cost by factor 0.5. As shown in Fig. 2(a) and (b), when users report disutility functions, the scheduler shifts loads such that that overall demand moves toward the mid-day period

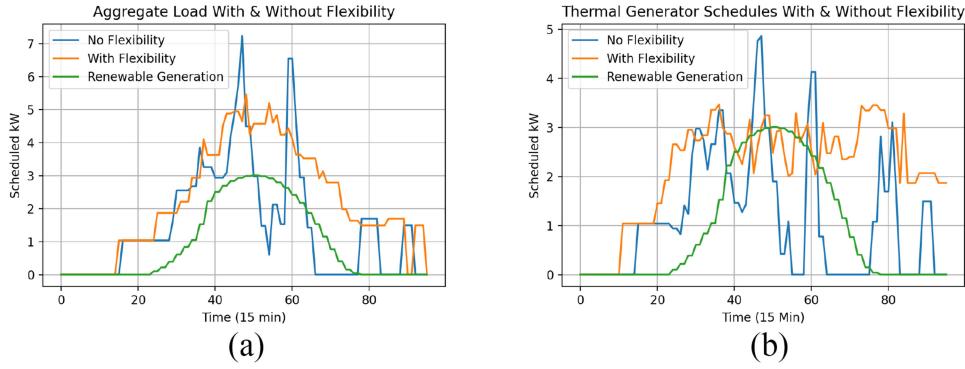


Fig. 2. Aggregate scheduled load (a) and generation (b) with and without load flexibility.

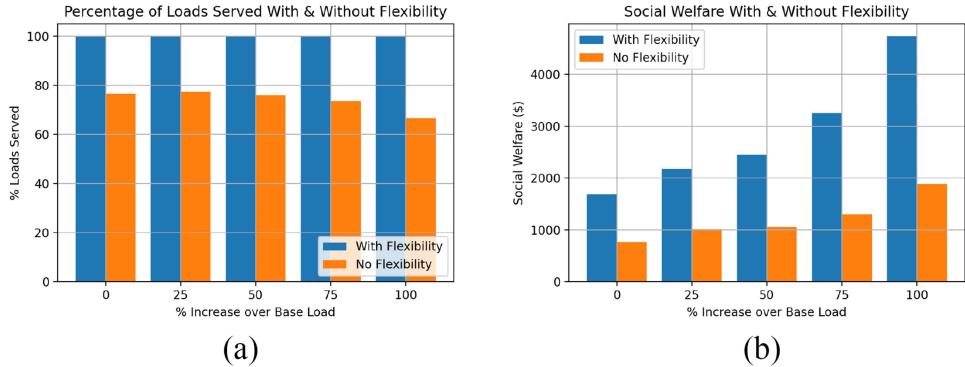


Fig. 3. (a) Percentage of loads served (a) and social welfare achieved (b) with and without flexibility.

of high renewable generation, thus relying less on the generator and incurring lower cost and higher social welfare overall. As seen in Fig. 2, the peak generation falls from 4.67 to 3.46 kW, a reduction of roughly 29%, while the peak demand falls from 7.23 to 5.46 kW, a reduction of roughly 24%.

To demonstrate the robustness of the disutility function approach to a surge in demand, we randomly sample loads served during the other weekdays of May 2018 in order to increase overall power demand. Specifically, we study demand scaled up in increments of 25% of the base load up to a 100% increase in load in terms of overall demand. Loads are randomly added to the base until the total power demand exceeds the desired level of increase, and the same random sets of loads are added in the cases with and without flexibility. The performance of both approaches are shown in Fig. 3(a) and (b). The flexibility-enabled schedule which makes use of the disutility functions continues to offer increased social welfare over on demand scheduling. Additionally, while disutility-based scheduling still includes all loads, the on-demand-based scheduler excludes between 25% and 33% of loads as the number of loads increases to double the base.

IX. CONCLUSION

In this work, we study how to schedule and price service for a set of flexible, but nonpreemptive loads, in the presence of renewable generation, as well as a dispatchable thermal

generator. Formulating a collection of mixed-integer optimization programs for the consumers, and the generator, we then study a centralized version of our setting with relaxed integer constraints, allowing for Lagrangian analysis and derivation of prices. A centralized solutions yields a competitive equilibrium, and conversely, a competitive equilibrium yields an efficient solution. Finally, we present a case study involving EV charging data to demonstrate the efficacy of our approach.

There are several directions for future work in this area. First, in terms of the scheduling aspect, it is desirable to determine a method for deriving at least an approximately optimal solution to the original integer-constrained setting, given an efficient solution to the relaxed social planner's problem presented here. In terms of pricing, properties such as fairness should be examined. For example, assuming that the disutility functions of each user can be at least partially ordered from less to more restrictive, is the compensation offered to more flexible users more than to those which are not as flexible? It will also be of interest to explore other types of loads, such as those which may be interrupted, as well as those which might accept less than an upper bound of total energy delivered. Strategic behavior should also be taken into account, as well as more detailed network modeling.

REFERENCES

[1] R. J. Aumann, "Markets with a continuum of traders," *Econometrica: J. Econometric Soc.*, vol. 32, no. 1/2, pp. 39–50, 1964.

[2] E. Bitar and Y. Xu, "Deadline differentiated pricing of deferrable electric loads," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp 13–25, Jan. 2017.

[3] N. Blair *et al.*, "System advisor model, SAM 2014.1.14: General description," Nat. Renewable Energy Lab. (NREL), Golden, CO, USA, Tech. Rep. NREL/TP-6A20-61019, 2014.

[4] P. Carrasqueira, M. J. Alves, and C. H. Antunes, "Bi-level particle swarm optimization and evolutionary algorithm approaches for residential demand response with different user profiles," *Inf. Sci.*, vol. 418, pp 405–420, 2017.

[5] N. Dahlin and R. Jain, "Scheduling of flexible nonpreemptive loads in Smart-Grid Networks," 2020, *arXiv:2003.13220*.

[6] C. Z. El-Bayeh, "A review on charging strategies of plug-in electric vehicles," *Eng. World*, vol. 1, 2019, pp. 9–33.

[7] A. Gupta, R. Jain, and R. Rajagopal, "Scheduling, pricing, and efficiency of nonpreemptive flexible loads under direct load control," in *Proc. IEEE 53rd Annu. Allerton Conf. Commun., Control, Comput.*, 2015, pp. 1008–1015.

[8] H. Abdulelah, J. Kleissl, and R. A. de Callafon, "Model predictive load scheduling using solar power forecasting," in *Proc. IEEE Amer. Control Conf.*, 2016, pp. 3200–3205.

[9] J. Han, J. Park, and K. Lee, "Optimal scheduling for electric vehicle charging under variable maximum charging power," *Energies*, vol. 10, no. 7, p. 933, 2017.

[10] M. U. Hashmi, "Load flexibility for price based demand response," 2018, Art. no. hal-01911254.

[11] M. Hijo and G. Frey, "Scheduling smart loads in modern buildings based on metaheuristic optimization," in *Proc. Int. Conf. Inform. Control, Automat. Robot.*, 2018, pp. 129–136.

[12] W. Hildenbrand, "On economies with many agents," *J. Econ. Theory*, vol. 2 no. 2, pp 161–188, 1970.

[13] R. Hledik, A. Faruqui, T. Lee, and J. Higham, "The national potential for load flexibility: Value and market potential though 2030," The Brattle Group, Tech. Rep., Jun. 2019.

[14] S. Hoff, "Electricity grid operators forecast load shapes to plan electricity supply," *Energy Central*, Jul. 29, 2016, Accessed: Oct. 1, 2020. [Online]. Available: <https://energycentral.com/c/iu/electric-grid-operators-forecast-load-shapes-plan-electricity-supply>

[15] Z. J. Lee, T. Li, and S. H. Low, "ACN-data: Analysis and applications of an open EV charging dataset," in *Proc. 10th ACM Int. Conf. Future Energy Syst.*, 2019, pp. 139–149.

[16] A. Mas-Colell *et al.* *Microeconomic Theory*, vol. 1. New York, NY, USA: Oxford Univ. Press, 1995.

[17] E. Maskin, "Introduction to mechanism design and implementation," in *Trans. Corporations Rev.*, vol. 11, no. 1, pp. 1–6, 2019.

[18] F.-L. Meng and X.-J. Zeng, "An optimal real-time pricing for demand-side management: A Stackelberg game and genetic algorithm approach," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2014, pp. 1703–1710.

[19] G. O'Brien and R. Rajagopal, "Scheduling nonpreemptive deferrable loads," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp 835–845, Mar. 2016.

[20] J. Qin, J. Mether, J.-Y. Joo, R. Rajagopal, K. Poolla, and P. Varaiya, "Automatic power exchange for distributed energy resource networks: Flexibility scheduling and pricing," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 1572–1579.

[21] D. Roberts, "Using electricity at different times of day could save us billions of dollars," *Vox*, Oct. 24, 2019, Accessed: Oct. 1, 2020. [Online]. Available: <https://www.vox.com/energy-and-environment/2019/8/7/20754430/renewable-energy-clean-electricity-grid-load-flexibility>

[22] F. Rossi, R. Iglesias, M. Alizadeh, and M. Pavone, "On the interaction between autonomous mobility-on-demand systems and the power network: Models and coordination algorithms," *IEEE Control Netw. Syst.*, vol. 7, no. 1, pp 384–397, Mar. 2020.

[23] K. Silverstein, "Solar-powered electric vehicle charging stations are just around the corner," *Forbes*, Feb. 10, 2020, Accessed: Oct. 1, 2020. [Online]. Available: <https://www.forbes.com/sites/kensilverstein/2020/02/10/solar-powered-electric-vehicle-charging-stations-are-just-around-the-corner/?sh=77f2a813320f>

[24] A. Subramanian, M. J. Garcia, D. S. Callaway, K. Poolla, and P. Varaiya, "Real-time scheduling of distributed resources," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp 2122–2130, Dec. 2013.

[25] B. Sun, Z. Huang, X. Tan, and D. HK Tsang, "Optimal scheduling for electric vehicle charging with discrete charging levels in distribution grid," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp 624–634, Mar. 2018.

[26] G. Wang, M. Negrete-Pincetic, A. Kowli, E. Shafeepoorfard, S. Meyn, and U. Shanbhag, "Dynamic competitive equilibria in electricity markets," in *Control and Optimization Theory for Electric Smart Grids*, A. Chakrabortty and M. Illic, eds. Berlin, Germany: Springer, 2011.



Nathan Dahlin (Student Member, IEEE) received the M.A. degree in applied mathematics and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, CA, USA, in 2020 and 2021, respectively.

He is currently a Postdoctoral Research Associate with the ECE Department, University of Illinois, Urbana-Champaign, IL, USA. His research interests include power systems market and mechanism design, stochastic optimization, and reinforcement learning.



Rahul Jain (Senior Member, IEEE) received the B.Tech. degree from IIT Kanpur, Kanpur, India, and the M.A. degree in statistics and the Ph.D. degree in EECS from the University of California, Berkeley, CA, USA, in 1993, 2002, and 2004, respectively.

He is a Professor of ECE, CS*, and ISE* (*by courtesy) with the University of Southern California, Los Angeles, CA, USA. His research interests include reinforcement learning, stochastic control, statistical learning, stochastic networks, and game theory, and power systems, transportation, and healthcare on the applications side.

Dr. Jain has received numerous awards, including the NSF CAREER award, the ONR Young Investigator award, an IBM Faculty award, the James H. Zumberge Faculty Research and Innovation Award, and was a US Fulbright Fellow from 2017 to 2020.