Proportional fairness for EV charging in overload

Martin Zeballos, Andres Ferragut and Fernando Paganini, Fellow, IEEE

Abstract—This paper studies a charging facility for Electrical Vehicles (EVs) at a parking garage, under the assumption that the infrastructure has a limited power capacity. In situations of overload, scheduling decisions must be made on which EVs to charge, possibly taking into account users’ sojourn times. We propose a fluid model that tracks service and sojourn times for a large population, enabling analytical studies of the distribution of service in steady state, for different policies. These results raise the issue of fairness in the distribution of partial service. We introduce a new policy called Least Laxity Ratio to achieve a suitable notion of proportional fairness. We test our results in simulation, including a real data-set with time-varying load. Our results show that our conclusions remain valid in this scenario, in particular the proposed policy performs well under an objective empirical measure of fairness.

Index Terms—Electric Vehicles, Stochastic Systems, Scheduling.

I. INTRODUCTION

The progressive deployment of Electrical Vehicles (EVs) is placing new requirements on the charging infrastructure [1]–[4]. Some of the demand will be covered by home charging, but another attractive option is to charge EVs at centralized parking lots, for instance at a large corporation. Initially, while EV penetration remains low, designing such facilities for peak power may be feasible. However as loads become higher it seems reasonable to provision power in a more conservative way; indeed, since occupation is statistically multiplexed, and EVs may tolerate some deferral of service, the need to turn on all chargers at the same time will be rare.

Operating such power constrained facilities requires a scheduling policy that, taking into account users sojourn times in the system, makes decisions on who receives charge at any given time. There is a rich literature on the scheduling of deadline constrained tasks, particularly in processor task scheduling [5], [6]; more recently the problem has received renewed attention in a smart-grid context [7]–[10] given the presence of deferrable loads. We highlight one peculiarity of the EV charging garage case: deadlines are inflexible (users leave the parking lot), but on the other hand partial service (charge) has value. One of the main features that a scheduling policy must have is, in the case of overload, to distribute the “pain” of partial service in an equitable manner.

In this paper we present an analytical framework to address these questions. In particular we introduce in Section II a dynamic model that treats the EV population as a fluid quantity, and allows for a unified representation of several scheduling policies, including some well known in the literature. With this machinery we can mathematically analyze, for the case of a stationary traffic load, the resulting steady-state conditions. In particular we show in Section III that in the case of overload, the distribution of partial service is highly dependent on the policy implemented, calling into question the fairness obtained by some popular policies. Aiming for a notion of proportional fairness in attained service, we introduce the new Least-Laxity-Ratio (LLR) policy.

To validate these results for practical systems we turn in Section IV to detailed simulations with a stream of individual charge requests arriving over time. We show first that the fluid approximation is accurate in systems of moderate scale. We also explore the important case of non-stationary load and heterogeneous EVs, using real data from the parking lots of a major tech company. Assuming a restricted capacity, we apply the different scheduling algorithms to these input traces; results show our theoretical conclusions remain valid during the peak, overload hours. We also adopt a quantitative index for comparison purposes, and apply it to the different policies. Conclusions are presented in Section V, and proofs are relayed to the Appendices. Partial versions of this work were presented in the conference papers [11], [12].

A. Related work

The analysis and design of charging policies for EV fleets is an active topic. In [13], the authors employ future demand estimation to maximize the state of charge of vehicles upon departure. In [7] the authors obtain competitive-ratio bounds for several scheduling policies with deadline constraints. Another relevant reference is [8], where the scheduling of a large aggregate of deferrable loads is studied, and many of the relevant policies such as earliest-deadline or least-laxity-first are brought into the smart grid context. In [9], [14], [15], EV scheduling is formulated as a dynamic program, with the optimization objective of minimizing reneged work; the resulting optimal policies in [9] take into account laxities and job sizes. Similar ideas in a time-varying environment with real data were analyzed in [10] where optimization based approaches are discussed. A proportional fairness approach with network considerations is given in [16]. More recently, [17] evaluates using the EV batteries to provide services to the grid, and [3], [4] discuss how to integrate the EV charging policies with the deployed infrastructure of the distribution network.

Closer to our work is the queueing approach of [18], further extended in [19] to include grid considerations. In comparison, our work covers more general scheduling policies when subject to deadlines, such as the ones introduced in [8]. In this regard [20], [21] laid the foundations for the analysis of reneged work in deadline systems with a single server.
A detailed study of deadline based policies such as earliest-deadline-first (EDF) is given in [22], [23] for both fluid and diffusion scales. A more comprehensive analysis of fluid limits for earliest deadline policies and many server queues is done in [24], [25], and constitutes an ongoing research subject. Our focus here is on the fluid scale itself, finding a common description for multiple policies that enables new designs. The empirical portion of this paper is closest to [10], where policies based on receding horizon optimization are proposed and tested with real data.

B. Contributions of this paper

Our first contribution is to introduce and develop a mathematical modeling strategy for the analysis of scheduling policies for EV charging. Based on a fluid approximation, our method applies to large scale systems, yielding mathematical expressions for the distribution of service and sojourn times among EVs in steady state, for a variety of policies.

The second contribution is to analyze the issue of fairness in partial service when power resources are scarce (overload). We show that the distribution of received service is highly dependent on the implemented policy, and arguably unfair in some popular policies. This motivates the design of the Least-Laxity-Ratio (LLR) policy, which is proved to achieve a notion of proportional fairness among jobs.

Our final set of results concerns the validation of our analytical conclusions in discrete simulations with real data. In particular we include practical, non-stationary load with heterogeneous EVs and of a realistic scale. We find that the mathematical model gives accurate predictions, in particular in regard to fairness during the peak overload hours. Using as a quantitative tool the well-known Jain’s index [26], we validate the fairness properties of different policies, and verify that LLR meets the design expectations.

II. ANALYTICAL MODEL

We consider a parking lot with an EV charging station per parking spot. The size of the parking lot is assumed large (infinite) so it never fills, but there is a limited power capacity that restricts the total consumption. The scheduling policy must allocate these limited resources among the EV clients currently present, taking into account their energy needs and their planned departure times.

In addition to the overall capacity limit, each charging station has a nominal power rating (maximum charging rate) that can be delivered to each EV. In our analytical model this quantity is assumed uniform across the parking lot; the assumption is relaxed for our simulation studies.

A. Discrete queueing model

As a first step toward our fluid population model we describe a discrete, stochastic counterpart. Here, vehicles arrive as a Poisson process of intensity λ, and arriving vehicles choose two random characteristics in i.i.d. fashion: a required service time $S_k$, i.e. the energy requested divided by the nominal power, and a sojourn time $T_k$, which is the time until the car leaves the parking lot. We assume that $S_k$ and $T_k$ follow general distributions, and $T_k \geq S_k$ with probability 1, which amounts to assuming that the demand of each EV is a priori feasible at the charging station. We denote by $S, T$ the random variables representing the characteristics of a general vehicle and by $f(\sigma, \tau)$ their joint density.

The garage operator must assign to each EV a charging rate $r_k(t)$, interpreted as a fraction of nominal power. We thus have

$$0 \leq r_k(t) \leq 1 \quad \text{for every } k, t.$$  (1)

Also, the capacity bound for the installation is expressed as

$$\sum_{k=1}^{n(t)} r_k(t) \leq C,$$  (2)

where $n(t)$ is the number of EVs present in the garage which still require service. $C$ can be interpreted as the maximum number of chargers that could be simultaneously turned on at full rate; we could, however, choose to activate more than $C$ chargers at a reduced rate.

We will consider charging policies that take into account the current population of EVs, and their residual times; for these, the system state can be represented as a counting measure on the service - sojourn space, as in [21], namely:

$$\Phi_t = \sum_{k=1}^{n(t)} \delta_{(\sigma_k(t), \tau_k(t))}.$$  

Here $\delta_{(\sigma_k(t), \tau_k(t))}$ is a point-mass measure in $\mathbb{R}^2$ located at the point $(\sigma_k(t), \tau_k(t))$, where $\sigma_k(t)$ is the remaining service time of each task and $\tau_k(t)$ is the remaining time until departure.

The dynamics is as follows: each point $k$ in the system consumes service time at a rate $r_k(t)$ and its lead time or time-to-deadline at rate 1, as depicted in Figure 1. The scheduling policy can thus be represented by a (possibly state-dependent) vector field on $\mathbb{R}_+^2$ given by:

$$\bar{u}(\sigma, \tau, \Phi) = - (r(\sigma, \tau, \Phi), 1).$$  (3)

Points follow this vector field up to $\sigma = 0$ (charge completed), or $\tau = 0$ (deadline expires). Completely charged EVs may stay in the parking lot, but we consider them served and out of our system. Vehicles that exhaust their time constraint depart the system with partial charge. We call reneged service the amount of charge that is not given to the vehicle, and corresponds to the value of the residual work $\sigma$ when $\tau$ reaches 0.

We restrict our attention to policies that do not waste charging opportunities; specifically, we require the following:

![Fig. 1. Dynamics for each job.](image-url)
Definition 1: A charging policy is called efficient if at every time \( t \), either (2) is satisfied with equality, or (1) is at its upper bound for every \( k = 1, \ldots, n(t) \).

As an example, consider that vehicles are served under the earliest deadline first (EDF) policy, where the first \( C \) vehicles with closer deadlines are served at rate \( r = 1 \). Then the system dynamics has the form depicted in Figure 2, and the rate \( r \) in equation (3) is:

\[
r(\sigma, \tau, \Phi) = 1_{\tau \leq \tau^*} \quad (4)
\]

with \( \tau^*(\Phi) = \sup \{ \tau : \Phi(\mathbb{R}^+ \times [0, \tau]) < C \} \). In words, given the state \( \Phi \) there are exactly \( C \) vehicles with residual sojourn times below \( \tau^*(\Phi) \); or, if the population is smaller than \( C \), the threshold \( \tau^*(\Phi) \) is infinite.

The process \( \Phi_t \), determined by the arrival distributions and the scheduling policy \( r \), is a measure-valued Markov process. Such a detailed system description is in general difficult to analyze. In a large scale situation, it is more tractable to consider a macroscopic, fluid description of the dynamics.

B. Fluid model

In the fluid scale, the number of points in the system is large and we replace \( g(t, \sigma, \tau) \) by a density function \( g(\sigma, \tau, g) \) over \( \mathbb{R}^2_+ \). Namely, \( g(t, \sigma, \tau) d\sigma d\tau \) represents the population with residual service/sojourn times in \([\sigma, \sigma + d\sigma] \times [\tau, \tau + d\tau]\). New mass arrives into the system at rate \( \lambda f(\sigma, \tau) \), where \( f \) is the joint density of service and sojourn times, as before. Mass is transported along the vector field \( \vec{u} = -(r(\sigma, \tau, g), 1) \) defined by the scheduling policy.

This implies that the density \( g \) should satisfy the following advection equation:

\[
\frac{\partial g}{\partial \tau} + \nabla \cdot (g\vec{u}) = \lambda f \quad (5)
\]

where \( \nabla \cdot (\cdot) \) is the divergence operator on \( \mathbb{R}^2_+ \), i.e. on the variables \( \sigma, \tau \). To explain this model, consider a region \( R \) of the \((\sigma, \tau)\) plane, with boundary \( \partial R \), depicted in Fig. 3. The total mass of particles (EVs) within this region at time \( t \) is given by

\[
\Phi_t(R) = \int_R g(t, \sigma, \tau) d\sigma d\tau.
\]

The variation of this quantity over time is due to arriving mass minus flow across the boundary. We therefore have:

\[
\frac{d\Phi_t(R)}{dt} = \lambda \int_R f(\sigma, \tau) d\sigma d\tau - \int_{\partial R} g(t, \sigma, \tau)[\vec{u} \cdot d\vec{n}].
\]

Here \( \vec{n} \) denotes the direction normal to the boundary; transforming this second term via the divergence theorem yields

\[
\int_R \frac{\partial g}{\partial \tau} d\sigma d\tau = \lambda \int_R f d\sigma d\tau - \int_R \nabla \cdot (g\vec{u}) d\sigma d\tau,
\]

which is the integral version of (5).

Remark 1: The formal relationship between stochastic and fluid models is beyond our scope here. Relevant references are [20]–[23], [25].

We are interested in steady-state solutions of eq. (5) so we set \( \frac{\partial g}{\partial \tau} = 0 \) and substitute the vector field, obtaining the equilibrium condition:

\[
\frac{\partial (r g)}{\partial \sigma} + \frac{\partial g}{\partial \tau} + \lambda f = 0. \quad (6)
\]

In the fluid model, the power constraints on the scheduling policy become:

\[
0 \leq r(\sigma, \tau, g) \leq 1; \quad \int_R r(\sigma, \tau, g) g(\sigma, \tau) d\sigma d\tau \leq C.
\]

In the next section we will analyze the steady state behavior of several scheduling policies by solving eq. (6) in each case.

C. Underload and Overload

The system load is defined as the product of the arrival rate and the mean service requirement,

\[
\rho := \lambda E[S] = \lambda \int_0^\infty \int_0^\infty f(\sigma, \tau) d\sigma d\tau.
\]

The load represents the average power requested to the garage; however, since we are representing service in units of time (normalizing energy by nominal charging power), then load is a dimensionless quantity. \( \rho \) represents the mean number of chargers needed to fully satisfy the demand. We say that the system is in underload when this mean number of chargers \( \rho < C \). If instead \( \rho > C \) we say that the system is in overload.

We now state some general results that apply to all efficient policies, depending only on the load conditions. Proof of both statements below can be found in Appendix A.

In the underload case the equilibrium conditions are independent of the policy. In particular all vehicles present will receive, full, immediate service:

Proposition 1: Assume \( \rho < C \). The steady state for any efficient policy is such that \( r = 1 \), the vehicle density is \( g(\sigma, \tau) = \lambda \int_0^\infty f(\sigma + x, \tau + x) dx \), and the total population is

\[
n = \int_0^\infty \int_0^\infty g(\sigma, \tau) d\sigma d\tau = \rho.
\]
In the overload case, the steady-state distribution will depend on the specific policy. However, the total amount of reneged work in the system, i.e., requested energy not delivered, is the same for all efficient policies.

**Proposition 2:** Assume that \( \rho > C \), and that the policy is efficient. Then the amount of reneged work in steady state is

\[
W = \int_0^\infty \sigma g(\sigma, 0) d\sigma = \rho - C.
\]

Again, reneged work is expressed here in dimensionless units.

### III. Scheduling Policies in Overload

Given the preceding results, the important pending question refers only to the overload situation (\( \rho > C \)). Namely: how is the overall reneged service \( W \) distributed between individual vehicles in the system? In this section we analyze this issue for different scheduling policies.

#### A. Earliest Deadline First

We begin by considering the EDF policy already described, depicted in Figure 2. The rate function for this policy is the fluid counterpart of (4):

\[
r(\sigma, \tau, g) = 1_{\{\tau \leq \tau^*(g)\}}.
\]

Eq. (8) says there is a threshold \( \tau^* \), dependent on the state \( g \) such that loads with remaining sojourner time \( \tau > \tau^* \) do not receive service. At equilibrium, this value is fixed. Hence, the typical service profile is to wait until the time-to-deadline is \( \tau^* \) and then be served up to completion or deadline expiration.

**Remark 2:** Note that the EDF policy only assigns extreme values of the rates, 0 or 1, and does not exploit the intermediate range. While some other policies we analyze below share this extreme property, it is not a requirement and indeed we will study one that employs partial service rates. Our modeling technique covers the general case.

The following result is obtained by explicitly solving (via characteristic curves) the PDE (6) under rate function (8). Its proof is found in Appendix B.

**Proposition 3:** Under the EDF policy in overload, the reneged work \( S_r \) per user is given by \((S - \tau^*)^+\), where the threshold \( \tau^* \) satisfies

\[
\lambda E[\min\{S, \tau^*\}] = C.
\]

#### B. Least Laxity First

The second policy we analyze is Least Laxity First (LLF) from the real-time scheduling literature [6], introduced in the smart-grid context in [8], and more recently discussed for the EV charging problem in [9]. Here, the laxity or spare time of each job is considered, defined by \( \ell_k = \tau_k - \sigma_k \), i.e., the amount of time that the job can be delayed and still be able to meet its deadline.

The LLF policy, fills capacity with the jobs of lowest laxity, so the rate function is

\[
r(\sigma, \tau, g) = 1_{\{\sigma \leq \ell^*(g)\}}.
\]

Figure 4 depicts the decisions of LLF for the same points of Figure 2; note the different EVs in service.

In this case, the system serves loads with laxity \( \ell_k \leq \ell^* \) at full rate, while the rest consume their spare time up to reaching this level. In equilibrium, this laxity level \( \ell^* \) is fixed, and in overload this laxity level becomes negative, implying that all jobs depart with reneged work. The equilibrium of (6) with rate function (10) is as follows (proof in Appendix B):

**Proposition 4:** Under the LLF policy in overload, the reneged work \( S_r \) per user is given by \( \min\{S, \sigma^*\} \), where the threshold \( \sigma^* = -\ell^* \) satisfies

\[
\lambda E[(S - \sigma^*)^+] = C.
\]

When the system is in overload, it finds a threshold \( \ell^* < 0 \) and all EVs with initial laxity \( \ell > \ell^* \) consume their spare time up to reaching \( \ell^* \) and leave the system with \( S_r = \sigma^* := -\ell^* \) when their deadlines expire. However, if a given EV arrives with service request \( S < \sigma^* \), it never attains the required laxity level for service and departs the system without being charged at all, leaving with \( S_r = S \). Thus, an LLF system in overload discriminates against small jobs. Again, the system ends up not discriminating by \( T \) (time in the system), only job size.

#### C. Fair Scheduling: Least Laxity Ratio

The policies analyzed so far do not show a fair behavior in overload; EDF discriminates against large jobs, by chopping their service, and LLF discriminates against small jobs, by not giving service at all. We now propose a new policy that we call Least-Laxity-Ratio (LLR). It works as follows: given
the current set of jobs with remaining service and deadlines \((σ_k, τ_k)\), construct the following index called laxity ratio:

\[
θ_k := \frac{τ_k}{σ_k} = 1 + \ell_k.
\]

(12)

Then serve the \(C\) jobs with smallest \(θ_k\) in the system at full rate. The policy serves the most urgent loads, i.e. those with more urgent deadlines, relative to their residual service time.

The behavior of the policy is depicted in Figure 5, again for the same points of Figures 2 and 4; note the differences.

The rate allocated to each job is given by:

\[
r(σ, τ, g) = 1\{g ≤ θ(g)\},
\]

(13)

where \(θ(g)\) is the threshold laxity ratio.

In equilibrium, this ratio reaches a value \(θ^*\); an EV will receive no service until \(τ/σ\) falls below the threshold \(θ^*\), and receive unit rate after that. In overload, \(θ^* < 1\) meaning that jobs must be lagging behind their deadline to get service, and will always have some reneged service. The following proposition is proved in Appendix B.

**Proposition 5:** Under the LLR policy in overload, the reneged work \(S_r\) per user is given by \((S - r^*T)^+\), where the equilibrium \(r^*\) satisfies

\[
λE[\min\{S, r^*T\}] = C.
\]

(15)

Once more, equation (15) has a single solution \(0 < r^* < 1\) under the overload condition. (15) is analogous to the fixed point equation derived in [21] for the single server PS queue, and in [18] for the EV case under exponential assumptions.

We find here that the reneged work depends on both the service requirement \(S\) and the sojourn time \(T\) offered to the system, rewarding EVs that offer more flexibility. Except for this fact, PS is similar to EDF: for equal sojourn time \(T\), the PS policy favors small jobs, which are served to completion, while large jobs only receive partial service.

Table 1 summarizes our analytical results. We end the section with brief remarks on an alternative set of policies.

### E. Optimization-based policies

Some recent references (e.g. [8], [10]) propose to schedule based on online optimization. In this receding horizon approach, at a given time a cost function that looks ahead a certain amount of time is optimized, with the information of EVs currently present. If the cost is such that EVs with equal residual times \((σ, τ)\) are treated equally, this policy would fit our framework, albeit with a complex rate function.

A major issue in this kind of method is to guarantee the feasibility of the constraints at every time step, otherwise the policy is undefined. [8] ensures feasibility through the inclusion of reserve generation, which we do not have here. The Smoothed Least Laxity First policy of [10] for EVs assumes feasibility at each time step, which would not apply to an overload scenario which, in their terminology, is not offline feasible.

### IV. Simulation Studies

We developed a discrete-event simulator over Julia [27] to carry out several experiments of a parking lot under the different scheduling policies.

A first aim would be to validate the fluid approximations made in our theory for the stationary case, against a scenario of discrete EVs. This was our main focus in [11]; we extract here one representative comparison. We simulated a stochastic, discrete system under Poisson arrivals of rate \(λ = 120\) EVs/hr, with job sizes \(S_k\) of exponential distribution, mean 1hr. The load is thus \(ρ = 120\), and we choose \(C = 60\) to have an

\[n(t) = 9 \text{ and } C = 3.\]

![Fig. 5. Least laxity ratio policy behavior for \(n(t) = 9\) and \(C = 3\).](image)

**TABLE I**

<table>
<thead>
<tr>
<th>Policy</th>
<th>Threshold equation</th>
<th>Attained service ((S_a))</th>
<th>Reneged service ((S_r = S - S_a))</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDF</td>
<td>(λE[\min{S, r^*}] = C)</td>
<td>(\min{S, r^*})</td>
<td>((S - r^*))</td>
</tr>
<tr>
<td>LLF</td>
<td>(λE[(S - σ^*)^+] = C)</td>
<td>((S - σ^*)^+)</td>
<td>(\min{S, σ^*})</td>
</tr>
<tr>
<td>LLR</td>
<td>(θ^*E[S] = C)</td>
<td>(θ^*S)</td>
<td>((1 - θ^*)S)</td>
</tr>
<tr>
<td>PS</td>
<td>(λE[\min{S, r^*T}] = C)</td>
<td>(\min{S, r^*T})</td>
<td>((S - r^*T)^+)</td>
</tr>
</tbody>
</table>

\[σ_t = θ^* σ\]
overload situation. The initial laxities $L_k$ were taken to be exponential, with mean 2 hours.

Figure 6 shows the service time $S$ demanded by each EV against the attained service time $S_a$ under EDF, LLF, LLR, compared with the fluid model predictions. The correspondence is seen to be very close, and the EDF and LLF policies discriminate against large and small jobs respectively. In contrast, our proposed LLR policy achieves the desired linear relationship, imposing proportional fairness across jobs.

We have tested the validity of the fluid approximation for lower levels of the load. Results (omitted due to space limitations) show that even for populations 3 times smaller, the fluid model captures well the mean values, albeit with higher dispersion around this mean.

The second, more important objective for our simulation environment is to explore scenarios which escape some of the simplifying assumptions of our theory:

- **Study time-varying** loads. In a practical system, arrival patterns and charging requests are not stationary, they reflect daily use cycles. And, since sojourn times are long with respect to these variations, we cannot assume that each EV sees an approximately steady situation. Ideally, in a typical day with varying levels of congestion, the scheduling policy should only play a major rule during the intervals of overload, and impose during those times the appropriate fairness in the curtailed service.

- **Cover heterogeneous** EVs, in regard to their nominal charging power.

A good way to investigate such practical conditions is through a data set of real experimental EV charging data.

### A. Data set and experimental results.

A dataset of about 100 EV parking lots of a major tech company was obtained, each with a capacity between 10 and 70 vehicles. The data contains values of: arrival times, sojourn times, consumed energy and the maximum charging power rate for each car, during a period of about 50 days.

We wish to represent the load of a large parking lot, for example a large office building or shopping mall, subject to an aggregate power capacity restriction. The dataset we obtained corresponded to many smaller lots, so we merged 20 of them to create a data set of about 500 cars, arriving during one day and reaching a maximum of 170 cars parked at the same time. Nominal charging power is heterogeneous, in the range 1.4 to 7.0 kW with an average of 5.0 kW. Sojourn times have a mean of $\bar{T} = 2.47$ hr and a mean requested energy of 10.3 kWh. The main feature is the variability of congestion levels: Fig. 7 shows the total requested system power in the parking lot within 24 hours. As expected the system presents more power demand in working hours, while there is almost no power requirement at night.

We assume that this aggregate load is applied to a parking lot with individual charging stations, but where the total maximum power is 150 kW, equivalent to $C = 30$ chargers at the mean nominal power of 5 kW. Regarding the merging of smaller lots mentioned above, we do not maintain this information, and in particular we do not impose capacity limits of the individual components. Analyzing such a situation would be an interesting topic for further research.

In our initial study [12] with this data we had artificially normalized car ratings to this value. For the present paper we have improved our simulator to accommodate individualized nominal ratings, keeping track of residual energy for each EV. Priorities are still established in terms of residual service times (energy/power) and sojourn times. Some curtailing appears when accommodating the last car within the power capacity constraint.

In Fig. 8 we plot a snapshot of the system in a condition of high congestion (with more than 100 cars in the parking lot), for each of the three policies, EDF, LLF and LLR. We classify cars according to whether they are in service, and for illustration we mark the empirical threshold between the two classes. The transition has the same qualitative behavior as in the stationary load case.

### B. Proportional fairness measurement

In our theory for stationary load we showed that all efficient policies have the same underload behavior, and differences occur when in overload. With a non-stationary load as in Fig. 7 we expect that EVs present during intervals of underload will receive full service, but curtailment will appear in the overload hours; the main fairness question is how partial service is allocated during these intervals.
This quantity \( J \in [0,1] \), and reaches unity only if all \( x_k \)'s are the same, i.e. under proportional fairness.

**Remark 3:** In the fluid limit of the stationary case, the LLR policy achieves the optimal Jain’s Index \( J = 1 \) by definition since \( S_{a,k} = \theta^* S \), so \( x_k = \theta^* \) for every \( x_k \).

Since by definition \( x_k \) can only be evaluated at the end of service, to obtain a time-varying measurement of fairness we will compute Jain’s index at time \( t \) between the set of cars which have finished service in a window of time \( [t-t_0, t] \).

In Figure 9 we plot the Jain fairness index computed across a typical day considering a half-hour window, i.e. with \( t_0 \) equal to half an hour. As expected the index is 1 when there is no congestion in the system, here \( x_k = 1 \) for all EVs.

During busy hours (see Fig. 7) the fairness index decreases, with different behavior across policies. EDF shows a sudden drop in the fairness index but recovers to values near 1 quickly. In the case of LLF, the fairness index suffers the most, staying near 0.5 for several hours. On the other hand, for LLR the fairness index stays close to unity at all times, with far less variability, evidence of the desired proportional fairness behavior.

**V. CONCLUSIONS AND OPEN QUESTIONS**

In this work we have analyzed through mathematical models and simulations the performance of charging policies for an EV garage. The scenario of interest is when limits apply to the total power of the installation, which imply that some curtailment takes place during intervals of overload.

We introduced a fluid PDE model for the EV population, which provides an elegant macroscopic characterization of different policies. In particular for stationary traffic load, we obtained analytic expressions for the service attained by an overloaded system under the standard EDF, LLF and PS policies, and introduced the LLR policy to achieve proportional fairness.

Our simulation studies with real data showed that conclusions remain valid under time-varying load and heterogeneous EVs. We also introduced a fairness index that may be used as a figure of merit in empirical studies of charging policies. An interesting open question would be the mathematical analysis of the time-varying situation.

As is always the case with issues of fairness, our proportional criterion is not the only possible choice. Alternatives of the “max-min fairness” kind can be considered. Maximizing the minimal service delivered would imply prioritizing small jobs (like EDF does), which does not seem motivated in this context. Instead one could think of maximizing the minimum state of charge of departing EVs; this suggests favoring large jobs which reflect low initial battery levels. In this viewpoint, the LLF policy becomes more attractive, and indeed the results in [9] follow essentially this criterion. An interesting question
for future research is the extent to which such alternatives can be studied within our fluid framework.

A broader open avenue of research would be to analyze, with an economic perspective, the provisioning decisions of the garage operator as a function of load patterns and customers’ valuation of partial service.

ACKNOWLEDGMENT

The authors would like to thank Steven Low and his team of the California Institute of Technology who kindly provided us the EV charging dataset.

APPENDIX A

EFFICIENT POLICIES IN STEADY STATE

Proof of Proposition 1 (underload case): Let \( \rho < C \). We verify that \( r \equiv 1 \) is an equilibrium solution, applying it to the advection equation (6) in steady state, which becomes:

\[
\frac{\partial g}{\partial \sigma} + \frac{\partial g}{\partial \tau} + \lambda f = 0. \tag{17}
\]

The preceding equation can be solved by the method of characteristics [28]. Consider the following total derivative:

\[
\frac{d}{dx} g(x + \tau, \sigma + x) = \frac{\partial g}{\partial \sigma} + \frac{\partial g}{\partial \tau} = -\lambda f(x + \tau, \sigma + x). \]

By integration, and using the boundary condition \( g(\sigma, \tau) \rightarrow 0 \) when \( \sigma, \tau \rightarrow \infty \), we have the solution to (17):

\[
g(\sigma, \tau) = \lambda \int_{\tau}^{\infty} f(\sigma, x + \tau, x + \tau) dx. \tag{18}
\]

In order to be a suitable equilibrium, we should verify that \( g \) in (18) yields a total rate less than \( C \). We have:

\[
\int_{0}^{\infty} \int_{0}^{\infty} g(\sigma, \tau) d\sigma d\tau = \lambda \int_{0}^{\infty} [\int_{0}^{\infty} P(S > x, T > x) dx] d\sigma
\]

\[
= \lambda \int_{0}^{\infty} [\int_{0}^{\infty} P(\min\{S, T\} > x) dx]
\]

\[
= \lambda E[\min\{S, T\}] = \lambda E[S] = \rho < C. \tag{19}
\]

Note that by assumption \( S \leq T \) a.s.

Proof of Proposition 2 (overload case): We first observe that in overload, the efficiency of the policy implies

\[
\int_{0}^{\infty} \int_{0}^{\infty} r(\sigma, \tau) g(\sigma, \tau) d\sigma d\tau = C. \tag{20}
\]

Indeed, the alternative for efficiency would be to set \( r = 1 \) for every point; but then the same calculation as (19) would yield, for the left-hand side of (20), a result \( \rho > C \), violating our capacity constraint.

\[
W = \int_{0}^{\infty} \sigma g(\sigma, 0) d\sigma = \int_{0}^{\infty} \sigma \int_{0}^{\infty} \left[ \frac{\partial g}{\partial \sigma}(\sigma, \tau) \right] d\tau d\sigma
\]

\[
= \int_{0}^{\infty} \int_{0}^{\infty} \sigma \left[ \lambda f + \frac{\partial g}{\partial \tau}(\sigma, \tau) \right] d\tau d\sigma
\]

\[
= \lambda \int_{0}^{\infty} \int_{0}^{\infty} f d\sigma d\tau + \int_{0}^{\infty} \int_{0}^{\infty} \sigma \frac{\partial g}{\partial \sigma}(\sigma, \tau) d\sigma d\tau
\]

\[
= \lambda E[S] - \int_{0}^{\infty} \int_{0}^{\infty} rg d\sigma d\tau = \rho - C,
\]

where we have used (6), integration by parts and (20).

APPENDIX B

DISTRIBUTION OF RENEGED WORK IN STEADY-STATE

Proof of Proposition 3 (EDF): Substituting the EDF rate function \( r = 1(\tau < \tau^*) \), the equilibrium PDE (6) becomes:

\[
\frac{\partial g}{\partial \sigma} + \frac{\partial g}{\partial \tau} + \lambda f = 0 \quad \tau < \tau^*,
\]

\[
\frac{\partial g}{\partial \tau} + \lambda f = 0 \quad \tau > \tau^*,
\]

which is a two part transport equation. The characteristic trajectories of this equation are derived from following the vector field; the first part of the equation is analogous to the underload case (parallel service), for the second part the trajectories are vertical lines with constant \( \sigma \). This leads to the following solution integrating along trajectories:

\[
g(\sigma, \tau) = \lambda \int_{\tau}^{\infty} f(\sigma, x + \tau, x + \tau) dx, \quad \tau > \tau^*;
\]

whereas for \( \tau < \tau^* \) the solution has two parts.

\[
g(\sigma, \tau) = \lambda \int_{0}^{\tau^* - \tau} f(\sigma + x, \tau + x) dx +
\]

\[
+ \lambda \int_{\tau^*}^{\infty} f(\sigma + \tau^* - \tau, \tau) dx, \quad \tau < \tau^*. \tag{21}
\]

To compute the distribution of reneged work we evaluate the above at the point \((\sigma_r, 0)\). Here the second formula (21) applies; now since we have assumed \( f(\sigma, \tau) = 0 \) for \( \sigma > \tau \), the first term in (21) disappears \((f(\sigma_r + x, x) = 0 \forall x)\), and in the second we can begin integrating at \( x = \sigma_r + \tau^* \). We obtain:

\[
g(\sigma_r, 0) = \lambda \int_{\sigma_r + \tau^*}^{\infty} f(\sigma_r + \tau^*, x) dx.
\]

The integral above amounts to marginalizing \( f(\cdot, \cdot) \) over its second variable (sojourn time), and reducing service time by \( \tau^* \), with positive truncation: i.e. it is the density of the random variable \((S - \tau^*)^+\).

Therefore, \( S_r \sim (S - \tau^*)^+ \) would be the reneged work per client, and the overall rate for reneged work is \( W = \lambda E[(S - \tau^*)^+] \). Using Proposition 2 we obtain the equilibrium condition for \( \tau^* \):

\[
\lambda E[(S - \tau^*)^+] = \rho - C.
\]

Equivalently, noting that \( S - (S - \tau^*)^+ = \min\{S, \tau^*\} \):

\[
\lambda E[\min\{S, \tau^*\}] = C.
\]

Since \( \min\{S, \tau^*\} \leq S \), the above equation always has a solution in the overload situation \( \rho > C \).

Proof of Proposition 4 (LLF): Under the rate function \( r(\tau) \) with equilibrium threshold \( \ell^* \), the vector field is \( \vec{u} = -1(\tau^* - \sigma < \ell^*), 1) \).

Since the system is in overload, we posit a solution with \( \ell^* < 0 \) (expired laxity on departure). This defines a threshold line \( \tau = \sigma + \ell^* \), below the diagonal; we discuss the dynamics in separate regions:

- Above the line there is no service so (6) becomes:

\[
\frac{\partial g}{\partial \tau} + \lambda f = 0 \quad \tau - \sigma > \ell^*, \tau, \sigma > 0.
\]
• When flow reaches the threshold line \( \tau - \sigma = \ell^* \), it moves along it and departs the system through \((0, -\ell^*) = (0, \sigma^*)\). Since mass accumulates in this line we have a singularity in the measure, strictly speaking not a density.

• In the region below the line, \( \tau < \sigma + \ell^* < \sigma \), so by assumption there is no arriving flow; also from the previous case no flow enters the region from above, so in steady-state the density in this region is zero.

Above the threshold line, the steady-state solution is simply

\[
g(\sigma, \tau) = \lambda \int_{-\infty}^{\infty} f(\sigma, x) dx, \quad \tau - \sigma > \ell^*, \tau, \sigma > 0.
\]

We will avoid the characterization of the singularity by simply focusing on the flow of reneged work, which is the integral of \( g(\sigma, 0) \) across the distribution variable \( \sigma \). For \( \sigma < \sigma^* \) (corresponding to \((\sigma, 0)\) above the threshold line), we have

\[
g(\sigma, 0) = \lambda \int_{0}^{\infty} f(\sigma, x) dx;
\]

this is the marginal density of \( S \) at \( \sigma \), times \( \lambda \). The remainder of the reneged work distribution is a point mass at \( \sigma = \sigma^* \), resulting from the singular flow at the threshold boundary.

So the reneged work per job is distributed as \( S_r \sim \min\{S, \sigma^*\} \), and \( W = \lambda E[\min\{S, \sigma^*\}] \). Using Proposition 2 we find the equilibrium condition for \( \sigma^* \):

\[
\lambda E[(S - \sigma^*)^+] = \rho - C;
\]

or equivalently, noting that \( S - \min\{S, \sigma^*\} = (S - \sigma^*)^+ \):

\[
\lambda E[(S - \sigma^*)^+] = C.
\]

As a function of \( \sigma^* \), the left-hand side starts at \( \rho > C \) for \( \sigma^* = 0 \), and decreases to zero as \( \sigma^* \to \infty \); hence the above equation always has a solution.

Proof of Proposition 5 (LLR): Imposing the rate function (13), the equilibrium PDE (6) takes the form

\[
\begin{align*}
\frac{\partial g}{\partial \sigma} + \frac{\partial g}{\partial \tau} + \lambda f &= 0 \quad \tau < \sigma^*, \\
\frac{\partial g}{\partial \tau} + \lambda f &= 0 \quad \tau > \sigma^*,
\end{align*}
\]

for a certain threshold \( 0 < \sigma^* < 1 \); this is similar to the EDF case, but with an oblique threshold line. Note that all characteristic trajectories cross this line since \( \sigma^* < 1 \).

Above the threshold, the solution is, as before:

\[
g(\sigma, \tau) = \lambda \int_{\tau}^{\infty} f(\sigma, x) dx, \quad \tau > \sigma^*.
\]

If we start with \((\sigma, \tau)\) below the threshold line, we integrate first along the characteristic curve \((\sigma + x, \tau + x)\) until reaching the line:

\[
g(\sigma, \tau) = \lambda \int_{0}^{\sigma - \sigma^*} f(\sigma + x, \tau + x) dx
\]

\[
+ g_-(\frac{\sigma - \tau - \theta^*}{1 - \theta^*}, \sigma - \tau),
\]

and then we would apply formula (22) for the term (24). However, due to flow preservation when crossing the line, we must impose \((1 - \theta^*)g_+ = g_- \) at the boundary; this is because the normal component of the vector field \( \tilde{u} \cdot \tilde{n} \) is, respectively, \( 1 - \theta^* \) and 1 below and above the line. So the correct expression for the term in (24) is:

\[
g_-(\cdot, \cdot) = \lambda \int_{1 - \theta^*}^{\infty} f \left( \frac{\sigma - \tau - \theta^*(\sigma - \tau)}{1 - \theta^*}, x \right) dx.
\]

With the above solution we can compute the distribution of reneged work \( g(\sigma, 0) \). Using the fact that \( f(\sigma, \tau) = 0 \) if \((\sigma, \tau)\), the term in (23) is 0, and the expression in (25) can be simplified as well, leading to:

\[
g(\sigma, 0) = \lambda \int_{0}^{\infty} f \left( \frac{\sigma - \tau}{1 - \theta^*}, x \right) dx.
\]

Except for the factor \( \lambda \), the expression above corresponds to the density of the random variable \((1 - \theta^*)S\). Indeed, the variable \( T \) has been marginalized from the joint density, and the change of variables applied to \( \sigma \). We conclude that the reneged work per client is \( S_r \sim (1 - \theta^*)S \).

We can now find the threshold \( \theta^* \) from the equation

\[
W = \lambda (1 - \theta^*)E[S] = \rho - C,
\]

applying Proposition 2. Equivalently, \( \theta^* = C/\rho \), which satisfies \( \theta^* < 1 \) when in overload, as desired.

References


Andres Ferragut obtained his PhD degree in Electrical Engineering (Telecommunications) from Universidad de la Republica, Uruguay (2011). He previously held teaching and research positions at the Mathematics and Electrical Engineering Depts. at Univ. de la Republica (2000 to 2009). He also spent two years (2004-2006) as a research intern at Télémcom Paris, France. Since 2011 he is Associate Professor at Universidad ORT Uruguay, working in the Mathematical Analysis in Telecommunications and Energy (MATE) research group. In 2012, Dr. Ferragut was awarded the annual Uruguayan Engineering Academy prize for the best doctoral thesis in Electrical Engineering. Since 2017, he is also Editor for IEEE/ACM Transactions on Networking. His research interests are stochastic processes and queueing theory applied to the mathematical modeling of networks.

Fernando Paganini (M’90–SM’05–F’14) received his Electrical Engineering and Mathematics degrees from Universidad de la Republica, Montevideo, Uruguay, in 1990, and his M.S. and PhD degrees in Electrical Engineering from the California Institute of Technology, Pasadena, in 1992 and 1996 respectively. His PhD thesis received the 1996 Wilts Prize and the 1996 Clauser Prize at Caltech. From 1996 to 1997 he was a postdoctoral associate at MIT. Between 1997 and 2005 he was on the faculty of the Electrical Engineering Department at UCLA, reaching the rank of Associate Professor. Since 2005 he is Professor of Electrical and Telecommunications Engineering at Universidad ORT Uruguay. Dr. Paganini has received the 1995 O. Hugo Schuck Best Paper Award, the 1999 NSF CAREER Award, the 1999 Packard Fellowship, the 2004 George S. Axelby Best Paper Award, and the 2010 Elsevier Scopus Prize. He is a member of the Uruguayan National Academies, both in Science and in Engineering, and a member of the Latin American Academy of Sciences. He is a Fellow of the IEEE. His research interests are control and networks.

Martin Zeballos obtained his Mechanical Engineer Degree in 2011 and is pursuing a Master’s Degree in Energy Engineering both from Universidad de la Republica (UDELAR), Uruguay. He joined the Mathematical Analysis in Telecommunications and Energy (MATE) Group of Universidad ORT del Uruguay in 2017 where he works as a research assistant in optimization of EVs charging systems. From 2007 to 2011, he was a research and teaching assistant in the Instituto de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA), UDELAR where besides lecturing to undergraduate, participated in research projects on wind power and wind engineering. He also held teaching position at the Universidad del Trabajo del Uruguay (UTU) for the Mechanical Technologist career (2011-2013). Since 2011, Ing. Zeballos works in the Energy and Industrial Division of the engineering services company CSI Ingenieros (Uruguay) where he has participated in the development of various industrial and energy projects, especially in renewable energy systems and energy efficiency studies.