Notes

January 12, 2025

Contents

1 On-Road EV Charging Management

1.1 Problem Statement

We aim to schedule a fleet of for-hire EVs to charge during periods of low charging price. The goal is twofold: first, to reduce electricity costs for the fleet operator, and second, to flatten the net load profile by distributing charging more evenly throughout the day. This helps avoid the scenario where all EVs charge at night, thereby easing pressure on the power grid.

1.2 MDP Model

We consider a for-hire EV fleet of size N, with a total number of M chargers available for charging in the region ($N \gg M$), and a scheduling horizon of T. The charging management problem for this fleet can be modeled as a discrete-time MDP with finite time steps $t = 1, 2, \cdots, T$. Each EV's state at time t is $\mathbf{s}_t^i$ and we denote the aggregate state as $\bar{\mathbf{s}}_t = \{\mathbf{s}_t^i\}_{i=1}^N$. At each time t, the dispatcher observes the system state $\bar{\mathbf{s}}_t$ and takes an aggregate action $\bar{\mathbf{a}}_t = \{a_t^i\}_{i=1}^N$ for all agents, deciding whether each agent should begin charging or continue charging.

We assume both time-varying charging prices and time-varying estimated payments that a driver could receive if assigned a ride order. To prevent any EV from consistently charging when costs are low or always being assigned ride orders when payments are high, we apply max-min fairness. This ensures that no EV's cumulative earnings under the aggregate policy fall too low. After the action is taken, the system provides a reward r_t, which includes both charging benefits (low or even negative charging costs) and ride order payments, and transitions to the next state $\bar{\mathbf{s}}_{t+1}$.

1) *State space*: The state of each agent (EV) at time t is represented by the tuple:

$$\mathbf{s}_t^i = (\alpha_t^i, \beta_t^i, \theta_t^i) \in \mathbb{Z}_{\geq 0} \times \{0, 1\} \times \{0, 1, 2, \ldots, 100\}.$$

The variable α_t^i denotes the remaining trip length if vehicle i has been assigned an order, and $\alpha_t^i = 0$ if no order is assigned. The variable β_t^i denotes the charging status. If assigned a ride order, the ride trip may span multiple time steps, but each charging decision only applies to a single time step. Therefore, if an EV chooses to continue charging in the next time step, it must make a new charging decision. While this may seem inconvenient, it provides the EV with the flexibility to leave at any time. Moreover, since decisions are made sequentially by the centralized dispatcher, the process is more manageable than it might initially seem. Note that the pair (α_t^i, β_t^i) represents the operational status of vehicle i:

$$\begin{array}{rc} \alpha_t^i = 0, \beta_t^i = 0: & \text{Vehicle } i \text{ is idle} \\ \alpha_t^i > 0, \beta_t^i = 0: & \text{Vehicle } i \text{ is on a ride} \\ \alpha_t^i = 0, \beta_t^i > 0: & \text{Vehicle } i \text{ is charging} \end{array}.$$

The variable θ_t^i represents the battery state of charge (SoC) (e.g., $\theta_t^i = 0.1$ represents a 10% SoC).

2) *Action space*: The action for each EV is $a_t^i \in \{0, 1\}$, where $a_t^i = 1$ indicates that EV i is scheduled to charge, and $a_t^i = 0$ means that EV i is unplugged or remains idle, ready to take ride orders.

3) *State evolution*: When a vehicle is idle, i.e., takes action $a_t^i = 0$ in the state $\alpha_t^i = 0, \beta_t^i = 0$, there is a probability ρ_t that it will be assigned an order. If assigned, the vehicle starts a trip with a random duration. Let $\tau_{\theta_t^i}$ represent this random duration, which follows the distribution:

   - With probability ρ_t, $\tau_{\theta_t^i}$ is drawn from a discrete distribution $f(\theta_t^i)$, dependent on the vehicle's current SoC θ_t^i, i.e., $\tau_{\theta_t^i} \sim f(\theta_t^i)$.
   - With probability $1 - \rho_t$, $\tau_{\theta_t^i} = 0$, meaning no order is assigned.

   The vehicle earns $w_t \cdot \tau_{\theta_t^i}$, where w_t is the estimated earnings per unit of ride time at time t.

   If EV i is on a ride and the remaining ride time is greater than 1 step, its action space is restricted to $\mathcal{A} = \{0\}$, meaning it must remain idle (i.e., not charge). If the vehicle is charging, it can return to idle status by taking the action $a_t^i = 0$ to unplug and stop charging.

   In summary, each EV i with state $(\alpha_t^i, \beta_t^i, \theta_t^i)$ transitions to the following states under action a_t^i:

$\boldsymbol{\alpha_t^i}$	$\boldsymbol{\beta_t^i}$	$\boldsymbol{a_t^i}$	**Next State Tuple**	**Reward**
≥ 2	0	0	$(\alpha_t^i - 1, 0, [\theta_t^i - \delta_i^-]_0^1)$	0
1	0	0	$(\tau_{\theta_t^i}, 0, [\theta_t^i - \delta_i^-]_0^1)$	$w_t \tau_{\theta_t^i}$
1	0	1	$(0, 1, [\theta_t^i - \delta_i^-]_0^1)$	$-(h_t + p_t \min(\delta_t^+, 1 - \theta_{t+1}^i) C_i^{\max})$
0	0	0	$(\tau_{\theta_t^i}, 0, [\theta_t^i - \delta_i^-]_0^1)$	$w_t \tau_{\theta_t^i}$
0	0	1	$(0, 1, [\theta_t^i - \delta_i^-]_0^1)$	$-(h_t + p_t \min(\delta_t^+, 1 - \theta_{t+1}^i) C_i^{\max})$
0	1	0	$(\tau_{\theta_t^i}, 0, [\theta_t^i + \delta_i^+]_0^1)$	$w_t \tau_{\theta_t^i}$
0	1	1	$(0, 1, [\theta_t^i + \delta_i^+]_0^1)$	$-p_t \min(\delta_t^+, 1 - \theta_{t+1}^i) C_i^{\max}$

Table 1: Transition Dynamics

   where δ_i^+ and δ_i^- are the charging and discharging percentages of EV i' SoC, respectively, $C_i^{\max}$ is the battery capacity of EV i, and θ_{t+1}^i is the battery SoC in next state. The notation $[x]_0^1$ represents the truncation of x to the range $[0, 1]$, defined as:

$$[x]_0^1 = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{if } 0 \leq x \leq 1, \\ 1 & \text{if } x > 1. \end{cases}$$

Note that the dynamics of β and θ can be written explicitly as:

$$\beta_{t+1}^i = a_t^i$$

$$\theta_{t+1}^i = \left[\theta_t^i + \beta_t^i \delta_i^+ - \beta_t^i \delta_i^-\right]_0^1$$

4) *Reward*: At time t, the reward received by each EV i with action a_t^i in state $\mathbf{s}_t^i = (\alpha_t^i, \beta_t^i, \theta_t^i)$ is:

$$r_t^i = w_t \cdot \tau_{\theta_t^i} \cdot (1 - a_t^i)\mathbb{1}(\alpha_t^i \geq 2) - h_t \cdot a_t^i \cdot (1 - \beta_t^i) - a_t^i \cdot p_t \cdot \min(\delta_t^+, 1 - \theta_{t+1}^i) C_i^{\max}$$

where $\mathbb{1}(\cdot)$ is the indicator function, h_t is the connection fee for plugging in to charge, and p_t is the per kWh charging fare. The first term represents the expected payment EV i could receive for remaining idle, the second term accounts for the connection fee when first connecting to a charger (if the EV connects once and charges for multiple consecutive time steps, it only pays the connection fee once), and the last term represents the cost for charging for one time step.

5) *Objective function*: Given an initial system state $\bar{\mathbf{s}}_0$ and a policy π that generates a sequence of actions $\bar{\mathbf{a}}_t$, $t = 0, 1, \cdots, T$, we aim to identify a policy that maximizes the expected accumulated rewards starting from $\bar{\mathbf{s}}_0$:

$$\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T-1} \sum_{i=1}^{N} r_i^t \mid \bar{\mathbf{s}}_0\right] + r_T \tag{1}$$

where r_T represents the reward for the terminal state. Currently, we set $r_T = 0$. In the future, we plan to incorporate max-min fairness by defining r_T as the minimum cumulative reward across all EVs:

$$r_T = \min_i \mathbb{E}\left[\sum_{t=0}^{T-1} r_i^t\right].$$

6) *Constraints*: First, Each EV's battery SoC must remain within the range $[0, 1]$, representing 0% to 100% of the maximum capacity C_i. This constraint is expressed as:

$$0 \leq \theta_t^i \leq 1, \quad \forall i, t$$

This condition is inherently satisfied, as the next state's SoC is constrained within $[0, 1]$ in Table 1.

Second, the total number of EVs requesting to charge at any time t must not exceed the charging station capacity M:

$$\sum_{i=1}^{N} a_t^i \leq M, \quad \forall t$$

To simplify the problem, this constraint can be relaxed and incorporated into the objective function using a penalty method. Let λ denote the penalty parameter. The modified objective function from (1) becomes:

$$\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T-1} \sum_{i=1}^{N} r_i^t \middle| \bar{\mathbf{s}}_0\right] + r_T - \lambda \sum_{t=0}^{T-1} \left|\sum_{i=1}^{N} a_t^i - M\right|$$

Here, the penalty term $\lambda \sum_{t=0}^{T-1} \left|\sum_{i=1}^{N} a_t^i - M\right|$ discourages violations of the charging capacity constraint.