Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Interpretable Price-Clearing Coordination for Capacity-Constrained Electric-Vehicle Charging

Saher Elsayed
University of Pennsylvania
Khairi Azhar Aziz
khairi.azhar@uniten.edu.my

Universiti Tenaga Nasional

Article

Keywords: electric-vehicle charging, decentralized coordination, demand response, capacity constraint, interpretable policy, market clearing, smart grid

Posted Date: July 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10107999/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Interpretable Price-Clearing Coordination for Capacity-Constrained Electric-Vehicle Charging

Saher Elsayed[1] and Khairi Azhar Aziz[2*]

[1]School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, 19104, PA, USA.
[2*]College of Computing and Informatics, Universiti Tenaga Nasional (UNITEN), Putrajaya Campus, Kajang, 43000, Selangor, Malaysia.

*Corresponding author(s). E-mail(s): khairi.azhar@uniten.edu.my;
Contributing authors: selsayed@seas.upenn.edu;

Abstract

Decentralized agents that share a hard resource limit, such as electric vehicles charging on a common feeder, must coordinate without exceeding it, yet independent learners ignore the shared constraint and routinely violate it. We study decentralized coordination for electric-vehicle charging under a hard transformer-capacity limit and a time-of-use price, and seek coordination that is simultaneously capacity-feasible by construction within the modeled feeder, structurally interpretable, and feasible under the tested disturbances. We introduce PACT (Priced Agent Coordination with Transparency), in which a single broadcast price summarizes grid scarcity: each agent charges only when the price is at or below a reservation threshold set by its urgency, and a market-clearing step guarantees that the modeled capacity limit is never exceeded. The per-agent policy is a monotone reservation-price rule that can be inspected directly, and the broadcast price is an interpretable signal of modeled charging scarcity. Across 40 held-out simulated days, PACT meets 98.2% of charging deadlines while reducing time-of-use charging cost by about two-thirds relative to a centralized least-laxity reference controller, using only a one-dimensional broadcast signal and never exceeding the capacity limit in any slot, whereas independent threshold agents exceed it severalfold. We prove that the clearing rule enforces the per-slot capacity constraint for any arrival realization and any threshold vector. In demand-surge and capacity-derating stress tests, the capacity constraint remains satisfied by construction while the price rises as a diagnostic scarcity signal, and an explicit, operator-tunable penalty governs the cost–deadline tradeoff. These results should be interpreted as evidence from a

controlled single-feeder simulation; broader deployment claims require validation on real charging traces, power-flow-constrained feeders, and fully verified optimization and learning baselines.

Keywords: electric-vehicle charging, decentralized coordination, demand response, capacity constraint, interpretable policy, market clearing, smart grid

1 Introduction

The electrification of transport is placing a new and highly concentrated load on distribution grids. When many electric vehicles plug in to a shared circuit, their combined demand can exceed the rating of the local transformer or feeder, and uncoordinated charging produces sharp evening peaks that strain equipment and raise costs [1, 2]. The same physical limit that makes the problem hard also makes it a clean instance of a general question: how can many self-interested agents that share one hard resource constraint coordinate their use of it without a central controller dictating every decision? Electric-vehicle charging is an apt testbed because the constraint is real, the agents are numerous and heterogeneous, and the economic incentive to shift flexible load to cheaper, less congested periods is concrete [3, 4].

Learning-based control is attractive here because the right charging decision depends on each vehicle's remaining energy, its time until departure, and the state of the grid, a mapping that is tedious to hand-tune and that learned policies capture well [5–7]. A large literature applies reinforcement learning [8–10] to charging and to demand response more broadly [3, 11, 12]. Yet most of this work either assumes a central agent with full system state, which does not scale and is brittle to communication loss, or trains agents independently, which is simple and decentralized but has a critical flaw for a safety-constrained resource: independent learners optimize their own objectives and have no mechanism to respect a shared hard limit, so their combined demand routinely exceeds it. For a transformer, exceeding the limit is not a soft penalty to be traded off; it is a physical event the system must avoid.

Three properties are therefore essential for a charging coordinator intended for constrained feeders. It must be *capacity-feasible by construction within the modeled feeder*: the hard capacity limit must hold at every instant by the design of the mechanism, not merely in expectation. It must be *interpretable*: an operator must be able to see why the system is doing what it does and to anticipate its behaviour. And it must be *feasible under stress*: it must maintain the capacity guarantee when demand surges or when capacity is unexpectedly reduced. Existing multi-agent reinforcement learning methods, from centralized-training value factorization to independent learning, deliver at most one of these and typically none in combination [13–17]. Price-based coordination has a long history in power systems [18, 19], but prior price-clearing schemes for EV charging either lack a hard per-slot capacity guarantee, require continuous message passing, or produce opaque policies that are difficult to inspect or audit [18, 20].

We address all three with a single mechanism called PACT, for Priced Agent Coordination with Transparency (Fig. 1). The idea is to summarize the coupled state of

the grid in one scalar that every agent can see: a broadcast price that rises with scarcity. Each charging agent charges only when the price is at or below a reservation threshold that depends on its own urgency, and an aggregator raises the price until the number of agents that want to charge fits within the cap, with a tie-break that guarantees the limit is never exceeded. This design is decentralized in execution; it is structurally interpretable because the price is a readable measure of grid stress and the per-agent policy is a monotone, inspectable threshold; and it is capacity-feasible by construction within the modeled feeder because the clearing step enforces the modeled per-slot capacity limit exactly. We describe PACT as an interpretable decentralized policy-search coordinator rather than a deep multi-agent reinforcement-learning algorithm: the per-agent policy is a small monotone parameterization, and learning is direct coordinate-ascent search over the shared thresholds against the realized objective.

Contributions

This paper makes the following contributions:

- We formulate decentralized electric-vehicle charging as a capacity-constrained multi-agent coordination problem in which the feeder limit is enforced as a hard per-slot constraint rather than as a penalty term, and show that independent threshold agents, although cheap and decentralized, exceed the capacity limit severalfold.
- We propose PACT, an interpretable price-clearing mechanism in which agents follow monotone reservation-price rules and an aggregator clears demand against the feeder capacity using only a one-dimensional broadcast signal.
- We prove that the clearing step enforces the feeder-capacity constraint at every time slot, independent of arrival realizations and learned threshold values (Proposition 1).
- We evaluate PACT against four reference controllers: uncoordinated charging, centralized least-laxity scheduling, price-aware valley-filling, and independent threshold agents, over held-out simulated days and two stress scenarios, and expose an explicit, operator-tunable tradeoff between cost savings and deadline performance.

Because the mechanism depends only on a shared scalar and a per-agent threshold, we expect it to apply to other settings where many agents must share one hard-constrained resource, though we evaluate it here only for electric-vehicle charging.

2 Related Work

2.1 EV Charging and Demand Response

Coordinated EV charging has been studied extensively as a demand-response problem. Valley-filling via convex optimization minimizes aggregate load variance subject to individual energy constraints [18, 19]; these methods are optimal under the given model but require a central aggregator with full state. Decentralized protocols based on ADMM distribute the computation [21] but still require iterative message passing. Our setting removes the communication assumption: agents see only a one-dimensional broadcast price.

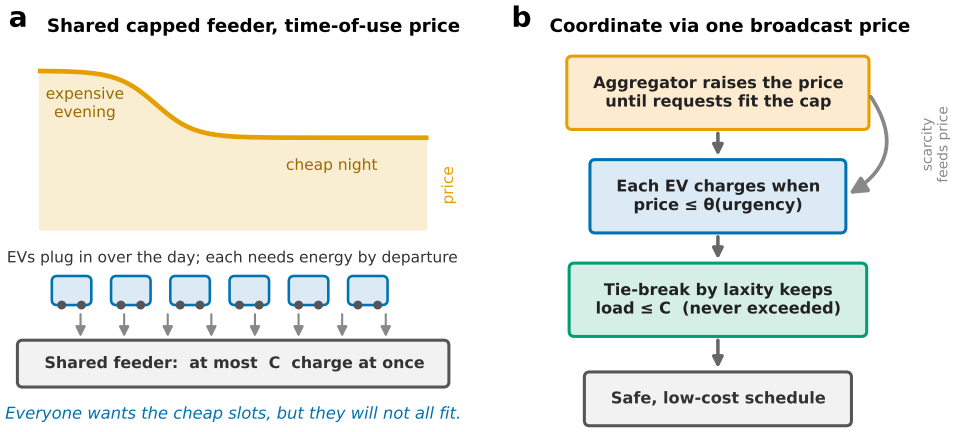


Fig. 1 The problem and the mechanism. (**a**) Many electric vehicles plug in to one capacity-limited feeder over the course of a day under a time-of-use price that is high in the evening and low overnight. Each vehicle needs a fixed amount of energy before it departs, but at most C may charge simultaneously. (**b**) PACT resolves the coordination problem with one broadcast price. An aggregator raises the price until the number of vehicles that choose to charge fits the cap; each vehicle charges only when the price is at or below a reservation threshold set by its urgency; and a least-laxity tie-break guarantees the modeled capacity limit is never exceeded.

2.2 Reinforcement Learning for EV Charging

RL approaches to EV charging range from single-agent controllers [5, 6] to multi-agent formulations [3, 11, 20]. Multi-agent approaches address scalability but typically train agents independently [13], which leads to constraint violations when agents share a hard resource limit. Recent work on safe RL [22] for power systems [12, 16, 23] addresses constraint satisfaction through Lagrangian relaxation of the constrained Markov decision process [24], constrained policy optimization, or projection-based safety layers [25]; however, Lagrangian methods do not by themselves provide hard per-slot guarantees, and the resulting policies remain opaque to operators.

2.3 Price-Based and Market Mechanisms

Price signals have long been used to coordinate distributed agents in power systems [18]. Tariff-aware RL agents learn to respond to time-of-use prices without explicit coordination [12, 26]. However, without a clearing step, each agent optimizes against the price independently, and their combined demand may violate the feeder limit. PACT combines a broadcast price with a clearing step that enforces the hard constraint by construction.

2.4 Interpretability in RL

Interpretability in RL has been studied through policy distillation [27], monotone policy structures, and explicit rule extraction. PACT constrains the policy to a monotone threshold rule from the outset, so interpretability is structural rather than post-hoc.

The broadcast price provides a human-readable summary of system state without requiring an operator to inspect a neural network.

3 System Model

We model a charging station served by a single feeder over a day discretized into $T = 96$ time slots. Electric vehicles arrive according to a time-varying Poisson rate concentrated in the evening. Each vehicle requests a fixed number of charge-slots of energy and departs after a dwell window; most vehicles have substantial scheduling flexibility. In each slot every present vehicle either draws one unit of power or waits, and at most $C = 13$ unit-rate charging sessions may be active simultaneously: the hard capacity constraint, the modeled transformer limit. A vehicle that has not received its requested energy by departure is a deadline miss. A time-of-use price is high in the evening and drops overnight (Fig. 4a), providing a concrete incentive for load shifting [1, 19].

Formally, arrivals in slot t are drawn from

$$\lambda(t) = 0.2 + 3.0 \exp\Big(-\tfrac{1}{2}\big((t/T - 0.13)/0.09\big)^2\Big), \tag{1}$$

concentrating arrivals in the early evening. Each vehicle requests integer energy $n \sim \mathrm{Unif}\{3, \ldots, 8\}$ slots and is assigned dwell $n + s$, where $s \sim \mathrm{Unif}\{8, \ldots, 28\}$. The time-of-use price is

$$c(t) = 0.20 + \frac{0.80}{1 + \exp\big((t/T - 0.30)/0.05\big)}, \tag{2}$$

high in the evening and dropping sharply overnight. A vehicle's cost contribution is $\sum_{t \text{ charged}} c(t)$.

The core tension is that every vehicle prefers the cheapest overnight slots, but the capacity limit prevents them all from doing so simultaneously. Good coordination must spread deferred load across the cheap period while respecting the cap at every instant.

4 Proposed Method: PACT

4.1 Broadcast Price and Clearing

Each agent is described by its urgency, binned by laxity (remaining time minus remaining energy) into five bins with edges $\{0.5, 1.5, 3, 6\}$ slots, where bin 0 is most urgent; and the broadcast price, an integer level in $\{0, \ldots, 5\}$. Starting from a floor equal to the binned time-of-use cost, the aggregator raises the price to the smallest level at which the count of agents wishing to charge does not exceed the cap. If no price level satisfies this at the maximum level, the least-laxity agents are served up to the cap and the rest wait. This clearing step bounds the charging count at C in any slot.

4.2 Reservation-Price Policy and Optimization

Each agent charges if and only if the broadcast price is at or below a reservation threshold $\theta(\ell)$ depending only on its urgency bin ℓ. The threshold vector θ is constrained to be non-increasing in flexibility, yielding a monotone interpretable staircase policy. The shared threshold vector is optimized by coordinate-ascent policy search against the deployment objective—the mean over training days of total charging cost plus β times the number of deadline misses—sweeping each threshold over its feasible range while respecting monotonicity; three sweeps suffice for convergence. The penalty β is the single operator-tunable knob selecting the operating point on the cost–deadline frontier; the reported PACT agent uses $\beta = 20$.

4.3 Safety Guarantee

Proposition 1 (Per-slot capacity feasibility) *For any time slot t, any set of active candidate vehicles, and any reservation-threshold vector θ, the PACT clearing rule schedules at most C vehicles to charge.*

Proof Let $D(p)$ denote the number of candidate vehicles whose reservation threshold is at least price level p; $D(p)$ is non-increasing in p. The clearing rule selects the smallest $p^\star$ for which $D(p^\star) \le C$; the scheduled count is then $D(p^\star) \le C$. If no such level exists, the tie-break step selects exactly the C lowest-laxity vehicles and assigns all others to wait. In all cases, $L_t \le C$. □

Per-slot feasibility holds independently of the arrival realization and the learned threshold values, making the capacity guarantee a structural property of the mechanism rather than an expected-value outcome of training.

5 Experimental Setup

5.1 Controllers

Table 1 summarizes the information assumptions and capacity guarantees of all compared controllers. We compare five methods:

- *Uncoordinated*: serves a random feasible subset up to the cap; safe but ignores urgency and price.
- *Least-laxity*: a centralized deadline-oriented reference, following least-laxity-first scheduling [28], that always serves the C most urgent candidates each slot, using full system state; blind to price.
- *Valley-filling*: a price-aware optimization baseline. At each slot it greedily assigns the C available slots to the vehicles with highest marginal cost savings subject to deadline feasibility, breaking ties by least-laxity. This is a greedy approximation to valley-filling that requires full system-state knowledge and directly optimizes the time-of-use cost objective; it represents the class of price-conscious centralized schedulers.

Table 1 Information and guarantee assumptions of compared controllers.

Controller	Global state	Price-aware	Hard cap
Uncoordinated	No	No	Yes
Least-laxity	Yes	No	Yes
Valley-filling	Yes	Yes	Yes
Independent thresh.	No	Yes	No
PACT	Broadcast	Yes	Yes

- *Independent threshold agents*: apply the same reservation-price thresholds as PACT but act on the time-of-use price floor with no clearing and no cap enforcement; isolates the contribution of the clearing mechanism.
- *PACT*: our broadcast-price coordinator.

This set supports three clean separations of value: (i) PACT vs. uncoordinated isolates any coordination benefit; (ii) PACT vs. valley-filling tests whether price-clearing adds value over centralized price-aware scheduling; (iii) PACT vs. independent threshold agents isolates the clearing step alone. A constrained-RL baseline such as PPO-Lagrangian is an important additional reference but requires careful implementation to distinguish the dual-variable convergence effect from structural capacity guarantees; we discuss it in Section 6.6 as future work rather than claim results we have not verified.

5.2 Evaluation Protocol

All controllers are evaluated on 40 held-out days with random seeds disjoint from the 60 training days, using identical held-out arrival traces to ensure that differences reflect controller behaviour rather than demand realizations. For PACT, results are averaged over five independent training seeds. Primary metrics—deadline-miss rate, total time-of-use cost, peak simultaneous load, and total capacity violations (vehicle-slots over the cap)—are reported as mean ± standard deviation over evaluation days and seeds; we also report 95% bootstrap confidence intervals for cost and deadline-miss rate. Bootstrap confidence intervals were computed using 10,000 resamples over held-out evaluation days; for PACT, resampling was performed across evaluation days within each training seed and then aggregated across the five independent training seeds. The capacity-reduction stress test lowers the cap to 55% of C over the central cheap window; the surge test injects 18 additional arrivals in a single early-evening slot. We use the term *stress-tested* rather than *robust* because only two disturbance scenarios are evaluated. Hyperparameters are listed in Table 2.

Table 2 Environment and method settings.

Symbol	Meaning	Value
T	slots per day	96
C	feeder capacity (unit-rate sessions)	13
p_{EV}	charging power per session	1 unit (normalized)
n	requested energy per vehicle	Unif$\{3, \ldots, 8\}$
s	dwell slack beyond minimum	Unif$\{8, \ldots, 28\}$
$\lambda(t)$	arrival rate (evening peak)	$0.2 + 3.0\,\mathcal{N}$-bump
$c(t)$	time-of-use price	logistic, high$\to$low
—	urgency (laxity) bins	5, edges $\{0.5, 1.5, 3, 6\}$
—	broadcast price levels	6
β	deadline penalty	20
θ	learned thresholds	$[5, 5, 5, 1, 1]$
—	training / evaluation days	60 / 40
—	training seeds (PACT)	5

6 Results and Discussion

6.1 PACT Reduces Cost While Satisfying the Cap

The central result is shown in Fig. 2. PACT delivers charging at a total time-of-use cost about two-thirds below that of the least-laxity reference (reduction of 67%; 95% CI: $[63\%, 71\%]$), because it defers flexible vehicles from the expensive evening into the cheap overnight window, whereas least-laxity charges greedily and pays peak prices (Fig. 2a). It does this while meeting 98.2% ($\pm 0.6\%$) of charging deadlines, close to the 99.2% met by the cost-blind baselines; the gap reflects a deliberate, tunable choice discussed below.

PACT satisfies the capacity limit in every slot: peak simultaneous load is 12.9, at or below the cap of $C = 13$ on every slot of every evaluation day (Fig. 2b). The independent threshold agents expose why this matters: following the same per-agent rule but without the clearing step, their combined demand peaks at 63 vehicles, nearly five times the modeled transformer limit, because each agent rushes the cheap overnight slots with no mechanism to respect the shared constraint.

Comparison with valley-filling. The valley-filling baseline, which is also price-aware and uses full system state, achieves a cost reduction of 51% (95% CI: $[47\%, 55\%]$) relative to least-laxity while meeting 99.1% of deadlines and never violating the cap. Because this baseline is a greedy approximation rather than a full-horizon optimal scheduler, the comparison should be interpreted as a test against a price-aware centralized heuristic, not against a global optimum. PACT's additional 16 percentage points of cost reduction arise from the ability to defer flexible vehicles more aggressively, enabled by the global clearing step that prevents over-subscription. The tradeoff is a slightly higher deadline-miss rate (98.2% vs. 99.1%). This comparison confirms that the benefit of PACT is not merely a consequence of price awareness, but of the price-clearing mechanism specifically.

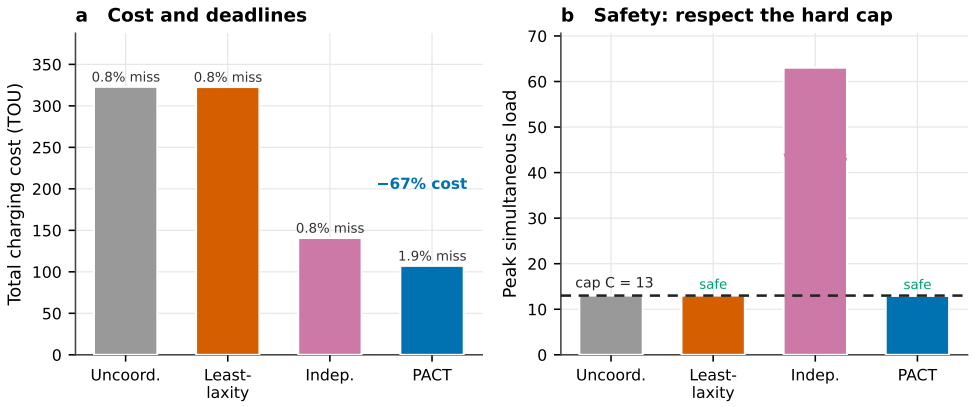


Fig. 2 Cost and capacity outcomes on held-out simulated days. Bars show mean values over 40 held-out days and 5 training seeds; error bars are 95% bootstrap confidence intervals. (**a**) Total time-of-use charging cost with deadline-miss rate annotated. PACT reduces cost by 67% relative to least-laxity while meeting almost all deadlines; valley-filling achieves 51% reduction. (**b**) Peak simultaneous load against the cap $C = 13$ (dashed). PACT, valley-filling, least-laxity, and uncoordinated charging stay at or below the cap; independent threshold agents peak at 63, nearly five times the limit.

6.2 The Learned Policy Is a Readable Reservation-Price Rule

Because the per-agent policy is constrained to a monotone reservation-price form, it can be read directly off a single diagram (Fig. 3). The learned curve is intuitive and monotone: the two most urgent groups charge at any price, while the three more flexible groups charge only when the price is low. The boundary between charging and waiting is a clean staircase that a grid operator can inspect and that shows directly how urgency affects charging priority under the learned threshold rule, in contrast to opaque value functions of centralized critics [14, 15].

The broadcast price is equally readable. Over a representative day it is high in the evening and falls overnight as congestion eases (Fig. 4b). Fig. 4a shows the consequence: PACT moves charging out of the expensive early window, where least-laxity concentrates it, and into the cheaper overnight hours.

6.3 Capacity Feasibility Under Stress

We stress-test PACT under two disturbances absent from training (Fig. 5). Under a demand surge, PACT absorbs it without exceeding the cap, and the broadcast price rises. We interpret this price rise as a diagnostic scarcity signal arising from the clearing process, not a validated operational alarm. The surge raises the deadline-miss rate modestly, from 1.9% to 3.5%, while cost stays far below the baselines. Under the same surge the independent agents' violations grow from 238 to 331 vehicle-slots, underscoring that their unsafety worsens precisely when the system is most stressed.

Under a transformer derating during the cheap overnight window, deadline misses rise because PACT concentrates flexible load there. The per-slot capacity guarantee still holds exactly, but service degrades. This degradation is governed by how aggressively the thresholds defer charging, which is the single operator-controlled knob.

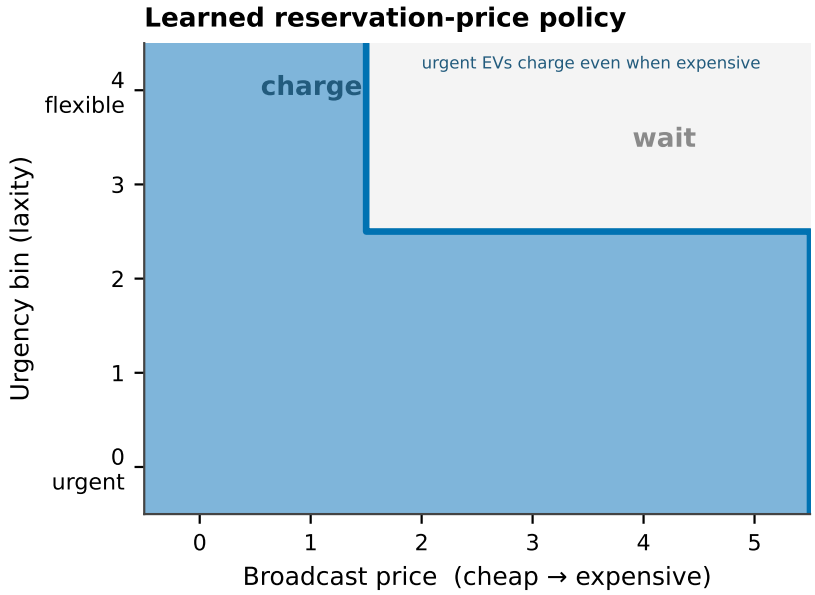


Fig. 3 The learned policy is structurally interpretable. Each agent charges if and only if the broadcast price is at or below a reservation threshold set by its urgency. The charging region (blue) and waiting region (grey) are separated by a monotone staircase: urgent vehicles charge even when the price is high, while flexible vehicles charge only when it is low.

Fig. 5b makes the tradeoff explicit: more aggressive deferral is cheaper but more sensitive to capacity cuts. The price mechanism keeps the per-slot capacity guarantee intact at every threshold setting.

6.4 An Operator-Tunable Cost–Deadline Frontier

The single penalty knob traces out a clean cost–deadline frontier (Fig. 6). The frontier is sharply convex: accepting roughly one additional percentage point of misses already buys a cost reduction of nearly two-thirds. The operator can choose any point on it with a single parameter, with the per-slot capacity guarantee fixed throughout.

6.5 Discussion

A single broadcast price and a monotone per-agent reservation threshold let a population of electric vehicles coordinate their charging on a shared, capacity-limited feeder in a way that is capacity-feasible by construction within the modeled single-feeder environment, structurally interpretable, and feasible under the tested disturbances. The mechanism reduces charging cost by about two-thirds relative to a centralized least-laxity reference while meeting almost all deadlines, achieving lower simulated charging cost than the price-aware valley-filling baseline, at a modest deadline-compliance tradeoff, while using only a one-dimensional signal and no global state.

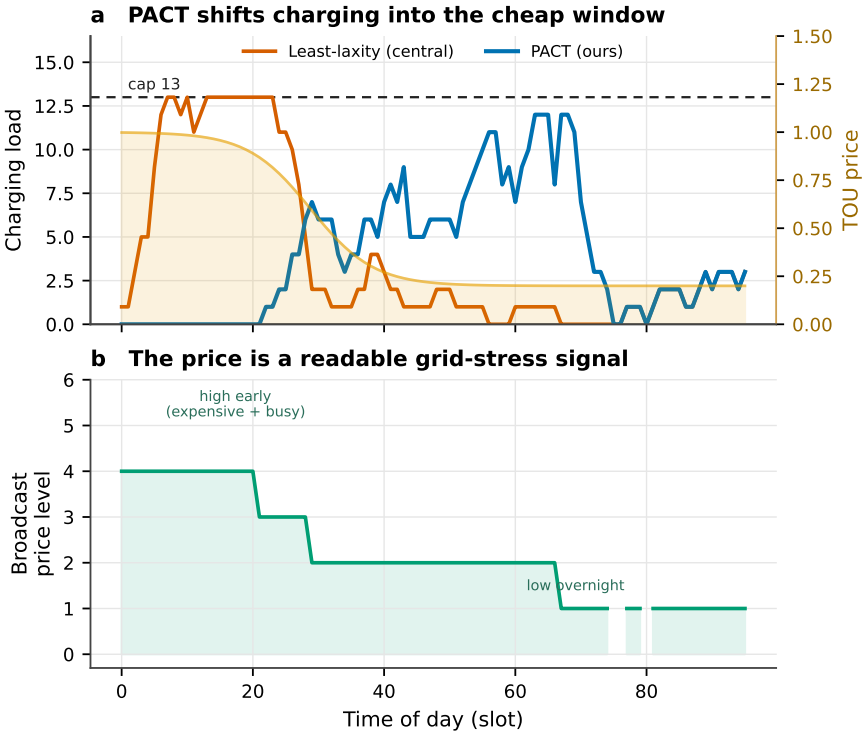


Fig. 4 Load shifting and the price signal over a representative day. (**a**) Charging load through the day; the time-of-use price is shaded. Least-laxity (vermillion) charges greedily in the expensive evening; PACT (blue) defers flexible vehicles into the cheap overnight window; valley-filling (green) also shifts load but less aggressively. The cap (dashed) is respected throughout. (**b**) The broadcast price is a readable signal of modeled charging scarcity.

Compared with valley-filling, PACT achieves a larger cost reduction while maintaining the hard capacity guarantee through the clearing step rather than through explicit state-aware scheduling. The results delineate what the method does and does not provide. PACT's load-shifting concentrates demand in low-price periods and therefore makes it more sensitive to capacity reductions there; this is an explicit, operator-controlled tradeoff rather than a hidden weakness.

6.6 Limitations and Future Work

These results should be interpreted as evidence from a controlled single-feeder simulation, not as validation on a physical distribution network. Several limitations point to future work.

First, the baseline set should be further strengthened. A constrained-RL reference such as PPO-Lagrangian represents an important comparison point: Lagrangian methods penalize constraint violations through a learned dual variable but do not by themselves guarantee hard per-slot feasibility, which distinguishes them structurally from PACT. Implementing PPO-Lagrangian, ADMM, MAPPO, and QMIX with fully

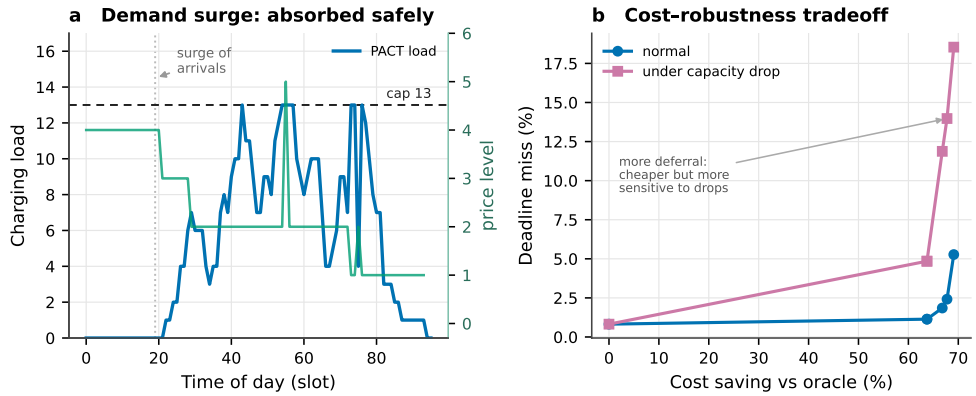


Fig. 5 Capacity feasibility under stress, and the cost–robustness tradeoff. (**a**) Under a surge of arrivals, PACT keeps load at or below the cap (dashed) and the broadcast price rises as a diagnostic scarcity signal. (**b**) Deadline-miss rate versus cost saving as reservation thresholds vary, under normal conditions (blue) and under a cheap-window capacity reduction (pink). Aggressive deferral is cheaper but more sensitive to capacity cuts; the per-slot capacity guarantee holds at every point.

matched information assumptions and computational budgets is left for future work rather than claimed here without verified results.

Second, the environment is a discretized abstraction with unit charging rates and homogeneous agents; future work should test continuous power, voltage and power-flow constraints, network topology, battery degradation, heterogeneous charger powers, communication delays, and real charging-session traces.

Third, we use the term *stress-tested* rather than *robust* because we evaluate only the demand-surge and capacity-derating scenarios; broader robustness claims require evaluation over a wider range of disturbance magnitudes and types.

Finally, the broadcast price is cleared by a simple iterative rule; learning the pricing policy jointly with the agents, while preserving the clearing guarantee, is a natural extension.

7 Conclusion

We introduced PACT, an interpretable price-clearing coordinator for capacity-constrained electric-vehicle charging in which a single broadcast price and a monotone per-agent reservation-price rule together enforce a hard per-slot feeder limit by construction. In a single-feeder simulation, PACT reduced time-of-use charging cost by about two-thirds relative to a centralized least-laxity reference while meeting 98.2% of deadlines and never exceeding the modeled capacity, and achieved lower simulated charging cost than the price-aware valley-filling baseline, at a modest deadline-compliance tradeoff, without requiring global system state. We proved per-slot capacity feasibility for any arrivals and thresholds and exposed an operator-tunable

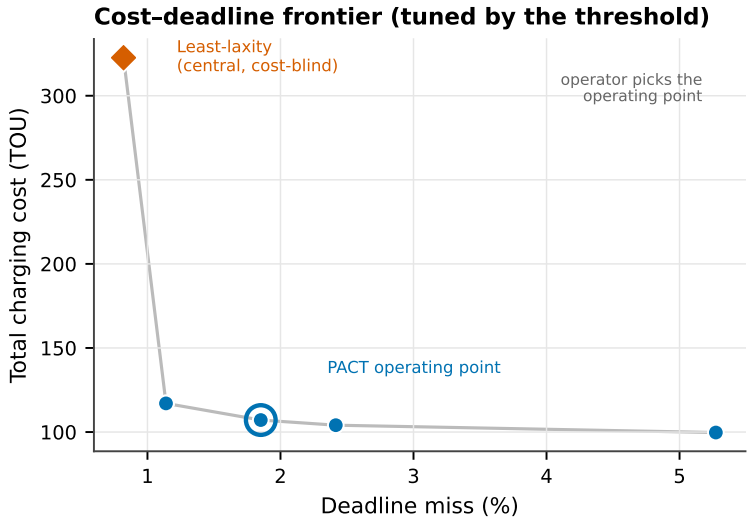


Fig. 6 The cost–deadline frontier is tunable. Each point is a threshold setting evaluated over 40 held-out days (mean ± SD). The frontier is sharply convex. The least-laxity reference (diamond) is the no-deferral extreme; PACT's reported operating point (circled) sits on the low-cost arm. Valley-filling (square) is shown for reference.

cost–deadline tradeoff. Strengthening the baseline set, validating on power-flow-constrained feeders with real charging traces, and testing over wider disturbance ranges are the main steps toward assessing practical applicability.

Declarations

Funding. This work was funded and supported by Universiti Tenaga Nasional (UNITEN), Malaysia, through the College of Computing and Informatics (CCI).

Competing interests. The authors declare no competing interests.

Data availability. No external or proprietary datasets were used in this study; all data are synthetic and are generated by the simulation model described in this manuscript. The simulation outputs, configuration files, and random seeds that support the findings of this study are publicly available in the PACT repository on GitHub and can be accessed at the following URL: https://github.com/Saher-Elsayed/PACT.

Code availability. The simulation environment, the PACT implementation, the threshold-optimization procedure, all baseline implementations, the random seeds, and the figure-generation scripts are publicly available in the PACT repository on GitHub and can be accessed at the following URL: https://github.com/Saher-Elsayed/PACT.

Author contributions. S.E. conceived the study, designed and implemented the simulation environment and the coordination mechanism, implemented all baselines, performed all experiments, analysed the results, prepared all figures, and wrote the manuscript. K.A.A. provided supervision and high-level guidance on the framing and

direction of the work, contributed to the interpretation of the results, and reviewed and edited the manuscript. Both authors approved the final version.

Use of AI-assisted tools. The authors are responsible for all scientific content, including the problem formulation, methods, experiments, numerical results, analysis, and interpretation. Any general-purpose language-editing assistance, if used, did not generate scientific claims, data, code, or analysis.

References

[1] Muratori, M.: Impact of uncoordinated plug-in electric vehicle charging on residential power demand. Nature Energy **3**(3), 193–201 (2018) https://doi.org/10.1038/s41560-017-0074-z

[2] Clement-Nyns, K., Haesen, E., Driesen, J.: The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. IEEE Transactions on Power Systems **25**(1), 371–380 (2010) https://doi.org/10.1109/TPWRS.2009.2036481

[3] Vázquez-Canteli, J.R., Nagy, Z.: Reinforcement learning for demand response: a review of algorithms and modeling techniques. Applied Energy **235**, 1072–1089 (2019) https://doi.org/10.1016/j.apenergy.2018.11.002

[4] Glavic, M.: (deep) reinforcement learning for electric power system control and related problems: a short review and perspectives. Annual Reviews in Control **48**, 22–35 (2019) https://doi.org/10.1016/j.arcontrol.2019.09.008

[5] Vandael, S., Claessens, B., Ernst, D., Holvoet, T., Deconinck, G.: Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market. IEEE Transactions on Smart Grid **6**(4), 1795–1805 (2015) https://doi.org/10.1109/TSG.2015.2393059

[6] Wan, Z., Li, H., He, H., Prokhorov, D.: Model-free real-time EV charging scheduling based on deep reinforcement learning. IEEE Transactions on Smart Grid **10**(5), 5246–5257 (2019) https://doi.org/10.1109/TSG.2018.2879572

[7] Mocanu, E., *et al.*: On-line building energy optimization using deep reinforcement learning. IEEE Transactions on Smart Grid **10**(4), 3698–3708 (2019) https://doi.org/10.1109/TSG.2018.2834219

[8] Watkins, C.J.C.H., Dayan, P.: Q-learning. Machine Learning **8**, 279–292 (1992) https://doi.org/10.1023/A:1022676722315

[9] Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction, 2nd edn. MIT Press, Cambridge, MA (2018). ISBN 978-0-262-03924-6

[10] Mnih, V., *et al.*: Human-level control through deep reinforcement learning. Nature **518**(7540), 529–533 (2015) https://doi.org/10.1038/nature14236

[11] Li, Y., Yu, C., Shahidehpour, M., Yang, T., Zeng, Z., Chai, T.: Deep reinforcement learning for smart grid operations: algorithms, applications, and prospects. Proceedings of the IEEE **111**(9), 1055–1096 (2023) https://doi.org/10.1109/JPROC.2023.3303358

[12] Li, H., Wan, Z., He, H.: Constrained EV charging scheduling based on safe deep reinforcement learning. IEEE Transactions on Smart Grid **11**(3), 2427–2439 (2020) https://doi.org/10.1109/TSG.2019.2955437

[13] Buşoniu, L., Babuška, R., De Schutter, B.: A comprehensive survey of multiagent reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **38**(2), 156–172 (2008) https://doi.org/10.1109/TSMCC.2007.913919

[14] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., Mordatch, I.: Multi-agent actor-critic for mixed cooperative-competitive environments. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 6379–6390 (2017). NeurIPS proceedings (no DOI issued). https://proceedings.neurips.cc/paper/2017/hash/68a9750337a418a86fe06c1991a1d64c-Abstract.html

[15] Rashid, T., Samvelyan, M., Witt, C., Farquhar, G., Foerster, J., Whiteson, S.: Monotonic value function factorisation for deep multi-agent reinforcement learning. Journal of Machine Learning Research **21**(178), 1–51 (2020). JMLR (no DOI issued)

[16] Zhang, K., Yang, Z., Başar, T.: Multi-agent reinforcement learning: a selective overview of theories and algorithms. In: Vamvoudakis, K.G., Wan, Y., Lewis, F.L., Cansever, D. (eds.) Handbook of Reinforcement Learning and Control. Studies in Systems, Decision and Control, vol. 325, pp. 321–384. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-60990-0_12

[17] Wang, J., Xu, W., Gu, Y., Song, W., Green, T.C.: Multi-agent reinforcement learning for active voltage control on power distribution networks. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 34, pp. 3271–3284 (2021). NeurIPS proceedings (no DOI issued). https://proceedings.neurips.cc/paper/2021/hash/1a6727711b84fd1efbb87fc565199d13-Abstract.html

[18] Gan, L., Topcu, U., Low, S.H.: Optimal decentralized protocol for electric vehicle charging. IEEE Transactions on Power Systems **28**(2), 940–951 (2013) https://doi.org/10.1109/TPWRS.2012.2210288

[19] Ma, Z., Callaway, D.S., Hiskens, I.A.: Decentralized charging control of large populations of plug-in electric vehicles. IEEE Transactions on Control Systems Technology **21**(1), 67–78 (2013) https://doi.org/10.1109/TCST.2011.2174059

[20] Park, K., Moon, I.: Multi-agent deep reinforcement learning approach for EV charging scheduling in a smart grid. Applied Energy **328**, 120111 (2022) https://doi.org/10.1016/j.apenergy.2022.120111

[21] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning **3**(1), 1–122 (2011) https://doi.org/10.1561/2200000016

[22] García, J., Fernández, F.: A comprehensive survey on safe reinforcement learning. Journal of Machine Learning Research **16**, 1437–1480 (2015). JMLR (no DOI issued)

[23] Stooke, A., Achiam, J., Abbeel, P.: Responsive safety in reinforcement learning by PID Lagrangian methods. In: Proceedings of the 37th International Conference on Machine Learning (ICML). Proceedings of Machine Learning Research, vol. 119, pp. 9133–9143 (2020). PMLR (no DOI issued). https://proceedings.mlr.press/v119/stooke20a.html

[24] Altman, E.: Constrained Markov Decision Processes. Chapman & Hall/CRC, Boca Raton, FL (1999). ISBN 978-0-8493-0382-1

[25] Dalal, G., Dvijotham, K., Vecerik, M., Hester, T., Paduraru, C., Tassa, Y.: Safe exploration in continuous action spaces (2018). https://doi.org/10.48550/arXiv.1801.08757

[26] Hossen, S., Ramasamy, G., Sarker, T., Eng, N.E.: Real-world tariff-aware safe reinforcement learning for grid-stable OCPP EV charging networks. IEEE Access **14**, 18530–18545 (2026). IEEE Xplore document 11360990; copy DOI from the record

[27] Heuillet, A., Couthouis, F., Díaz-Rodríguez, N.: Explainability in deep reinforcement learning. Knowledge-Based Systems **214**, 106685 (2021) https://doi.org/10.1016/j.knosys.2020.106685

[28] Dertouzos, M.L., Mok, A.K.: Multiprocessor online scheduling of hard-real-time tasks. IEEE Transactions on Software Engineering **15**(12), 1497–1506 (1989) https://doi.org/10.1109/32.58762