

UC Berkeley

UC Berkeley Previously Published Works

Title

Optimal Control of Differentially Private EV Charging: A Scalable Learning Approach Under Uncertainty

Permalink

<https://escholarship.org/uc/item/0zb2d9s8>

Journal

IEEE Internet of Things Journal, 12(24)

ISSN

2327-4662

Authors

Wu, Tong
Scaglione, Anna
Ravi, Nikhil
[et al.](#)

Publication Date

2025-01-15

DOI

10.1109/jiot.2025.3622650

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Optimal Control of Differentially Private EV Charging: A Scalable Learning Approach Under Uncertainty

D Arnold

February 2026

IEEE Internet of Things

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Optimal Control of Differentially Private EV Charging: A Scalable Learning Approach under Uncertainty

Tong Wu, Anna Scaglione, Nikhil Ravi, Sean Peisert, Daniel Arnold

Abstract—IoT-enabled electric vehicles (IoEVs) enable intelligent charging coordination that accounts for grid congestion. However, increased data exchange raises privacy concerns, as charging patterns can reveal sensitive driver behavior to grid operators. We propose a differentially private (DP) EV charging framework that enables coordinated control while protecting driver data with theoretical privacy guarantees. Nevertheless, integrating DP inevitably introduces uncertainty into the control strategy for EVs, which can lead to infeasible solutions. To tackle this challenge, we develop a feasible and scalable control algorithm based on constrained reinforcement learning (CRL) and convex hulls. While our framework is designed to handle the uncertainty introduced by DP, it is general and also applicable to other sources of uncertainty in EV charging, such as the stochastic nature of driver behavior and renewable variability. This ensures feasible and privacy-preserving coordination of EV charging at scale. Our method constructs convex hulls within the action space to guarantee feasibility under stochastic constraints and incorporates constraint reduction techniques to improve scalability. Case studies based on IEEE benchmark systems demonstrate that the proposed approach effectively balances feasibility under uncertainty, scalability, and privacy in large-scale EV charging control.

Index Terms—IoEV, Differential Privacy, Reinforcement Learning

I. INTRODUCTION

A. Background and Motivation

The Internet of Things (IoT) enables IoT-enabled EVs (IoEVs) to coordinate charging with renewable DERs by sharing real-time data with control systems through sensors and robust communication networks [1]. IoEV charging, particularly at scale, offers considerable spatio-temporal flexibility in terms of both charging locations and time, with recent advances demonstrating enhanced coordination strategies across multiple network domains [2]. IoT integration transforms EV charging systems through three key characteristics: (1) **Enhanced real-time visibility and control** where IoT-enabled EVs provide continuous state monitoring and bidirectional communication, enabling granular control of individual EV charging rates based on grid conditions—impossible in traditional uncoordinated systems; (2) **Data-driven coordination capabilities** that transform EVs from passive loads to

active grid resources, where real-time signals directly influence charging decisions and enable alignment with renewable availability through adaptive, learning-based control strategies, with emerging approaches incorporating advanced forecasting and market participation mechanisms [3]; and (3) **New privacy and security challenges** where IoT connectivity creates unprecedented concerns as fine-grained charging data reveals sensitive behavioral patterns. While this connectivity enhances situational awareness for EVs, Demand Response (DR) aggregators, and the grid [3, 4], it also necessitates privacy-preserving mechanisms for system-wide coordination. As illustrated in Fig. 1, these IoT-enabled characteristics create both opportunities and challenges that motivate our approach to provide differential privacy protection for EV drivers while maintaining coordinated charging capabilities.

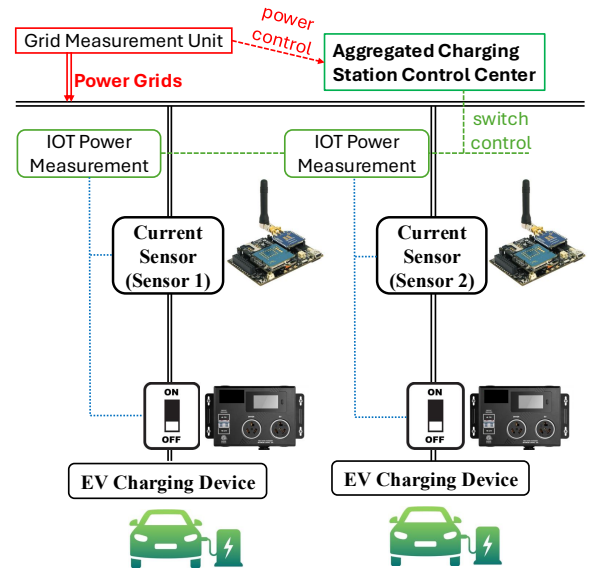


Figure 1. IoT integration with EV architecture for communications with grids and control systems.

To mitigate privacy concerns, this paper considers the adoption of differential privacy (DP) [5], a methodology that provides theoretical bounds on the potential for information leakage when disseminating aggregate data. Recent developments in DP have expanded its applicability to complex IoT architectures and high-dimensional scenarios [6]. The method we propose is tailored to queries that allow to optimize EV scheduling for a population of cars. The mechanism maintains the overall pattern of charging requirements (which are in any case stochastic) without revealing individual data points by adding pseudo-random noise (usually Laplacian or Gaussian) to the actual values. Despite minor changes to individual records that do not change key statistics, the added noise prevents specific inferences about any individual demand patterns from a power system operator that has to continuously

Tong Wu is with the University of Central Florida, Anna Scaglione is with the Department of Electrical and Computer Engineering, Cornell University, 10044 USA (e-mail: {tw385, as337}@cornell.edu). Nikhil Ravi was with the Department of Electrical and Computer Engineering, Cornell University, 10044 USA. Sean Peisert is with Lawrence Berkeley National Laboratory. Daniel Arnold is with Lawrence Livermore National Laboratory. This research was supported in part by the Director, Cybersecurity, Energy Security, and Emergency Response (CESER) office of the U.S. Department of Energy, via the Privacy-Preserving, Collective Cyberattack Defense of DERs project, and in part by the CESER Risk Management and Tools and Technologies (RMT) program of the U.S. Department of Energy via the Mitigation via Analytics for Grid-Inverter Cybersecurity (MAGIC) project under contract DE-AC02-05CH11231. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsors of this work.

track the aggregate available DR flexibility [7]. However, this privacy introduces trade-offs. The added noise introduces further uncertainty into EVs control models, which can lead to infeasible or suboptimal decisions. The inherent variability in EV charging demand—compounded by the randomization introduced by DP—further complicates real-time coordination. Managing this dual challenge of demand uncertainty and privacy preservation is the central objective of this paper.

B. Related Work

In this section, we review related work in the following areas: (1) the use of a DP mechanism in communicating the EV charging flexibility to the operator controlling generation resources, and (2) the control of EVs through deep reinforcement learning (DRL) under uncertainty in DR settings.

a) Differential Privacy in EV Charging: The exploration of DP in EV charging is nascent, with key studies focusing on modifying the charging rate with DP noise and exploring distributed constrained optimization for EV charging under DP [8, 9]. Recent advancements include an adaptive DP federated learning framework for EV charging infrastructures [7, 10], aiming to detect anomalous traffic and enhance the privacy provisioning mechanism. However, these methods often neglect broader privacy issues related to user behavior patterns, including charging times, deadlines, capacity, and initial states of EVs batteries, which leak information about how long the user has traveled. Exposure of this data could reveal detailed information about users' lifestyles and preferences. Although DP can effectively prevent privacy leakage, the uncertainty introduced by integrating existing DP approaches could result in infeasible control actions due to the real-time supply-demand balance and physics-hard constraints inherent in power systems [11]. Despite this challenge, few studies explicitly address the impact of DP-induced uncertainty on the feasibility and performance of EV charging coordination.

b) Federated Learning with Differential Privacy in EV Charging: Existing federated learning approaches for EV charging systems can be categorized into four primary application domains. **Forecasting applications** focus on infrastructure planning, where [12] develops FL-based mobile charging station placement strategies using distributed demand predictions. **Detection methods** enhance system security through anomaly identification, with [13] proposing FedEVCP for detecting charging pile anomalies and [14] designing privacy-preserving threat detection schemes. **Privacy preservation applications** address data protection challenges, where [15] tackles privacy in vehicle-to-grid (V2G) systems and [16] develops secure charging station recommendation systems. **Control applications** focus on charging coordination, where [4] implements federated reinforcement learning for real-time EV charging control on distribution networks. FL+DP methods provide dual-layer privacy protection: (1) federated learning keeps raw training data local to each participant, avoiding centralized data collection, and (2) differential privacy adds calibrated noise to shared model updates, providing mathematical guarantees against inference attacks. However, existing FL+DP approaches primarily focus on offline training scenarios rather than online operational control, leaving a gap in

real-time privacy-preserving EV coordination under dynamic grid constraints.

c) Control of EVs with Demand Response: The domain of EV charging integrated with DR mechanisms can be classified into two main research threads. The first explores the collective potential of EVs within the wholesale market, highlighting the flexibility, aggregate energy demand, and ramping capabilities of DR device groups [17]. The second category [18] investigates the synergy between real-time pricing and DR, focusing on households equipped with high-consumption appliances such as air conditioners and refrigerators, along with EVs and batteries that can all change their load profile in response to economic incentives. The latter studies typically adopt a utility-maximization framework for DR, maximizing the system surplus (users utilities minus cost) to balance the system through dynamic pricing. Despite their contributions, leveraging EVs' flexibility at a large scale introduces three significant challenges for traditional control methods [19]. First, these methods, which have relied on predictive data such as expected EV arrivals, fall short in real-time EV management due to the existence of randomness [20], prompting the development of more robust optimization frameworks that better handle uncertainties. Second, scaling predictive control methods to manage large-scale aggregated EV charging in real time, especially when coupled with power grids, introduces a large-scale nonconvex optimization problem that is impractical for real-time computation [21]. Third, due to the stochastic nature of the problem, predictive control strategies may not ensure feasibility in the future [20].

DRL emerges as a promising alternative, enabling decision-making in uncertain environments, using neural networks (NN) as a function approximator for both policies and value functions [22]. Although the training process of DRL is time-intensive, once trained offline, the NN policy closely approximates the optimal results and can be easily implemented in real-time [23]. Despite these advantages, effectively addressing uncertainty during decision-making remains a critical challenge for ensuring feasible control.

TABLE I: Comparison of EV Charging Control Approaches

Method	Priv.	Feas.	Scale	RT	Unc.
<i>Privacy-Preserving</i>					
DP Rate [8]	P	N	Y	Y	N
Fed. DP [10]	T	-	L	N	P
Dist. DP [11]	Y	L	N	N	N
FL-EV [7]	T	-	Y	N	P
EPDB [24]	Y	L	Y	L	N
V2G [15]	P	-	L	N	N
Sec. FL [16]	T	-	Y	N	P
<i>Traditional Control</i>					
Market [17]	N	Y	L	N	N
Pred. Ctrl [21]	N	Y	N	N	N
DRL [25]	N	N	Y	Y	Y
Fed. RL [4]	N	L	Y	Y	P
Mobile FL [12]	N	-	Y	N	L
CRL+Hull	F	Y	Y	Y	Y

F: Full T: Training only P: Partial
Y: Yes N: No L: Limited
-: N/A RT: Real-time Unc.: Uncertainty

In summary, existing approaches exhibit fundamental limitations: **Differential Privacy in EV Charging** methods provide

theoretical privacy guarantees through DP mechanisms but suffer from operational privacy gaps (protecting only training data, not real-time states), scalability barriers, and feasibility blindness where privacy noise violates power system constraints; **Control of EVs with Demand Response** approaches that handle uncertainty through predictive control but face computational intractability for large-scale systems, real-time infeasibility, and privacy incompatibility requiring direct access to individual EV states; and **Deep Reinforcement Learning** methods that enable decision-making in uncertain environments using neural networks as function approximators and provide real-time implementation once trained offline, but exhibit constraint violation tendencies, action space explosion, and fundamental privacy incompatibility. As demonstrated in Table I, existing approaches address only subsets of the critical requirements for large-scale EV coordination, with no method simultaneously achieving privacy protection, constraint feasibility, scalability, real-time capability, and uncertainty handling—a fundamental gap that motivates this work.

C. Contributions and Organization

To address these challenges, this paper introduces novel strategies for the safe control of EV charging under uncertainty, with integrated DP mechanisms. These strategies enhance the privacy of IoEVs charging data while explicitly addressing the uncertainty inherent in aggregated EVs charging models. Additionally, we improve upon the framework of CRL in [22] to manage the scalability challenges posed by high-dimensional action spaces that would result from scheduling individually each EV. The core contributions include a privacy-preserving communication protocol for EV owners and aggregators, as well as a safe control strategy for the Power System Operator (PSO). Specifically we propose:

- An EV control model is developed to coordinate aggregate EV charging DR with power grid system operations, utilizing the flexibility of EV charging to facilitate renewable energy integration.
- A DP mechanism is designed to protect the privacy of information related to EV charging events such as arrivals, ensuring that power system operators do not have access to data describing the use of individual EVs. This mechanism theoretically proves the ϵ -DP guarantee.
- A method that constrains the EV schedule to be within the intersection of random sample scenario-based constraints. These sampling scenario intersections can encompass real-world unseen scenarios, thereby reliably producing feasible control solutions under uncertainty.
- To solve the stochastic optimal power with uncertainty on net load and uncertain EV charging constraints, we further extend the CRL framework we proposed in [22] by introducing a new scalable representation for the constraints set that governs the population of EVs. The intersection of the possible feasible set is approximated with convex polyhedron that is inscribed in it, and that has a fixed number of constraints. This technique greatly reduces the complexity of both actions and states, thereby enhancing the tractability of the learning-based process.

The structure of the paper is as follows: Section II discusses aggregated EV charging models. Section III introduces the concept of DP, detailing our specific mechanism designed to safeguard the privacy of IoEV charging data. Section IV presents the V2G model, focusing on controlling DP-EV charging, and discusses our scalable control methodology for DP-enhanced EV models. In Section V, we propose the CRL with convex hulls method, laying the groundwork for our control strategies under uncertainty. Section VI showcases the applications of the proposed CRL with convex hulls for the DP-EV charging control, illustrating the implementation and benefits of our approach. The paper concludes with Section VII.

II. AGGREGATE EV CHARGING MODEL

This section introduces a model that illustrates the charging flexibility of a population of EVs within the context of DR. Additionally, we detail the dynamic constraints imposed by EV charging behaviors.

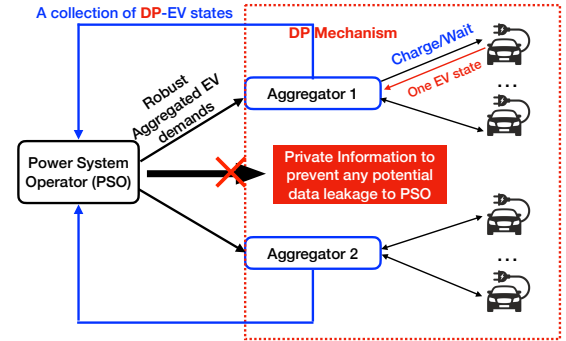


Figure 2. The three-tier system architecture. In this architecture, the users share their states with their aggregator, who then passes on differentially private arrival counts to the PSO. The PSO performs robust control optimization and communicates the optimal control actions back to the aggregators.

A. System Architecture

In this paper, we consider a three-tier infrastructure for the EV charge scheduling problem (see Fig. 2), which focuses on dynamic interactions among the key stakeholders within the system: the PSO, the aggregators, and the EV owners (or the users/consumers), each driven by distinct yet interrelated objectives:

- 1) The PSO aims to minimize the power generation cost and ensure efficient load management while also setting electricity prices. To achieve these goals, the PSO can schedule the aggregate load of EVs based on a description of their feasible aggregate load profile. Since renewable energy generation has no marginal cost, the PSO can ensure that charging of EVs uses as much as possible this non-dispatchable resource.
- 2) The aggregators are intermediaries between the PSO and EV owners, and aim to allocate loads based on factors such as charging demands, user preferences, and locations. It is important to note that while aggregators gather the actual EV states, they transmit DP EV states to the PSO to safeguard the privacy of each EV driver, preventing any potential data leakage to the PSO.
- 3) EV owners prioritize the seamless availability of cost-effective charging that does not interfere with their travel

needs. Significantly, our framework addresses user privacy concerns, safeguarding sensitive information such as arrival times, charging requirements, and departure deadlines – a novel aspect exclusive to our approach.

Notations: In the following, time is discretized and represented by the index t , the time horizon is denoted by $\mathcal{T}$, the sampling interval is normalized by 1; variables that are scalar are indicated in italics, such as $z(t)$, while vectors or matrices are in boldface, like $\mathbf{z}$. Temporal differences are represented by $\partial z(t) = z(t+1) - z(t)$. The symbols $u(t)$ and $\delta(t)$ denote the unit step and Kronecker delta functions, respectively.

B. EV Charging Model

To address the challenges of modeling an EV population and capturing its uncertainty, we modify the *slack-charging* aggregate EV model introduced in [26]. Specifically, we introduce a DP-based uncertainty component that captures randomness in two key parameters: the initial residual energy and the slack times, denoted as $\boldsymbol{\tau} = (\tau_r, \tau_s)$. These parameters characterize individual drivers and introduce uncertainty into the feasibility set of the aggregate model. The residual and slack times $\boldsymbol{\tau} = (\tau_r, \tau_s)$ are discretized as follows:

- $\tau_r \in \{0, \dots, T_r - 1\}$ denotes the time required to charge the EV from its arrival SoC to the target level $\overline{SoC}$ at a fixed rate ρ , i.e., $\tau_r = \lfloor (\overline{SoC} - SoC) / \rho \rfloor$. All vehicles are assumed to charge at the same rate ρ . For example, with a typical Level 2 charger ($\rho = 7.2$ kW), an EV with 40 kWh battery at 20% SoC requiring 80% SoC needs $\tau_r = \lfloor (0.8 - 0.2) \times 40 / 7.2 \rfloor = 3$ hours.
- Given the charging deadline t^d , the slack time $\tau_s \in \{0, \dots, T_s - 1\}$ represents how long the EV can wait before charging, defined as $\tau_s = t^d - t^a + \tau_r$.

For the p th vehicle, $\boldsymbol{\tau}_p(t)$ denotes its current state at time t , starting from $\boldsymbol{\tau}_p(t^a)$. The EV charging process evolves along a path in the $\boldsymbol{\tau}$ space: charging reduces τ_r by 1, while waiting reduces τ_s by 1.

C. System State

In this model, at any time $t \in \mathcal{T}$, an EV's state is described by $(x_r, x_s) \in \mathcal{U}_{rs} \triangleq \mathbb{R}^{\{0, \dots, T_r - 1\} \times \{0, \dots, T_s - 1\}}$, with x_r indicating the remaining charging duration and x_s the remaining slack time an EV driver has to stop charging and await further decisions. For simplicity, in subsequent sections, we use $\mathbf{x} = (x_r, x_s) \in \mathcal{U}_{rs}$ to represent a two-dimensional EV state. Our objective is to manage the progression of $n_{\mathbf{x}}(t)$, the number of EVs in the state $\mathbf{x} = (x_r, x_s)$ at time t . We use $a_p(t) = u(t - t_p^a)$ to indicate whether EV p has arrived by time t . The count of new vehicles that have arrived by time t with specific charging and slack time is:

$$a_{\mathbf{x}}(t) = \sum_{p \in \mathcal{P}} \delta(\tau_{p,r} - x_r) \delta(\tau_{p,s} - x_s) a_p(t), \quad (1)$$

where the process $a_{\mathbf{x}}(t)$ is typically modeled as a non-stationary Poisson process, and $\mathcal{P}$ represents the set of all EVs in the system. The arrival process $a_{\mathbf{x}}(t)$ is modeled as a non-stationary Poisson process with rate $\lambda_{\mathbf{x}}(t)$, where $\lambda_{\mathbf{x}}(t)$

denotes the average number of EVs that, at time t , request a specific amount of charge with a given slack time [27].

The number of EVs transitioning from state $\mathbf{x}$ to $\mathbf{x}'$ at time t is denoted by $\partial d_{\mathbf{x}, \mathbf{x}'}(t)$. Thus, the expression $((\mathbf{x}' - \mathbf{x}) \partial d_{\mathbf{x}, \mathbf{x}'}(t))_r$ indicates the EVs actively charging from $\mathbf{x}$ to $\mathbf{x}'$, and $\rho((\mathbf{x}' - \mathbf{x}) \partial d_{\mathbf{x}, \mathbf{x}'}(t))_r$ represents the total charging power, reflecting the combined EV demands. Furthermore, the total number of EVs transitioning from state $\mathbf{x}$ equals the current EVs at that state. This is because the remaining time for each EV decreases, irrespective of whether they opt for charging or waiting. Thus, the combined feasible set for EVs at charging rate ρ is described by [26]:

$$\begin{aligned} \mathcal{L} = \left\{ \ell_{ev}(t) \mid \ell_{ev}(t) = \sum_{\forall \mathbf{x}} \sum_{\mathbf{x}' \in \mathcal{U}_{\mathbf{x}}} \rho((\mathbf{x}' - \mathbf{x}) \partial d_{\mathbf{x}, \mathbf{x}'}(t))_r, \right. \\ \left. 0 \leq \partial d_{\mathbf{x}, \mathbf{x}'}(t) \leq \bar{d}_{\mathbf{x}, \mathbf{x}'}, \partial d_{\mathbf{x}, \mathbf{x}'}(t) \in \mathbb{Z}^+, \right. \\ \left. \sum_{\mathbf{x}' \in \mathcal{U}_{\mathbf{x}}} \partial d_{\mathbf{x}, \mathbf{x}'}(t) = n_{\mathbf{x}}(t), \forall t \in \mathcal{T} \right\}, \end{aligned} \quad (2)$$

where $(\cdot)_r$ denotes the value associated with the charging dimension and $\ell_{ev}(t)$ the aggregated EV demand at time t . The set $\mathcal{U}_{\mathbf{x}}$ is given by:

$$\mathcal{U}_{\mathbf{x}} = \left\{ \mathbf{x}' \mid \|\mathbf{x}' - \mathbf{x}\|_1 = \min(\|\mathbf{x}\|_1, 1), (\mathbf{x} - \mathbf{x}') \geq \mathbf{0} \right\}, \quad (3)$$

where $\mathcal{U}_{\mathbf{x}}$ defines permissible state transitions, ensuring EVs can either charge or wait (slack) at any time, with transitions to states one unit apart (equality constraint), except near the origin, and maintaining movement within the state space's upper right quadrant (inequality constraints). These constraints require that movements should be restricted to one of the two axes – either charging or waiting.

The system dynamics are governed by the state population tensor $n_{\mathbf{x}}(t)$ and transitions $d_{\mathbf{x}, \mathbf{x}'}(t)$:

$$\begin{aligned} n_{\mathbf{x}}(t) &= a_{\mathbf{x}}(t) + \sum_{\mathbf{x}' \in \mathcal{U}_{\mathbf{x}}} [d_{\mathbf{x}', \mathbf{x}}(t) - d_{\mathbf{x}, \mathbf{x}'}(t)], \forall \mathbf{x}, \\ \partial d_{\mathbf{x}, \mathbf{x}'}(t) &= d_{\mathbf{x}, \mathbf{x}'}(t) - d_{\mathbf{x}, \mathbf{x}'}(t-1), \forall \mathbf{x}, \mathbf{x}' \in \mathcal{U}_{\mathbf{x}}, \end{aligned} \quad (4)$$

where $n_{\mathbf{x}}(t)$ quantifies the EVs present in the system at time t with state $\mathbf{x}$. In our framework, the term $\mathbf{d}_{\mathbf{x}, \mathbf{x}'}^i(t)$ denotes the vector aggregating individual transitions $d_{\mathbf{x}, \mathbf{x}'}$ for EVs located at bus $i \in \mathcal{N}$ across all pairs $\mathbf{x}, \mathbf{x}'$ within the set $\mathcal{U}_{\mathbf{x}}$, at time t . Similarly, $\mathbf{n}_{\mathbf{x}}^i(t)$ specifies the vector that enumerates all EVs at bus i , in state $\mathbf{x}$. Rather than controlling individual EVs, our approach uses population-based control where the state space dimension remains fixed regardless of EV count. We only track the number of EVs in each state $\mathbf{x}$ category rather than individual vehicle states. This fundamental design choice makes our method inherently scalable to large EV populations without increasing computational complexity.

III. DIFFERENTIAL PRIVACY OF EV ARRIVALS

This section explores the application of a pseudo-randomized mechanism to provide DP guarantees for users whose charging patterns are kept confidential from PSOs, who might otherwise manipulate electricity prices.

Threat Model: A particular privacy threat emerges with the reporting of $n_{\mathbf{x}}(t)$, the number of EVs in each charging

state over time. Consider the worst-case scenario where exactly one car arrives between two consecutive time intervals, t and $t + 1$. Then, if the count of EV in state $\mathbf{x}$ at time $t + 1$, $n_{\mathbf{x}}(t + 1)$, increases by one unit from $n_{\mathbf{x}}(t)$, it can be directly inferred that the new EV has arrived in state $\mathbf{x}$: $n_{\mathbf{x}}(t + 1) = n_{\mathbf{x}}(t) + 1$. This direct correlation allows the PSO to determine the state $\mathbf{x}$ of the new arrival, posing the classical privacy risk of resolving individual records from repeated queries to the population. This example illustrates how real-time data sharing, even when aggregated, can potentially lead to the re-identification of individual behaviors or states, which could be exploited to infer sensitive user information [28].

To mitigate this threat, we employ DP, which ensures the privacy of individual data entries in statistical databases through controlled random noise:

Definition 1 (Differentially Private (DP) Mechanism) For $\varepsilon > 0$, a pseudo-randomized mechanism, $\mathbb{A}$, is said to be ε -DP if, for all $\mathcal{Q} \subseteq \text{Range}(\mathbb{A})$ and for all adjacent¹ vectors $X, X' \in \text{Domain}(\mathbb{A})$, the following holds:

$$\mathbb{P}[\mathbb{A}(X) \in \mathcal{Q}] \leq e^\varepsilon \mathbb{P}[\mathbb{A}(X') \in \mathcal{Q}]. \quad (5)$$

DP ensures that for any two adjacent datasets X and X' , the distributions of $\mathbb{A}(X)$ and $\mathbb{A}(X')$ are very similar, with the total variation distance between them at most ε .

Proposed DP Mechanism: To prevent PSOs or third-party analysts from inferring an EV's initial state τ_p from counts $n_{\mathbf{x}}(\cdot)$, it is imperative to obscure each EV's state in a privacy-preserving manner and ensure that details of individual customers are concealed. We consider a mechanism $\mathbb{A}$ that takes X , a series of EV arrival states, and outputs a time series of counts $n_{\mathbf{x}}(t)$ for all states $\mathbf{x}$ across time steps t from 1 to T . These counts are calculated by modifying the state of each EV at the time of their arrival using the Randomized Response (RR) mechanism, with these randomized states then used consistently in subsequent counts without further modification. Originally developed by Warner [29] and adapted for DP [30], the RR mechanism is used to modify each EV's state τ_p to a DP state $\tilde{\tau}_p$ using the following probability distribution:

$$\Pr[\tilde{\tau}_p = \mathbf{x} \mid \tau_p] = \begin{cases} \frac{e^\varepsilon}{e^\varepsilon + K - 1}, & \text{if } \mathbf{x} = \tau_p, \\ \frac{1}{e^\varepsilon + K - 1}, & \text{if } \|\mathbf{x} - \tau_p\|_1 = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Here, the privacy parameter ε controls the probability of concealing the ground-truth charging state, and K represents the number of possible states $\mathbf{x}$ satisfying $\|\mathbf{x} - \tau_p\|_1 \leq 1$. We then compute the randomized arrival counts $\tilde{a}_{\mathbf{x}}(t)$ as follows:

$$\tilde{a}_{\mathbf{x}}(t) = \sum_{p \in \mathcal{P}} \delta(\tilde{\tau}_{p,r} - x_r) \delta(\tilde{\tau}_{p,s} - x_s) a_p(t), \quad (7)$$

where $\tilde{\tau}_p$ is the output of the RR mechanism for each EV p . Note that each EV can locally apply the mechanism randomizing the time stamp of its arrival and required deadline.

Theorem 1 The mechanism $\mathbb{A}$ that takes X , a series of EV arrival states, and outputs a time series of counts $n_{\mathbf{x}}(t)$ for

all states $\mathbf{x}$ and time steps t from 1 to T , by utilizing the RR mechanism in Eq. (6) at the time of EV arrival and consistently using these randomized states in subsequent counts without additional randomization, is ε -DP for all EV arrivals X and X' that differ in the state of exactly one EV at any given time step.

Proof 1 Consider any two adjacent inputs τ and τ' and any possible output $\tilde{\mathbf{x}}$ by Eq. (6). $\Pr[\tilde{\tau}_p = \mathbf{x} \mid \tau_p]$ is maximized when $\mathbf{x} = \tau_p$, whereas $\Pr[\tilde{\tau}_p = \mathbf{x} \mid \tau'_p]$ is minimized when $\mathbf{x} \neq \tau'_p$. This implies that

$$\frac{\Pr[\tilde{\tau}_p = \mathbf{x} \mid \tau_p]}{\Pr[\tilde{\tau}_p = \mathbf{x} \mid \tau'_p]} \leq \frac{\frac{e^\varepsilon}{e^\varepsilon + K - 1}}{\frac{1}{e^\varepsilon + K - 1}} = e^\varepsilon.$$

Thus, the RR mechanism in (6) is ε -DP. Furthermore, even though we are releasing the entire time series of counts $n_{\mathbf{x}}(t)$ for all states $\mathbf{x}$, we apply the mechanism individually to each EV, where an EV's state is fed through the RR mechanism only once during its charging lifetime. Consequently, we employ the parallel composition [31, Theorem 4] of ε -DP to provide ε -DP for the entire system.

Fig. 3(a), illustrates the mechanism, where the modified response has a 90% chance of being the true label of (3, 1), and 2.5% chance of being one of its four neighbors at distance one. In our mechanism, the parameter K depends on the true state τ_p . When the true state is away from the boundaries as shown in Fig. 3(a), K is equal to 5 (having 4 neighbors at a distance of one and itself). In scenarios where the state is at a corner, $K = 3$, and $K = 4$ when the state is at an edge. Fig. 3(b) provides an example of the original and differentially private states of six EV drivers.

The DP mechanism introduces randomness into EV arrival counts observed by the PSO. However, it is crucial to address the inherent uncertainty and ensure that the optimal actions derived from these counts remain feasible for the actual arriving EVs. In Section IV, we present a scalable method that incorporates a comprehensive set of sampling-based constraints, thereby ensuring that optimal solutions adhere to network constraints under all possible conditions.

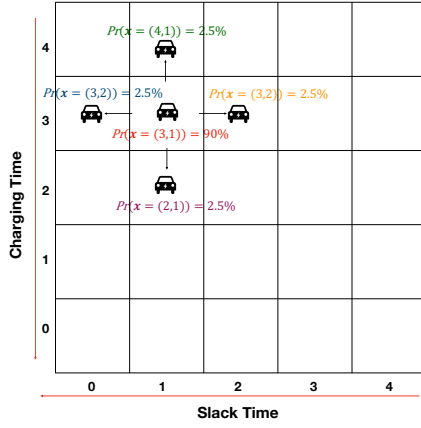
IV. PROBLEM FORMULATION: CONTROL OF DP-EV MODELS UNDER UNCERTAINTY

Building on the foundation laid by the integration of DP to safeguard EV arrival patterns, we now turn our focus to the challenges introduced by the added uncertainty in the optimization process. In this section, we formulate the control algorithm **from the perspective of the PSO**, who must make real-time decisions based on DP-protected and uncertain system information.

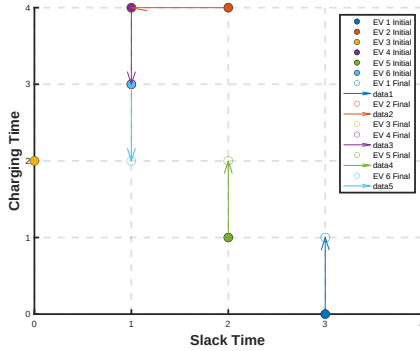
A. Predictive Stochastic Optimal Control

In this section, we formulate a multi-stage stochastic optimal control problem that addresses the integration of EVs with power grid operations, taking into account the complexities introduced by power flow constraints and adaptable EV charging. Our focus is to solve OPF with aggregated EVs to minimize fuel costs, underpinned by a formulation that

¹Adjacent vectors differ in only one record.



(a) An illustration of the DP pseudo-random mechanism.



(b) Visualization: Original and Differential Privacy States.

Figure 3. DP mechanism and Visualization

seeks to balance economic dispatch within network constraints while managing EV and battery operations. The problem is formalized as:

$$\max_{\pi(s(t-1))} \mathbb{E}_{\ell_d(t)} \left[\sum_{t=t_o}^{t_o+T-1} r(s(t), \mathbf{A}(t), \ell_d(t)) \right] \quad (8a)$$

$$s(t) = f_t(s(t-1), \mathbf{A}(t-1), \ell_d(t-1)), \quad (8b)$$

$$\mathbf{A}(t) = \pi(s(t-1)), \quad (s(t), \mathbf{A}(t)) \in \Omega(t), \quad (8c)$$

where $s(t)$ denotes the state vector, $\mathbf{A}(t)$ the control vector including all controllable devices, and $\ell_d(t)$ the demands and renewable energies. The cost function is represented by $r(s(t), \mathbf{A}(t), \ell_d(t))$, with $\Omega(t)$ representing network and device bound constraints. The randomized control policy, $\pi(s(t-1))$, guides the decision-making process to achieve network-constrained economic dispatch of power flows by controlling power generations, EVs, and batteries.

1) *Actions and States*: Let $\mathcal{N}$ denote the set of nodes within the power system, $\mathcal{G}$ denote the set of nodes connected to power generators, and $\mathcal{B}$ denote the set of nodes connected to battery storage and EV charging substations. In particular, the system state at time $t-1$ is defined as

$$s(t-1) \triangleq [v(t-1), \text{soc}(t-1), (\tilde{\mathbf{a}}_{\mathbf{x}}^i(t-1)) \forall i \in \mathcal{N}]^\top, \quad (9)$$

where $v(t-1)$ denotes the voltage phasors, $\text{soc}(t-1)$ is the SoC vector of all batteries in the system, and $(\tilde{\mathbf{a}}_{\mathbf{x}}^i(t-1)) \forall i \in \mathcal{N}$ represents the DP-protected version of the EV

arrival matrix available to the PSO. While EV aggregators at charging stations have access to the actual arrival data in the past, they only share a DP-protected version with the PSO to preserve user privacy. The control vector, $\mathbf{A}(t) = [g^p(t); g^q(t); p_{ch}(t); p_{dis}(t); (\partial \tilde{\mathbf{a}}_{\mathbf{x}, \mathbf{x}'}^i(t))_{i \in \mathcal{N}}]^\top$, encompasses various elements. It includes both active and reactive generations, denoted as $g^p(t) = (g_i^p(t))_{i \in \mathcal{G}}$ and $g^q(t) = (g_i^q(t))_{i \in \mathcal{G}}$, respectively. Additionally, it includes the charge and discharge rates of batteries, which are represented as $p_{dis}(t) = (p_{dis}^i(t))_{i \in \mathcal{B}}$ and $p_{ch}(t) = (p_{ch}^i(t))_{i \in \mathcal{B}}$, respectively. Furthermore, the vector includes EV control actions under uncertainty, denoted by $\partial \tilde{\mathbf{a}}_{\mathbf{x}, \mathbf{x}'}^i(t)$.

2) *Reward Function*: In Eq. (8a), we define our reward function as $r(s(t), \mathbf{A}(t), \ell_d(t))$, or more compactly, $r(t)$. This function reflects our objectives, which include minimizing fuel costs, represented by $\varsigma_f(t)$ (equivalently, maximizing the negative of fuel costs), and smoothing the EV charging demand, denoted by $\varsigma_m(t)$. The formulation is as follows:

$$r(t) = -\varsigma_f(t) - \varsigma_m(t), \quad (10)$$

indicating that higher rewards are obtained for lower fuel costs and smoother EV demand profiles. To achieve the goal of reducing fuel costs, we delve into optimizing the operation of power grids through the integration of EVs, focusing on OPF with the two goals of reducing fuel costs and accommodating power flow constraints along with adaptable EV charging schedules. The primary aim is to explore the economic and operational benefits of aggregated EV participation in the power grid. To this end, we formulate the objective of minimizing fuel costs as follows:

$$\varsigma_f(t) = \sum_{i \in \mathcal{G}} (v_i(g_i^p(t))^2 + \varrho_i g_i^p(t) + \iota_i), \quad (11)$$

where v_i , ϱ_i , and ι_i are positive coefficients, representing the quadratic and linear cost components of generation at generator i at time t . To smooth EV charging demand, we incorporate DR mechanisms via EV charging management to enhance grid flexibility and reduce peak demand. This involves smoothing the charging demands of EVs to avoid sharp increases or decreases, thereby facilitating a more stable load profile across the grid. The objective is designed to measure the variability of EV charging demand:

$$\varsigma_m(t) = \sum_{i \in \mathcal{B}} \|\ell_{ev}^i(t) - \ell_{ev}^i(t-1)\|_2^2, \quad (12)$$

where $\ell_{ev}^i(t)$ denotes the charging demand of EVs at battery i and time t , aiming to minimize the variation in demand from one time step to the next.

After formulating the stochastic control problem, we face the challenge of a hybrid action space, which includes both integer variables for EV charging and continuous variables for power generation. To address this, we utilize the Shapley-Folkman Lemma, as noted in [17], to approximate the aggregate feasible action space using continuous variables².

²The details of the Shapley-Folkman Lemma [17] are beyond the scope and focus of this work, and therefore, are not detailed further.

B. Feasible V2G Charging Control under Uncertainty

Building on the stochastic control problem formulated in Section IV-A, we now propose the scenario-based security constrained OPF for V2G charging control under uncertainty. This shift is necessitated by the DP mechanism's inherent uncertainty, which, while ensuring privacy, introduces unpredictability that could lead to infeasible control solutions. Given this context, our objective expands to ensuring that optimal actions not only account for DP-induced uncertainty but also remain feasible for the actual EV arrivals.

Again, $\tilde{a}_x(t)$ denotes the DP-protected EV arrival information, which remains uncertain to the PSO. To address this, we adopt a scenario sampling approach from the PSO's perspective. Let $\mathcal{S}$ denote the sampling space consisting of scenarios $s \in \mathcal{S}$, each generated by applying the DP mechanism (as described in Section III) to the ground-truth EV arrival matrix a_x , $\forall x$, resulting in scenario-specific realizations a_x^s .

For each scenario s and each bus i , we derive the corresponding EV arrival data ($a_{x,i}^{s,i}(t)$) and charging demands ($\partial d_{x,i}^{s,i}(t)$), which induce the scenario-specific aggregate EV demand feasible set $\mathcal{L}_i^s$ at each node i . To ensure feasibility, we define the aggregate feasible set as the intersection of all scenario-based sets:

$$\ell_{i,ev}(t) \in \cap_{s \in \mathcal{S}} \mathcal{L}_i^s, \quad \ell_{ev}(t) = (\ell_{i,ev}(t))_{i \in \mathcal{B}}. \quad (13)$$

This intersection guarantees feasibility under the unknown ground-truth realization, as any solution valid across all sampled scenarios is also valid under the actual (but private or uncertain) inputs.

To this end, we cast the OPF problem into a scenario-based security constrained optimization framework, ensuring that the control solution satisfies all network constraints across the sampled arrival scenarios. The resulting scenario-based OPF model leverages the intersection of scenario-induced feasible sets as follows:

$$\max_{A(t), t \in [t_o, t_o + T - 1]} \mathbb{E}_{\ell_d(t)} \left[\sum_{t=t_o}^{t_o + T - 1} r(t) \right] \quad (14a)$$

$$\mathbf{M}_b(\mathbf{p}_{dis}(t) - \mathbf{p}_{ch}(t) + \ell_{ev}(t)) + \mathbf{M}_g \mathbf{g}^p(t) - \ell_d^p(t) = \Re\{D(\mathbf{v}(t)(\mathbf{v}(t))^H \mathbf{Y}^H)\}, \quad \ell_{i,ev}(t) \in \cap_{s \in \mathcal{S}} \mathcal{L}_i^s, \quad \mathcal{L}_i^s \in (2) - (4) \quad (14b)$$

$$\mathbf{M}_g \mathbf{g}^q(t) - \ell_d^q(t) = \Im\{D(\mathbf{v}(t)(\mathbf{v}(t))^H \mathbf{Y}^H)\}, \quad (14c)$$

$$\underline{\mathbf{g}}^p \leq \mathbf{g}^p(t) \leq \bar{\mathbf{g}}^p, \quad \underline{\mathbf{g}}^q \leq \mathbf{g}^q(t) \leq \bar{\mathbf{g}}^q, \quad \underline{\mathbf{v}} \leq |\mathbf{v}(t)| \leq \bar{\mathbf{v}} \quad (14d)$$

$$0 \leq \mathbf{p}_{ch}(t) \leq \mathbf{P}_{rated}^{ch}, 0 \leq \mathbf{p}_{dis}(t) \leq \mathbf{P}_{rated}^{dis}, |\mathbf{K}\mathbf{v}(t)| \leq \mathbf{s}j_{\max} \quad (14e)$$

$$\text{SoC}_{\min} \leq \text{SoC}(t) \leq \text{SoC}_{\max}, \forall t \in [\tau, \tau + T - 1] \quad (14f)$$

$$\text{SoC}(t) = \text{SoC}(t-1) + \frac{\Delta t}{E_{cap}} \left(\eta_{ch} \mathbf{p}_{ch}(t) - \frac{\mathbf{p}_{dis}(t)}{\eta_{dis}} \right). \quad (14g)$$

where the active power demand vector is $\ell_d^p(t) = [\ell_d^p(n, t)]_{\forall n \in \mathcal{N}}$, the reactive power demand vector is $\ell_d^q(t) = [\ell_d^q(n, t)]_{\forall n \in \mathcal{N}}$, $\mathbf{Y}$ is the admittance matrix, and $\mathbf{v}(t) = [v(n, t)]_{\forall n \in \mathcal{N}}$ is the grid state in the AC power flow, i.e. $\mathbf{v}(t) = |\mathbf{v}(t)| \circ e^{j\theta(t)}$, $v(n, t) = |v(n, t)| e^{j\theta(n, t)}$, where $v(n, t) =$

$|v(n, t)| e^{j\theta(n, t)}$ corresponds to the measured bus voltage magnitude and phase angle at bus n . The matrix $\mathbf{K} \triangleq \mathbf{Y}\mathbf{I}$, where $\mathbf{I} \in \mathbb{R}^{m \times n}$ is a network-directed graph incidence matrix, where m is the number of lines in the network. The complex power injection is given by $\mathbf{s}j = (\mathbf{v} \circ \mathbf{i}^*) = D(\mathbf{v}(\mathbf{i})^H)$, where $\mathbf{v} \in \mathbb{C}^N$ represents the voltage phasor vector, and $\mathbf{i} \in \mathbb{C}^N$ symbolizes the current phasor vector. $(\cdot)^*$ denotes the conjugate of the complex vector or matrix and $\circ$ denotes the Hadamard product (element-wise product). The function $D(\cdot)$ extracts the vector of diagonal elements from a matrix, while $(\mathbf{i})^H$ and $(\mathbf{i})^*$ represent the Hermitian and the conjugate of the vector $\mathbf{i}$, respectively. The current phasor vector can be further expressed as $\mathbf{i} = \mathbf{Y}\mathbf{v}$, thereby allowing $\mathbf{s}j$ to be rewritten as $D(\mathbf{v}\mathbf{v}^H \mathbf{Y}^H)$. Therefore, the active and reactive power injections are expressed as $\Re(\mathbf{s}j) = \Re\{D(\mathbf{v}\mathbf{v}^H \mathbf{Y}^H)\}$ and $\Im(\mathbf{s}j) = \Im\{D(\mathbf{v}\mathbf{v}^H \mathbf{Y}^H)\}$. $\mathbf{s}j_{\max}$ denotes the vector of long-term branch rating limits. Let $\mathbf{M}_g$ the matrix $\{0, 1\}^{N \times G}$ that maps the generation vector $\mathbf{g}^p \in \mathbb{R}^G$ to $\mathbb{R}^N$ as follows:

$$\begin{aligned} [\mathbf{M}_g \mathbf{g}^p]_i &= 0, [\mathbf{M}_g \mathbf{g}^q]_i = 0, \quad \forall i \in \mathcal{N} \setminus \mathcal{G} \\ [\mathbf{M}_g \mathbf{g}^p]_i &= g_j^p, [\mathbf{M}_g \mathbf{g}^q]_i = g_j^q, \quad \forall i \in \mathcal{G}, \quad \forall j \in [1, \dots, G] \end{aligned} \quad (15)$$

and the matrix $\mathbf{M}_b \in \{0, 1\}^{N \times B}$ functions as a mapping tool for the vectors of charging power $\mathbf{p}_{ch}(t)$, discharging power $\mathbf{p}_{dis}(t)$, and EV demands $\ell_{ev}(t)$ across the entire network. It assigns zero to those buses that lack battery storage or EV charging substations. The variable SoC denotes the SoC of the battery storage systems, with Eq. (14g) detailing the dynamic evolution of the SoC. The feasible set of Eq. (14) is denoted by Ω_t . The above optimization models are usually computationally intractable due to the complicating constraints Eqs. (2) - (4), compounded by the addition of a large number of scenario-based constraints denoted by $s \in \mathcal{S}$. We address this problem in the next section.

V. METHODOLOGY: CRL WITH CONVEX HULLS FOR ACTION SPACE DIMENSION REDUCTION

Albeit relying on an aggregate population model for the EVs, the problem formulation in (14) suffers from the curse of dimensionality, because each $\mathcal{L}_i$ at bus i in Eq. (2) and (even more so) the intersection $\cap_{s \in \mathcal{S}} \mathcal{L}_i^s$ across all possible random s , includes a large number of linear equality and inequality constraints. The goal of this section is to make the formulation tractable. To do so, we propose a low-order polyhedron approximation strategy for representing $\cap_{s \in \mathcal{S}} \mathcal{L}_i^s$, termed CRL with Convex Hulls (CRL+Hull for short). Notably, the proposed algorithm is general and applicable beyond DP-induced uncertainty—it can handle various sources of uncertainty in EV population behavior.

A. Constrained TD3

First, let us briefly review the constrained Twin Delayed Deep Deterministic Policy Gradient (TD3) framework. The TD3 method, which falls within an actor-critic framework, updates policy function parameters based on a critic or approximate value function [32]. The actor, π_ϕ , selects actions, while the critic, Q_ξ , evaluates them. Q-learning uses temporal

difference learning to derive the value function from the Bellman equation.

For our implementation, we adopt two target networks, specifically $Q_{\xi'_1}$ and $Q_{\xi'_2}$, and complement them with two critic networks, Q_{ξ_1} and Q_{ξ_2} . The methodology involves selecting the minimum value between the two value estimates, which promotes a consistent update process for the critic networks. Upon the suitable definition of the critic, we proceed to configure the actor network, introducing a constrained action space. Conventionally, the action network is trained to maximize the critic network's output. This is articulated as:

$$\phi \leftarrow \arg \max_{\phi} Q_{\xi_1}(s(t-1), \pi_{\phi}(s(t-1))). \quad (16)$$

where ϕ represents the parameters of the action network. We can employ either Q_{ξ_1} or Q_{ξ_2} to guide $\pi_{\phi}(\cdot)$ in updating ϕ . An action $\mathbf{A}(t) = \pi_{\phi}(s(t-1))$ is deemed feasible if $\mathbf{A}(t)$ adheres to all its constraints, $\Omega(t)$. Consequently, the policy π_{ϕ} is derived by maximizing the critic network while ensuring compliance with all the constraints $\Omega(t)$:

$$\max_{\phi} Q_{\xi_1}(s(t-1), \pi_{\phi}(s(t-1))) \quad s.t. \quad \mathbf{A}(t) \in \Omega(t). \quad (17)$$

where $\mathbf{A}(t)$ is taken according to policy $\mathbf{A}(t) = \pi_{\phi}(s(t-1))$.

B. Refined Polyhedron Constraints for Reduced Order

In this subsection, we introduce the details of the low-order convex hulls within the CRL+Hull framework, which effectively addresses the limitations of CRL or general DRL in high-dimensional action spaces.

Specifically, the intersections of feasible sets, i.e., Eqs. (2) - (4), is a polytope which means $\mathcal{L}$ can be replaced by convex combination of vertices $\nu \in \mathcal{V}$. In discussing the simplification of EV constraints, it is crucial to understand that the feasible region of a problem constrained by linear equations forms a polytope. This insight enables us to map the EV feasible set into a lower-dimensional space, significantly reducing the complexity of dimensional control actions. Fig. 4 depicts the process of low-dimensional reduction. For the intersection of a series of linear constraints, we can select four feasible vertices from the boundaries of this polytope, and the convex combination of these vertices is also feasible within the region of this intersection. Therefore, instead of controlling the high-dimensional actions $(\partial \mathbf{d}_{x,x'}^{s,i}(t))_{s \in \mathcal{S}, i \in \mathcal{N}}$, we can manage the convex combination scalars, significantly simplifying the control mechanism.

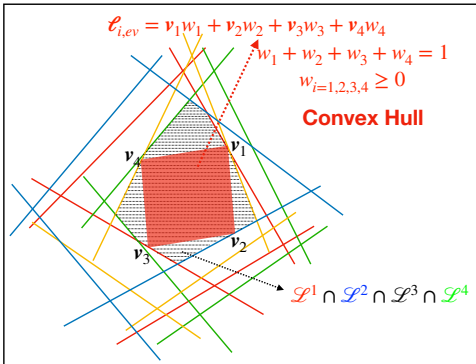


Figure 4. An illustration of the reduced Polyhedron approximation of intersections.

We denote t^+ by a time period, and t by a time point. Let $\ell_{i,ev}(t^+) = [\ell_{i,ev}(\tau)]_{\tau=t}^{t+T-1}$ for a time period, and let $\ell_{ev}(t^+) = [\ell_{i,ev}(t^+)]_{i \in \mathcal{N}}$ for all buses. When we describe a convex combination of vertices, represented as $\nu_{\beta}(t^+) \in \mathcal{V} = \text{Vert}(\cap_{s \in \mathcal{S}} \mathcal{L}^s)$, the expression is formalized as:

$$\ell_{i,ev}(t^+) = \sum_{\beta=1}^{|\mathcal{V}|} \nu_{\beta}(t^+) w_{\beta}(t), \quad \sum_{\beta=1}^{|\mathcal{V}|} w_{\beta}(t) = 1. \quad (18)$$

where $\mathbf{w}(t) \triangleq [w_{\beta}(t)]_{\forall \beta}$. We can see that the convex hull for the aggregate EV demands is much easier to manage. The high-dimensional control variables of $\cap_{s \in \mathcal{S}} \mathcal{L}^s$ (linear constraints) are significantly simplified into these scalars $w_{\beta}(t)$.

This reduction necessitates identifying the vertices of this region $\cap_{s \in \mathcal{S}} \mathcal{L}^s$. With a set of price scenarios given as $\kappa_{\beta} \in \mathbb{R}^T, \beta \in \mathbb{R}$, our starting point is with $\mathcal{V} = \emptyset$. For each price scenario, we determine the optimal power profile:

$$\nu_{\beta} = \arg \min \kappa_{\beta}^T \ell_{i,ev}(t^+), \quad (19a)$$

$$\nu_{\beta} \in \cap_{s \in \mathcal{S}} \mathcal{L}^s. \quad (19b)$$

Following this, we incorporate the outcome, $\ell_{i,ev}(t^+)$, into $\mathcal{V}$. It is important to recognize that these convex hulls serve as an effective method for reducing the action space in CRL, capable of representing either Eqs. (2) - (4) from various sampling scenarios or the full constraints of (14). In this work, we choose to employ the convex hulls only for (2) to (4), effectively balancing complexity and optimality. Another heuristic approach to further simplify the region is excluding vertices that closely align with others, thereby reducing the polyhedron facets count. Therefore, $\ell_{i,ev}(t^+) \in (14b)$ can be replaced by $\ell_{i,ev}(t^+) = \sum_{\beta=1}^{|\mathcal{V}|} \nu_{\beta}(t^+) w_{\beta}(t)$, where $\mathbf{w}$ are the control variables. In light of the above, the revised Eq. (14b) is

$$\begin{aligned} & \mathbf{M}_b(\mathbf{p}_{dis}(t) - \mathbf{p}_{ch}(t) + \ell_{ev}(t)) + \mathbf{M}_g \mathbf{g}^p(t) - \ell_d^p(t) = \\ & \mathbb{R}\{D(\mathbf{v}(t)(\mathbf{v}(t))^H \mathbf{Y}^H)\}, \quad \ell_{ev}(t) = [\ell_{i,ev}(t)]_{i \in \mathcal{B}} \\ & \ell_{i,ev}(t^+) = \sum_{\beta=1}^{|\mathcal{V}|} \nu_{\beta}^i(t^+) w_{\beta}^i(t), \quad \sum_{\beta=1}^{|\mathcal{V}|} w_{\beta}^i(t) = 1. \end{aligned} \quad (20)$$

1) *Advantages of Dimensionality Reduction:* Although linear reduction techniques like PCA [33] exist for dimensionality reduction, nonlinear methods such as autoencoders [34] learn low-dimensional latent representations with decoders that map back to action spaces. However, these decoders are trained purely for reconstruction accuracy without incorporating feasibility constraints. Consequently, decoded actions from arbitrary latent points can easily violate our linear constraints on timing, energy requirements, and charging rates. Our convex hull approach uniquely addresses these challenges by providing guaranteed feasibility, where any convex combination of feasible vertices remains feasible—a mathematical property that no data-driven method can guarantee. Additionally, convex hulls offer simplicity in learning as they only require learning the convex weights rather than complex nonlinear mappings. Another major advantage of this approach is its ability to simultaneously perform constraint dimensionality reduction for CRL across various charging stations that are

situated in the same load bus, and even residential charging. In contrast, the original problems required simultaneous attention to all control variables associated with different charging stations, which made the learning process extremely challenging to train. Thus, the optimization problem is formulated as Eq. (14), excluding Eq. (14b), combined with Eq. (20).

Remark 1 *Through this dimensionality reduction, the dimensions of the state and action spaces are significantly reduced. Specifically, we reduce from:*

Original state: $s(t) \triangleq [v(t), \text{soc}(t), (\mathbf{a}_{\mathbf{x}}^{s,i}(t))_{\forall s \in \mathcal{S}, \forall i \in \mathcal{N}}]^\top$

Original action:

$\mathbf{A}(t) \triangleq [\mathbf{g}^p(t); \mathbf{g}^q(t); \mathbf{p}_{ch}(t); \mathbf{p}_{dis}(t); (\partial \mathbf{d}_{\mathbf{x},\mathbf{x}'}^{s,i}(t))_{\forall s \in \mathcal{S}, \forall i \in \mathcal{N}}]^\top$

to:

Reduced state: $s(t) \triangleq [v(t), \text{soc}(t), \mathbf{v}_\beta(t^+)]^\top$

Reduced action:

$\mathbf{A}(t) \triangleq [\mathbf{g}^p(t); \mathbf{g}^q(t); \mathbf{p}_{ch}(t); \mathbf{p}_{dis}(t); \mathbf{w}_\beta(t)]^\top$

Using the convex hulls to approximate extensive scenario-based constraints effectively eliminates numerous variables, such as $\partial \mathbf{d}_{\mathbf{x},\mathbf{x}'}^s(t), \forall \mathbf{x}, \mathbf{x}'$, simplifying the learning process. While this approach is bound to have lower or equal cost with respect to the original one, our simulation results in Section VI demonstrate that the convex hulls in CRL approximates optimality with a very small gap.

2) *Why These Convex Hulls Promotes Feasibility:* Similar to security-constrained unit commitment problems with scenario constraints [35], which optimize the on/off status of each unit across multiple wind power scenarios, our philosophy is analogous but more intelligent. We identify a convex hull within the intersection of scenario-based security constraints to ensure feasibility. This convex combination approach promotes feasibility because: (1) each vertex $\mathbf{v}_\beta(t^+)$ represents a feasible solution under *all* DP scenarios (i.e., vertices of the intersection $\cap_{s \in \mathcal{S}} \mathcal{L}_s^i$), (2) any convex combination of feasible points remains feasible within the polytope, and (3) by using vertices from the intersection of all scenarios, we ensure validity across *numerous possible DP realizations*. Our CRL framework then optimizes over the low-dimensional weights $\mathbf{w}(t)$ rather than high-dimensional EV actions, achieving both computational efficiency and feasibility.

3) *Disaggregation of Aggregated EV Demands:* We now shift our focus to the EV charging stations (i.e., the aggregator perspective). After uploading the DP-protected EV arrival data to the PSO, the PSO returns the optimal aggregate EV charging demand to the aggregators—this demand should be achievable given the actual EV arrivals observed locally. While the PSO operates based on uncertain information, each EV charging station has access to the actual arrival data and charging patterns of its local EVs. This allows the aggregator to disaggregate the received aggregate demand into individual-level charging decisions.

Let the desired aggregate EV demand be $\ell_{i,ev} = \sum_{i=1}^{|\mathcal{V}|} \nu_i w_i^*$. The term $\ell_{i,ev}(t)$ indicates the aggregated EV demand at time t for bus i , and it corresponds to the first

element of $\ell_{i,ev}$. For the EV charging control the aggregator needs to map the load profile back onto the control signal $\partial \mathbf{d}_{\mathbf{x},\mathbf{x}'}^i(t), \forall \mathbf{x}, \forall \mathbf{x}' \in \mathcal{U}_{\mathbf{x}}$. This can be achieved solving, $\forall i, t$, the projection problem:

$$\min \left\| \ell_{i,ev}(t) - \sum_{\forall \mathbf{x}} \sum_{\mathbf{x}' \in \mathcal{U}_{\mathbf{x}}} (\rho(\mathbf{x}' - \mathbf{x}) \partial \mathbf{d}_{\mathbf{x},\mathbf{x}'}^i(t))_r \right\|_2^2 \quad (21)$$

$$\text{s.t. } \partial \mathbf{d}_{\mathbf{x},\mathbf{x}'}^i(t) \in (2)$$

In this context, the disaggregation problem leverages current actual arrivals to make localized decisions. In this localized disaggregation approach, there is no interaction with centralized agents. Conversely, in the centralized OPF problem, the DP arrival matrix is transmitted to the central agent.

C. Primal-Dual OPF Formulation in the CRL+Hull

In this subsection, to make the paper self-contained, we adapt the strategy in [22] and describe on how to embed the convex hulls into the constrained policy function, denoted as $\pi_\phi(\cdot)$. We introduce the normalized active power generation $\mathbf{g}^p(t)$ and the normalized reactive power generation $\mathbf{g}^q(t)$. Moreover, we define the normalized discharging power as $\mathbf{p}_{dis}(t)$ and the charging power as $\mathbf{p}_{ch}(t)$. Rather than directly controlling the high-dimensional variable $\partial \mathbf{d}_{\mathbf{x},\mathbf{x}'}^{s,i}(t)$, our control actions are represented by $\mathbf{w}_\beta$, which are convex combination weights. The actions $\mathbf{A}(t) = [\mathbf{a}_t, \dots, \mathbf{a}_{t+T-1}]$ are expressed as:

$$\begin{aligned} \mathbf{a}(t) &\triangleq [\hat{\mathbf{g}}^p(t), \hat{\mathbf{g}}^q(t), \hat{\mathbf{p}}_{ch}(t), \hat{\mathbf{p}}_{dis}(t), \mathbf{w}(t)]^\top, \\ \mathbf{g}^p(t) &\triangleq (1 - \hat{\mathbf{g}}^p(t)) \underline{\mathbf{g}}^p + \hat{\mathbf{g}}^p(t) \overline{\mathbf{g}}^p, \\ \mathbf{g}^q(t) &\triangleq (1 - \hat{\mathbf{g}}^q(t)) \underline{\mathbf{g}}^q + \hat{\mathbf{g}}^q(t) \overline{\mathbf{g}}^q, \\ \mathbf{p}_{ch}(t) &\triangleq \hat{\mathbf{p}}_{ch}(t) \overline{\mathbf{p}}_b, \quad \mathbf{p}_{dis}(t) \triangleq \hat{\mathbf{p}}_{dis}(t) \overline{\mathbf{p}}_b \end{aligned} \quad (22)$$

where the power system environmental is controlled by the unnormalized power generation, unnormalized discharging and charging powers, and the unnormalized aggregated EV demand. The vector $\mathbf{w}(t)$ inherently lies within $[0, 1]$ due to $\sum_\beta w_\beta(t) = 1$ and $w_\beta(t) \geq 0$. This can also be equivalently articulated through inequality constraints:

$$0 \leq \sum_{\beta=1}^{|\mathcal{V}|-1} w_\beta(t) \leq 1 \quad (23)$$

The last weight, $w_{|\mathcal{V}|}(t)$, is $w_{|\mathcal{V}|}(t) = 1 - \sum_{\beta=1}^{|\mathcal{V}|-1} w_\beta(t)$.

We introduce dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ in relation to Eq. (14), alongside the augmented penalty parameters $\boldsymbol{\alpha}$:

$$\begin{aligned} \boldsymbol{\lambda} &= [\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3]^\top, \quad \boldsymbol{\mu} = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3, \boldsymbol{\mu}_4]^\top, \\ \boldsymbol{\alpha}_\lambda &= [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3]^\top, \quad \boldsymbol{\alpha}_\mu = [\boldsymbol{\alpha}_5, \boldsymbol{\alpha}_6, \boldsymbol{\alpha}_7, \boldsymbol{\alpha}_8]^\top \end{aligned} \quad (24)$$

Furthermore, we reformulate the equality and inequality constraints of Eq. (14) in a more compact manner:

$$\begin{aligned} \boldsymbol{\omega}_\lambda(t) &= \\ &\begin{bmatrix} \mathbf{M}_b(\mathbf{p}_{dis}(t) - \mathbf{p}_{ch}(t) + \ell_{ev}(t)) + \mathbf{M}_g \mathbf{g}^p(t) - \ell_d^p(t) \\ -\Re\{D(\mathbf{v}(t)(\mathbf{v}(t))^H \mathbf{Y}^H)\}, \\ \mathbf{M}_g \mathbf{g}^q(t) - \ell_d^q(t) - \Im\{D(\mathbf{v}(t)(\mathbf{v}(t))^H \mathbf{Y}^H)\}, \\ \text{soc}(t) - \text{soc}(t-1) + \frac{\Delta t}{E_{cap}} (\eta_{ch} \mathbf{p}_{ch}(t) - \frac{\mathbf{p}_{dis}(t)}{\eta_{dis}}) \end{bmatrix} \end{aligned} \quad (25)$$

$$\omega_\mu(t) = \begin{bmatrix} |\mathbf{K}\vartheta(t)| - s_{\max} \\ soc(t) - soc_{\max} \\ soc_{\min} - soc(t) \\ \sum_{\beta=1}^{|\mathcal{V}|-1} w_\beta(t) \leq 1 \end{bmatrix}_+ \quad (26)$$

where the last term in Eq. (26) is the convex hull. With the definition of Eq. (22), the augmented Lagrangian is:

$$\min_{\phi} \mathcal{L}_\phi = -Q_{\xi'_1}(s(t-1), \pi_\phi(s(t-1))) + \left(\lambda^\top \omega_\lambda(t) + \mu^\top \omega_\mu(t) + \left\| \begin{bmatrix} \text{diag}(\alpha_\lambda) & \mathbf{0} \\ \mathbf{0} & \text{diag}(\alpha_\mu) \end{bmatrix} \begin{bmatrix} \omega_\lambda(t) \\ \omega_\mu(t) \end{bmatrix} \right\|_2^2 \right) \quad (27)$$

where λ and μ are the dual variable vectors, and α_λ and α_μ are positive scalars that penalize the augmented terms.

Algorithm 1: CRL+Hull for OPF with aggregate EVs

```

1 Initialize critic network  $Q_{\xi_1}, Q_{\xi_2}$ , and actor network  $\pi_\phi$ 
  with random parameters  $\xi_1, \xi_2$  and  $\phi$ ;
2 Initialize target networks  $\xi'_1 \leftarrow \xi_1, \xi'_2 \leftarrow \xi_2, \phi' \leftarrow \phi$ ;
3 Initialize replay buffer  $\mathcal{B}$ , and set primal and dual update
  periods  $pu$  and  $du$ ;
4 for  $t = 1 : T$  do
5   Select action  $\mathbf{A}(t) \sim \pi_\phi(s(t-1))$ ;
6   Observe reward  $r(t)$  by Eq. (10);
7   Obtain the new state  $\mathbf{s}(t) = \text{env}(\mathbf{A}(t))$  by taking action
    $\mathbf{A}(t)$ ;
8   Store transition tuple  $(s(t-1), \mathbf{A}(t), r(t), s(t))$  in  $\mathcal{B}$ ;
9   Sample mini-batch of  $N$  transitions
    $\{(s^{n-1}, \mathbf{A}^n, r^n, s^n) | n = 1, \dots, N\}$  from  $\mathcal{B}$ ;
10   $y \leftarrow r(t) + \gamma \min_{i=1,2} Q_{\xi'_i}(s^n, \pi_{\phi'}(s^n))$ ;
11  Update critics:  $\xi_{i=1,2} \leftarrow \arg \min_{\xi_{i=1,2}} \frac{1}{N} \sum (y - Q_{\xi_{i=1,2}}(s^{n-1}, \mathbf{A}^n))^2$ ;
12  if  $t \bmod pu$  then
13    Update  $\phi$  by the deterministic policy gradient:
     $\phi \leftarrow \phi - \eta \nabla \mathcal{L}_\phi(s^{n-1}, s^n)$ , where  $\eta$  is the
    learning rate and  $\mathcal{L}_\phi$  is defined in Eq. (27);
14    Update target networks by [22];
15  if  $t \bmod du$  then
16    Update the dual variables by Eq. (28).
```

The primal dual update requires minimizing the Lagrangian function and then maximizing the dual function. With the definition of Eq. (22), the dual variables gradient update is:

$$\begin{aligned} \lambda^{k+1} &\leftarrow \lambda^k + \text{diag}(\alpha_\lambda) \omega_\lambda(t), \\ \mu^{k+1} &\leftarrow \mu^k + \text{diag}(\alpha_\mu) \omega_\mu(t), \end{aligned} \quad (28)$$

where λ^{k+1} and μ^{k+1} are updated by batch samples. We conduct the primal-dual update alternatively to optimize ϕ of $\pi_\phi(\cdot)$ while enforcing the feasibility of both equality and inequality constraints. The training is summarized in Algorithm 1.

VI. CASE STUDIES

For our numerical experiments we used the IEEE 14-bus system, equipped with three EV charging stations and battery systems at Buses 7, 8 and 12, with three wind power generation units and the IEEE 30-bus system, which has four EV charging stations and battery systems located at Buses 6, 12, 20, and 25, alongside another four wind power generation units. We used actual demand profiles from Texas, the wind

power profiles were based on data from NREL Wind [36], while the EV arrival pattern was simulated using the model in [27].

In the context of reinforcement learning we used PyTorch for training the CRL+Hull algorithm. We have set the buffer size at 500, with a discount factor for the reward of 0.99. Following each time step, the networks are trained on a mini-batch of 100 transitions that are uniformly sampled from the replay buffer. The number of sampled arrival matrices generated for the DP EV arrivals is 20. We select four vertices to approximate intersections within the reduced polyhedron. The actor network's architecture is given as follows. For both the IEEE 14-bus and IEEE 30-bus systems, the setup includes an initial cplx-STGCN layer for feature extraction, succeeded by a Complex-valued Neural Network (cplx-NN) layer, as mentioned in [25]. This is then followed by a Real-valued Neural Network (real-NN) layer for output creation. On the other hand, the critic network, which estimates the long-term discounted reward, features a straightforward design: a three-layer real-NN, each furnished with 256 neurons.

We evaluate our method against the following benchmarks for optimality and feasibility comparison:

- **MPC** [21]: Model Predictive Control that solves Eq. (14) at each time step using expected EV arrival data without scenario-based constraint considerations.
- **STC** [37]: Traditional stochastic optimization approach for EV charging without scenario-based constraint considerations.
- **OPT** [38]: Centralized optimization solver that computes the theoretical optimal solution for Eq. (14) with perfect information.
- **TD3** [32]: Twin Delayed DDPG implementation without constraint handling, using buffer size 500 and discount factor 0.99.
- **DDPG** [39]: Deep Deterministic Policy Gradient baseline employing unconstrained policy optimization.
- **DQN** [40]: Deep Q-Network implementation without constraint-aware mechanisms.

Benchmarks for different DP methods' impact comparison are listed as follows:

- **Gaussian DP** [41]: Adds Gaussian noise $\mathcal{N}(0, \sigma^2)$ to actual EV system states before transmission to the PSO, where σ is calibrated to achieve ϵ -DP guarantees for EV arrival data protection.
- **Laplacian DP** [42]: Applies Laplacian noise with scale parameter $b = \Delta f / \epsilon$ to EV arrival counts $a_x(t)$ to protect individual charging patterns while maintaining aggregate statistics for grid optimization.
- **Exponential DP** [43]: Uses the exponential mechanism to select perturbed EV states from neighboring possibilities with probability $\propto \exp(\epsilon \cdot \text{utility} / 2\Delta)$, where utility reflects the state's contribution to optimal charging coordination.

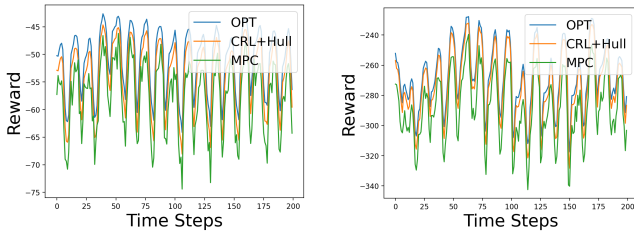
A. ϵ -DP, Optimality and Feasibility Validation

After training, we test our policy function with unseen cases. For the IEEE 14-bus system, the derived outcomes, illustrated

in Fig. 5(a), present the testing results of our CRL+Hull algorithm. This demonstrates that the results produced by our policy function are significantly better than those from MPC and show a consistent trend toward near-optimal outcomes, with an optimality gap of only 3.06%. Likewise, we scale the proposed algorithm to the IEEE 30-bus systems. For the training process, the difference between our DRL method and standard optimization is a only 2.15%. For the testing process, a curated selection from 200 sample outcomes is depicted in Fig. 5(b). The result implies our policy's approximation for near-optimal decisions, even in the absence of future EV arrival data and renewable energy generations, with 2.01% optimality gap. In comparison, the MPC method yields inferior results due to the difficulty in predicting EV demands. Besides, each time new data is collected, solving an optimization problem using the MPC method becomes necessary, which is very time-consuming. However, once trained, the CRL+Hull algorithm can produce near-optimal results within milliseconds, significantly enhancing efficiency and responsiveness.

1) *Validation of Feasibility:* In Fig. 6(a), we illustrate the feasibility validation of the proposed CRL+Hull algorithm, using voltage magnitude bounds as a specific example. While additional constraints are also satisfied, they are not shown due to page limitations. This figure highlights that other benchmark algorithms, such as DDPG, DQN, and TD3, which do not consider constraints, violate the voltage bound constraints. The proposed CRL+Hull algorithm effectively ensures that voltage magnitudes remain well within the boundaries.

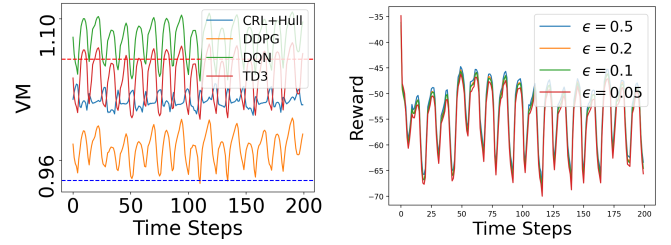
2) *Validation of Differential Privacy and Optimality:* Theorem 1 establishes ϵ -DP for all EV arrivals. In this part, we explore how varying levels of ϵ (representing different levels of DP) impact optimality. Notably, a smaller ϵ indicates stronger privacy. In Fig. 6(b), we show the tradeoff between privacy levels and optimality by selecting ϵ values of 0.5, 0.2, 0.1, and 0.05. These values demonstrate how optimal outcomes and aggregated EV demands vary. The findings indicate average optimal outcomes of -54.24 , -54.87 , -55.37 , and -55.68 , which are closely aligned. This suggests that privacy level has minimal impact on optimal outcomes.



(a) Testing curves of rewards (the 14-bus system).

(b) Testing curves of rewards (the 30-bus system).

Figure 5. Testing performance curves for CRL with EV demand responses in IEEE power systems. (a) IEEE 14-bus system: Reward values over 200 testing time steps (x-axis), where each time step represents a 15-minute interval, showing instantaneous reward values (y-axis, dimensionless) achieved by different control methods on unseen test cases. (b) IEEE 30-bus system: Similar testing reward trajectories over the same time horizon. The curves compare CRL+Hull (proposed method), MPC (Model Predictive Control), and other benchmark approaches against the theoretical optimal solution labeled as “OPT” (calculated through optimization methods). Higher reward values indicate better performance in minimizing fuel costs while maintaining grid constraints and coordinating EV charging with renewable energy sources.



(a) Voltage Magnitude Feasibility Validation. (b) Trade-off between Differential Privacy and Optimality.

Figure 6. Feasibility validation and differential privacy trade-off analysis. (a) Voltage magnitude validation over 200 time steps (15-minute interval per step): Per-unit voltage values (y-axis) showing CRL+Hull maintains bounds (0.95-1.05 p.u.) while unconstrained methods (DDPG, DQN, TD3) violate constraints. (b) Privacy-optimality trade-off across test cases: Reward values (y-axis) for different ϵ -DP levels ($\epsilon = 0.5, 0.2, 0.1, 0.05$), where smaller ϵ indicates stronger privacy. Results show minimal performance impact from privacy protection.

3) Demand Response of EV Controls under Uncertainty:

In this section, we underscore the advantages of employing demand response in EV control. By optimizing the use of wind power for EV charging, our primary objective is to substantially cut down on fuel costs. The performance of DRL without demand responses is depicted in Fig. 7. In this simulation, we set the DP level at $\epsilon = 0.1$. Clearly, the CRL+Hull integrated with demand responses yields higher rewards than its counterpart without them. Fig. 7 presents a comparative analysis between the projected aggregate EV demand incorporating demand response strategies and that without such strategies. The outcomes indicate that the policy function can effectively forecast actions: when wind power is abundant, it prioritizes charging more vehicles. In contrast, during low wind power periods, the policy directs a greater number of EVs towards a slacking mode rather than charging. Our tests on average fuel costs showed that with the implementation of CRL+Hull alongside EV demand response controls, we observed a 13.28% reduction in fuel expenses compared to setups without DRL demand response. This significant reduction in average fuel cost is attributed to the control policy's adeptness at harnessing renewable energy sources for EV consumption, consequently reducing fuel costs.

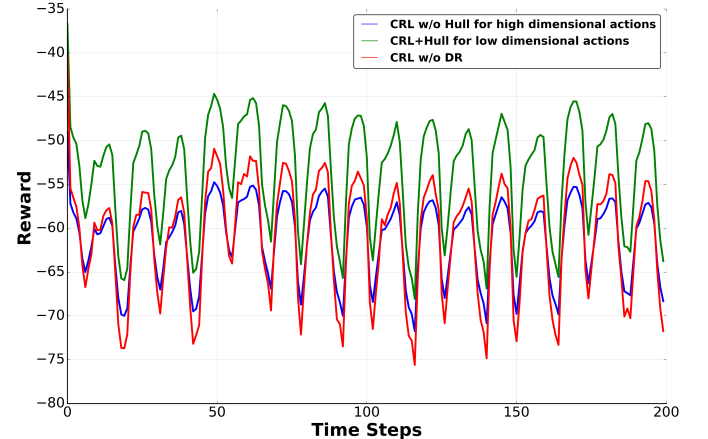
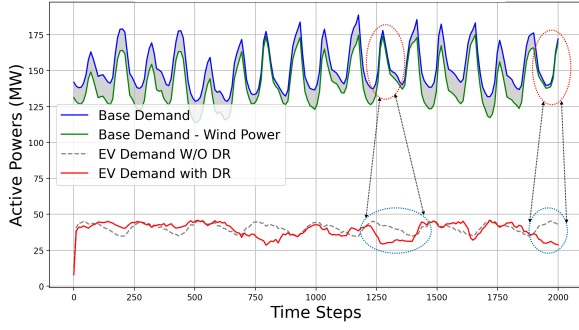


Figure 7. Comparative testing trajectories of different CRL variants: “CRL w/o Hull for high dimensional actions” represents the baseline method, “CRL+Hull for low dimensional actions” shows the method with action space reduction, and “CRL w/o DR” indicates CRL without demand response integration.



(a) Demand Curves without/with Demand Response.

Figure 8. Active power demand curves with and without demand response (DR) control over time. The x-axis shows time steps (each representing 15-minute intervals) and the y-axis shows power in MW. The curves display base demand (fixed load), wind power generation (shaded green areas), and EV charging demand with/without DR coordination. Red circles highlight periods of low wind power generation, which correspond to reduced EV charging power under DR control (indicated by blue circles), demonstrating how the intelligent DR policy adapts EV charging to renewable energy availability.

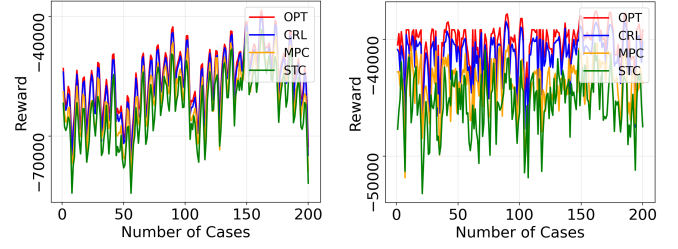
B. Scalability to Large Cases

To demonstrate the scalability of our CRL+Hull framework, we extend our evaluation to larger power systems: the IEEE 118-bus system and the IEEE 500-bus South Carolina synthetic system. These test cases validate our approach's ability to handle real-world scale deployments while maintaining computational efficiency and control performance.

1) *System Setting*: We evaluate our method on two power system testbeds. The IEEE 118-bus system features 9,966 MW generation capacity and 4,242 MW base demand, with approximately 30% of base demand (1,273 MW) allocated for EV charging across 12 charging hubs. This infrastructure supports roughly 177,000 EVs using Level 2 charging at 7.2 kW per vehicle. The renewable energy component consists of 8 wind power generation units positioned throughout the network, providing a combined renewable capacity of 800 MW. Wind power data are sourced from [NREL](#). Training requires 18 hours 32 minutes on average for 400,000 samples, while testing 3,000 samples takes 392.3 seconds.

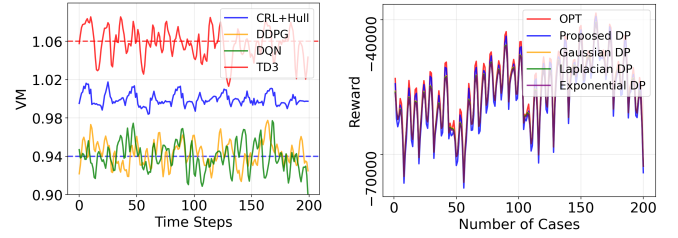
The larger IEEE 500-bus system operates with 12,189 MW generation capacity and 7,751 MW base demand, allocating approximately 30% of base demand (2,325 MW) for EV charging infrastructure that supports roughly 323,000 vehicles. This South Carolina synthetic system represents a realistic large-scale power grid with substantial geographic diversity, incorporating 25 EV charging hubs strategically positioned at major load centers. The renewable energy component comprises 20 wind farms with a combined capacity of 2,000 MW. To evaluate control system adaptability under renewable uncertainty, we implement wind farm profiles selected from the NREL dataset that exhibit high variability, with maximum ramping rates reaching 21.9% of base demand per time step. Training time averages 47 hours 57 minutes for 400,000 samples, while testing 3,000 samples requires 1,082.7 seconds.

2) *Testing Performance and Comparison*: Figure 9(a) shows the testing convergence results for both large-scale systems. For the IEEE 118-bus system, our CRL method demonstrates superior performance with an optimality gap of only 2.40% from the optimal solution, significantly outperforming both MPC (7.92% gap) and STC (12.64% gap).



(a) Testing curves of rewards (the 118-bus system). (b) Testing curves of rewards (the 500-bus system with more renewable generations).

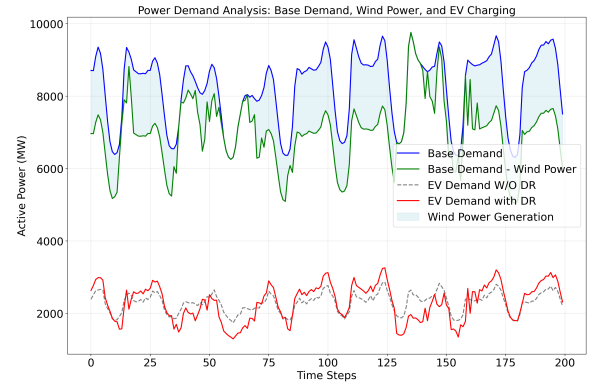
Figure 9. Testing curves for the CRL+Hull with EV demand responses in the IEEE 118-bus system and the South Carolina synthetic 500-bus system are illustrated in figures (a) and (b), respectively. The optimal rewards are calculated through the optimization method, labeled as "OPT". Compared with other benchmarks, including stochastic optimization and MPC, the proposed CRL+Hull method demonstrates superior performance.



(a) Voltage magnitude feasibility validation. (b) Comparison with differential privacy method.

Figure 10. Both figures are based on the IEEE 118-bus system. Figure (a) presents an example of feasibility validation for voltage magnitudes using the embedded convex hulls, demonstrating the constraint satisfaction capability of the proposed CRL method compared to traditional RL approaches. Figure (b) illustrates the trade-off between privacy protection and performance when applying different DP techniques to the CRL framework.

CRL achieves better performance than MPC in 98.0% of test cases and outperforms STC in 100.0% of cases. Similarly, the testing performance shown in Figure 9(b) demonstrates that our CRL+Hull algorithm maintains near-optimal performance under renewable energy fluctuations. For the 500-bus system, CRL achieves an optimality gap of 2.36% from the optimal solution, while MPC and STC show gaps of 8.36% and 12.47%, respectively. CRL outperforms MPC in 96.5% of test cases and consistently outperforms STC in 100.0% of cases.



(a) Demand Curves without/with Demand Response.

Figure 11. Total active power demand curves for the IEEE 500-bus system under high wind power penetration scenarios. The curves encompass base demands, wind power contributions, and EV demands with and without DR. The shaded areas represent wind power generation output. Red circles highlight periods of low wind power generation, which correspond to reduced EV charging power under DR control, as indicated by the blue circles.

3) *Demand Response under High-Penetration Renewable Energy*: Figure 11 illustrates how our CRL+Hull-based DR framework adapts EV charging patterns to renewable energy availability in the IEEE 500-bus system. The intelligent DR policy demonstrates remarkable adaptability: during periods of high wind generation, the system maximizes EV charging to absorb excess renewable energy, effectively using the EV fleet as a distributed energy storage resource. Conversely, when renewable output is low, the DR mechanism strategically reduces charging demand by transitioning EVs to idle states, preventing grid overload and minimizing dependence on costly fossil fuel generation. This renewable-aware demand response strategy achieves a substantial 14.5% reduction in overall fuel costs compared to conventional approaches without intelligent DR coordination, demonstrating the economic and environmental benefits of aligning EV charging with renewable energy patterns.

VII. CONCLUSION AND FUTURE WORK

In this work, we developed a control model that harmonizes EV charging DR seamlessly with power grid operations, leveraging the inherent flexibility of EV charging. We addressed privacy concerns by implementing DP mechanisms to prevent any potential EV drivers' data leakage to the PSO. However, the uncertainty introduced by the DP mechanism posed challenges, occasionally leading to infeasible solutions. To overcome this, our robust CRL strategy constrains the EV scheduling action to be within any stochastic potential arrival scenarios. To reduce the high-dimensional nature of the feasible action space we proposed the derivation of a low-order polyhedron to approximate the extensive scenario-based constraints. Through case studies using IEEE standard systems, we demonstrated our approach's capability of optimizing use of renewable resources for EV charging while simultaneously preserving the privacy of EV drivers. Several promising research directions emerge from our work. First, developing neural network-based approximations for feasible regions represents a natural extension that would complement our current scenario-based sampling approach. Such neural approximations of feasible regions induced by differential privacy mechanisms would provide alternative computational pathways while preserving the reinforcement learning framework. Second, we plan to extend our differentially private control framework to other power systems applications and cyber-physical systems (CPS), including smart grid management, IoT-enabled building energy management, autonomous vehicle coordination, and industrial control systems.

REFERENCES

- [1] M. M. Rana, "Iot-based electric vehicle state estimation and control algorithms under cyber attacks," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 874–881, 2019.
- [2] Y. Zhang, Z. Guo, F. Jiao, H. Li, X. Zhang, and A. Ukil, "Joint ev eco-charging strategy considering spatio-temporal flexibility in transportation-distribution networks," *IEEE Transactions on Transportation Electrification*, 2025.
- [3] H. J. Kim and M. K. Kim, "Data-driven virtual power plant bidding strategy in electricity markets integrating hybrid forecasting model and customized incentive demand response," *IEEE Internet of Things Journal*, 2025.
- [4] J. Qian, Y. Jiang, X. Liu, Q. Wang, T. Wang, Y. Shi, and W. Chen, "Federated reinforcement learning for electric vehicles charging control on distribution networks," *IEEE Internet of Things Journal*, 2023.
- [5] C. Dwork, "Differential privacy," in *International colloquium on automata, languages, and programming*. Springer, 2006, pp. 1–12.
- [6] X. Ye, G. Yang, H. Deng, P. Jie, R. Wu, and H. Jiang, "High-dimensional data release with local differential privacy under iot architecture," *IEEE Internet of Things Journal*, 2025.
- [7] X. Kong, L. Lu, and K. Xiong, "Privacy-preserving estimation of electric vehicle charging behavior: A federated learning approach based on differential privacy," *Internet of Things*, vol. 28, p. 101344, 2024.
- [8] S. Han, U. Topcu, and G. J. Pappas, "Differentially private distributed constrained optimization," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 50–64, 2016.
- [9] H. Wang, S. Chen, M. Li, C. Zhu, and Z. Wang, "Demand-driven charging strategy-based distributed routing optimization under traffic restrictions in internet of electric vehicles," *IEEE Internet of Things Journal*, vol. 11, no. 22, pp. 35 917–35 927, 2024.
- [10] Z. Ju and Y. Li, "Local differential privacy-based privacy-preserving data range query scheme for electric vehicle charging," *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 1, pp. 673–684, 2023.
- [11] T. Ding, S. Zhu, J. He, C. Chen, and X. Guan, "Differentially private distributed optimization via state and direction perturbation in multiagent systems," *IEEE Trans. Autom. Control*, vol. 67, no. 2, pp. 722–737, 2021.
- [12] L. Liu, Z. Xi, K. Zhu, R. Wang, and E. Hossain, "Mobile charging station placements in internet of electric vehicles: A federated learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24 561–24 577, 2022.
- [13] Z. Lin and J. Li, "Fedevcp: Federated learning-based anomalies detection for electric vehicle charging pile," *The Computer Journal*, vol. 67, no. 4, pp. 1521–1530, 2024.
- [14] S. Islam, S. Badsha, S. Sengupta, I. Khalil, and M. Atiquzzaman, "An intelligent privacy preservation scheme for ev charging infrastructure," *IEEE Trans. Ind. Inform.*, vol. 19, no. 2, pp. 1238–1247, 2022.
- [15] S. R. Pokhrel and M. B. Hossain, "Data privacy of wireless charging vehicle to grid (v2g) networks with federated learning," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp. 9032–9037, 2022.
- [16] Z. Teimoori, A. Yassine, and M. S. Hossain, "A secure cloudlet-based charging station recommendation for electric vehicles empowered by federated learning," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, pp. 6464–6473, 2022.
- [17] K. Hreinsson, A. Scaglione, M. Alizadeh, and Y. Chen, "New insights from the shapley-folkman lemma on dispatchable demand in energy markets," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4028–4041, 2021.
- [18] H. J. Kim and M. K. Kim, "New customized bidirectional real-time pricing mechanism for demand response in predictive home energy management system," *IEEE Internet of Things Journal*, vol. 11, no. 14, pp. 24 497–24 510, 2024.
- [19] G. G. Zanvetor, M. Fochesato, M. Casini, J. Lygeros, and A. Vicino, "A stochastic approach for ev charging stations in demand response programs," *Applied Energy*, vol. 373, p. 123862, 2024.
- [20] H. Karimianfard, "A robust optimization framework for smart home energy management: Integrating photovoltaic storage, electric vehicle charging, and demand response," *Journal of Energy Storage*, vol. 110, p. 115259, 2025.
- [21] B. Hermans, S. S. Walker, J. H. Ludlage, and L. Özkan, "Model predictive control of vehicle charging stations in grid-connected microgrids: An implementation study," *Applied Energy*, vol. 368, p. 123210, 2024.
- [22] T. Wu, A. Scaglione, and D. Arnold, "Constrained reinforcement learning for predictive control in real-time stochastic dynamic optimal power flow," *IEEE Trans. Power Syst.*, 2023.
- [23] T. Wu, A. Scaglione, A. P. Surani, D. Arnold, and S. Peisert, "Network-constrained reinforcement learning for optimal ev charging control," in *2023 IEEE SmartGridComm*.
- [24] D. Zhai, J. Liu, T. Zhang, J. Wang, H. Du, T. Liu, T. Wang, C. Zhang, J. Kang, and D. Niyato, "Epdh: An efficient and privacy-preserving electric charging scheme in the internet of robotic things," *IEEE Internet of Things Journal*, 2024.
- [25] T. Wu, A. Scaglione, and D. Arnold, "Complex-value spatio-temporal graph convolutional neural networks and its applications to electric power systems ai," *IEEE Trans. Smart grid*, 2023.
- [26] K. Hreinsson, A. Scaglione, and V. Vittal, "Aggregate load models for demand response: Exploring flexibility," in *2016 IEEE GlobalSIP*.
- [27] M. Alizadeh, A. Scaglione, J. Davies, and K. S. Kurani, "A scalable stochastic model for the electricity demand of electric and plug-in hybrid vehicles," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 848–860, 2013.
- [28] N. Ravi, A. Scaglione, S. Kadam, R. Gentz, S. Peisert, B. Lunghino,

- E. Levijarvi, and A. Shumavon, "Differentially Private K-Means Clustering Applied to Meter Data Analysis and Synthesis," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4801–4814, Nov. 2022.
- [29] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *JASA*, vol. 60, no. 309, pp. 63–69, 1965.
- [30] B. Ghazi, N. Golowich, R. Kumar, P. Manurangsi, and C. Zhang, "Deep learning with label differential privacy," *NeurIPS*, vol. 34, pp. 27 131–27 145, 2021.
- [31] F. D. McSherry, "Privacy integrated queries: an extensible platform for privacy-preserving data analysis," in *Proc. ACM SIGMOD Int. Conf. Manag. Data.*, 2009, pp. 19–30.
- [32] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *ICML*, 2018, pp. 1587–1596.
- [33] A. Bajpai, H. Verma, and A. Yadav, "Optimizing data aggregation and clustering in internet of things networks using principal component analysis and q-learning," *Data Science and Management*, vol. 7, no. 3, pp. 189–196, 2024.
- [34] Y. Chandak, G. Theodorou, J. Kostas, S. Jordan, and P. Thomas, "Learning action representations for reinforcement learning," in *International conference on machine learning*. PMLR, 2019, pp. 941–950.
- [35] J. Feng, Z. Ren, and W. Li, "Soft actor-critic combined with logic-based benders decomposition algorithm for monthly security constrained unit commitment under wind power uncertainty," *IEEE Transactions on Power Systems*, 2025.
- [36] C. Draxl, A. Clifton *et al.*, "The wind integration national dataset (wind) toolkit," *Applied Energy*, vol. 151, pp. 355–366, 2015.
- [37] J. Yu, J.-H. Ryu, and I.-b. Lee, "A stochastic optimization approach to the design and operation planning of a hybrid renewable energy system," *Applied energy*, vol. 247, pp. 212–220, 2019.
- [38] J. Rivera, C. Goebel, and H.-A. Jacobsen, "Distributed convex optimization for electric vehicle aggregators," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1852–1863, 2016.
- [39] F. Zhang, Q. Yang, and D. An, "Cddpg: A deep-reinforcement-learning-based approach for electric vehicle charging control," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3075–3087, 2020.
- [40] Y. Han, T. Li, and Q. Wang, "A dqn based approach for large-scale evs charging scheduling," *Complex & Intelligent Systems*, vol. 10, no. 6, pp. 8319–8339, 2024.
- [41] V. Dvorkin, F. Fioretto, P. Van Hentenryck, P. Pinson, and J. Kazempour, "Differentially private optimal power flow for distribution grids," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2186–2196, 2020.
- [42] M. Ryu and K. Kim, "A privacy-preserving distributed control of optimal power flow," *IEEE Transactions on Power Systems*, vol. 37, no. 3, pp. 2042–2051, 2021.
- [43] S. Gopi, Y. T. Lee, and D. Liu, "Private convex optimization via exponential mechanism," in *Conference on Learning Theory*. PMLR, 2022, pp. 1948–1989.