Optimizing Electric Vehicle Charging Through Determinization

Sandhya Saisubramanian and **Shlomo Zilberstein** and **Prashant Shenoy**
College of Information and Computer Sciences
University of Massachusetts Amherst
Amherst, MA 01003, USA
{saisubramanian, shlomo, shenoy}@cs.umass.edu

Abstract

We propose a determinization based approach to optimize the charging policies of an electric vehicle (EV) operating in a *vehicle-to-grid* (V2G) setting. By planning when to charge or discharge electricity from the vehicle, the long-term cost of operating the EV can be minimized, while being consistent with the owner's preferences. For an EV operating under price uncertainty caused by the dynamic pricing of electricity, this problem needs to be solved on-the-fly. Therefore, we model this problem as a *Stochastic Shortest Path* (SSP) problem and employ a determinization technique to solve it. Since it is hard to predict *a priori* the performance of a determinization method on a given problem, we introduce the notion of *Lossless Determinization* (LLD) that produces optimal action selection via determinization and present an approach that achieves lossless determinization by adjusting the cost of actions to account for the ignored outcomes. We also present *Approximate Lossless Determinization* (ALLD)—an effective method for approximating the cost of actions based on state features. We evaluate the performance of ALLD and demonstrate its effectiveness on a range of settings for the electric vehicle charging problem.

Introduction

Electric vehicles function primarily as consumers of electricity from the grid. However, recent developments in cyber-physical systems allow electric vehicles to act as both consumers and producers of electricity when connected to a smart grid. Specifically in the *Vehicle-to-Grid* (V2G) setting, connections are added to electric vehicles to allow the flow of electricity from the vehicles to the smart grid, thus enabling electric vehicles to act as consumers and producers of electricity (Guille and Gross 2009; Kempton and Letendre 1997).

The efficiency of an electric vehicle largely depends on its efficient battery charging schedule. Donadee and Ilic model the EV charging problem under price uncertainty as an MDP with continuous space of decision variables and solve it using stochastic dynamic programming. The price uncertainty is modeled using a Gaussian copula (Donadee and Ilic 2014). Donadee, Ilic, and Karabasoglu model the EV charging problem operating under price uncertainty with stochastic driver behavior as an infinite horizon average reward MDP. The price uncertainty is modeled using a Gaussian copula and the MDP is solved offline using the value iteration algorithm (Donadee, Ilic, and Karabasoglu 2014). Ruelens et al. consider the stochasticity in the arrival and departure time for a fleet of Plug-in Hybrid Vehicles (PHEVs) and optimize the charging schedule for the fleet, using approximate policy iteration to minimize the cost (Ruelens et al. 2012). Shi and Wong optimize the charging policies of an EV operating under price uncertainty using Q-learning technique (Shi and Wong 2011). Most researchers have focused on devising policies for EV charging in the traditional setting only (Donadee, Ilic, and Karabasoglu 2014; Sortomme and El-Sharkawi 2011; Vayá and Andersson 2012; Donadee and Ilic 2014; Ruelens et al. 2012). Since electric vehicles can charge and discharge electricity in a V2G setting, it is possible to exploit this feature to further minimize the long-term costs associated with battery charging (Ma et al. 2012).

Hence, our objective is to optimize the charging schedule for an electric vehicle that is parked and connected to a smart grid in a V2G setting. By planning when to buy or sell electricity, the EV can devise a robust schedule for charging and discharging that is consistent with the owner's preferences, while minimizing the long-term cost of operating the vehicle. This problem needs to be solved quickly and on-the-fly due to price uncertainty caused by the dynamic pricing of electricity. Hence, we model it as a *Stochastic Shortest Path* (SSP) problem.

Solving large SSPs is an active research area in automated planning. Among the different techniques for solving SSPs that have been explored, *determinization* has attracted significant interest because it greatly simplifies the problem and can quickly solve large SSPs on-the-fly. Determinization ignores the stochastic transitions, leverages efficient off-the-shelf classical planners to solve the corresponding deterministic problem, and uses online replanning when an unexpected state is encountered. Since the policies for the EV charging problem needs to be obtained on-the-fly, we solve the EV charging SSP using determinization.

While determinization could be extremely effective, it is often hard to predict when it will work particularly well as policies produced via existing determinization techniques do not guarantee bounded-optimal performance. Since the value of the deterministic policy is a loose lower bound on the optimal value of the SSP (i.e., the expected cost of reaching the goal), a large difference between the values may be indicative of the deviation of the optimal policy for the deterministic problem from the optimal policy for the SSP. We call this difference the *loss*. For example, in the EV charging problem, a suboptimal policy for the SSP yielded by determinization could be very expensive for the owner or could even lead to battery depletion at an unfavorable time. Therefore, it is beneficial to minimize the loss. When the loss is zero, it means that action values according to the deterministic policy match action values according to the optimal solution of the SSP, allowing for an easy derivation of the optimal actions. We examine the conditions under which this can be achieved and, more broadly, how to minimize the loss and thereby derive better policies.

To this end, our contributions in this paper are as follows,

1. We model the EV charging problem in a V2G setting as an SSP and consider different cost function scenarios;
2. We present the notion of *Lossless Determinization* (LLD) that requires the loss to be zero, and an approach called *Cost Adjustment for Lossless Determinization* (CALLD) that achieves zero loss by altering the costs of actions to account for the cost of ignored outcomes;
3. Naturally, it is challenging to accurately estimate the value discrepancy associated with each outcome and adjust the cost of each action without solving the original SSP. Hence, we propose an approximation technique to adjust the cost of actions in each state without calculating the true value of the outcomes. Specifically, in a factored MDP, the states are represented by feature vectors. The *Approximate Lossless Determinization* (ALLD) exploits the feature vector to derive an approximate cost for every action in a state;
4. We test the performance of ALLD on a range of settings for the EV charging problem.

We begin with a description of the model of EV charging as an SSP in Section 2. Section 3 defines determinization of an SSP and the notion of lossless determinization, and explains the CALLD algorithm for achieving zero loss. In Section 4 we describe our approach to approximating cost adjustments. Section 5 summarizes the performance of our approach in different settings of the EV charging domain.

The Model

By modeling the EV charging problem as an SSP, we aim to derive a sequence of actions that would minimize the long-term operational cost for an electric vehicle that is parked in a parking lot (parked and connected to a smart grid). On average, vehicles are parked for about 96% of the time (Kempton and Tomić 2005). Therefore, we restrict the decision process to the duration for which the vehicle is parked and connected to a smart grid. We begin with a formal background description of an SSP followed by a detailed explanation of modeling EV charging problem as an SSP.

Stochastic Shortest Path (SSP)

An SSP is a *Markov Decision Process* (MDP) with a start state and goal or terminal states, where the objective is to find a policy that minimizes the expected cost of reaching a goal state from the start state. A Stochastic Shortest Path MDP is denoted by the tuple $M = \langle S, A, T, C, s_0, S_G \rangle$, where,

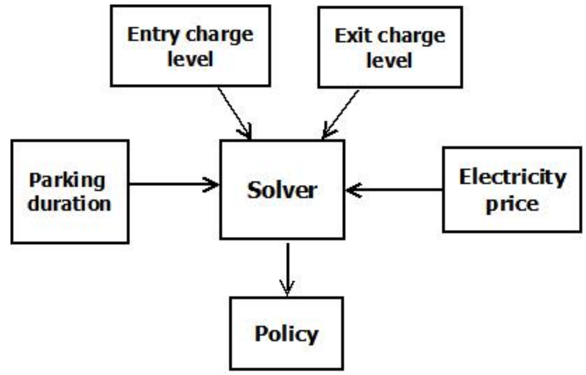

Figure 1: An illustration of EV charging problem

- S is a finite set of states;
- A is a finite set of actions with A_s denoting the set of actions available in state $s \in S$;
- T : $S \times A \times S \rightarrow [0, 1]$ is the transition function specifying the probability of moving to a state s' by executing an action $a \in A_s$ in state $s \in S$, denoted by $T(s, a, s')$;
- C : $S \times A \rightarrow \mathbb{R}^+ \cup \{0\}$ is the cost of executing action a in state s, denoted by $C(s, a)$. The cost of an action is positive in all states except goal states, where it is zero;
- s_0 is the initial state of the SSP, $s_0 \in S$; and
- S_G is the set of goal states of the SSP, $S_G \subseteq S$.

The solution of an SSP is a policy $\pi : S \rightarrow A$ that minimizes the expected cost of reaching a goal state. The Bellman equation defines a value function over states, $V^*(s)$, from which the optimal policy π^* can be extracted by:

$$V^*(s) = \min_a \; Q^*(s, a) \qquad \forall s \qquad (1)$$

$$Q^*(s, a) = C(s, a) + \sum_{s'} T(s, a, s') V^*(s') \quad \forall s, a \qquad (2)$$

where $Q^*(s, a)$ denotes the optimal Q-value of the action a in state s in the SSP M.

Modeling EV Charging Problem as an SSP

Since the decision process is restricted to the duration of parking for the electric vehicle, we model the EV charging problem as a finite horizon SSP with the parking duration as the horizon H. We assume that the vehicle can charge to a maximum limit (l_{max}) which is either the battery capacity (B_c) of the vehicle or some maximum threshold set by the vehicle owner, $0 < l_{max} \leq B_c$. Since we consider the electric vehicle in a V2G setting, we assume that the vehicle can discharge energy up to a minimum threshold level (l_g) which is either zero or some threshold set by the vehicle owner, $0 \leq l_g \leq B_c$. Assuming the parameters – l_{max}, l_g, H – are known, we can model this problem as an SSP with:

- S is the finite set of states that an electric vehicle can be in. It is defined by the tuple $\langle l, t, d, p \rangle$,where l denotes the current level of charge of the vehicle, $l \in [0, l_{max}]$, $t \in H$ denotes the current timestep, d denotes the current demand level for electricity, and p denotes the price distribution of electricity.

- A is the set of actions available to the vehicle. The vehicle can charge (Ch_i^+) and discharge (Ch_i^-) at three different speed levels, where i denotes the speed level, or remain idle (NOP). Therefore, there are seven actions in total, $A = \{Ch_1^+, Ch_2^+, Ch_3^+, Ch_1^-, Ch_2^-, Ch_3^-, NOP\}$. A_s denotes the set of actions available to the vehicle in state s. The charging and the discharging actions are stochastic, while the NOP action is deterministic.
- $T : S \times A \times S \rightarrow [0, 1]$ is the transition function denoted by $Pr(s'|s, a)$. It denotes the probability of reaching state s' by executing action a in state s. The state transition function also accounts for the demand level transitions and the pricing distribution transitions, as each state encapsulates the current demand level and current pricing distribution.
- $C : S \times A \rightarrow \mathbb{R}^+ \cup \{0\}$ is the cost function denoted by $C(s, a, s')$. It denotes the cost of executing action a in state s and reaching state s'. The costs for charging and discharging depend on the electricity pricing, and the speed setting. The cost for a NOP action is a constant. Based on real-world, the cost for discharging is modeled as a negative value (since the user profits by selling electricity), the cost for charging is modeled as a positive value (since the user has to pay for charging), and the cost for NOP action is modeled as zero.
- $s_0 \in S$ is the start state. It it defined by the tuple $s_0 = \langle l_0, t_0, d, p\rangle$, where $l_0 \in [0, l_{max}]$, and t_0 denote the charge level of the vehicle and the time when the vehicle is parked, respectively. t_0 also denotes the beginning of the decision process. d denotes the demand level at time t_0 and p denotes the price distribution at time t_0.
- $S_G \subseteq S$ is the set of goal states. It is denoted by the set of all states that match the tuple $\langle l, t_g, d, p\rangle$, where $l_g \leq l \leq l_{max}$, and t_g denotes the end of decision process when the vehicle is unplugged from the smart grid. d denotes the demand level at time t_g and p denotes the price distribution at time t_g.

The objective is to find a cost minimizing policy $\pi^* : S \rightarrow A$ that maximizes goal reachability, given the schedule of the vehicle owner. Figure 1 illustrates the EV charging problem.

Modeling Price Uncertainty

We consider four different types of cost functions that model the price uncertainty in a progressively more realistic way.

- Case 1 : The cost of discharging is the negation of the cost of charging and the costs are assumed to be known in advance. $C(s, Ch_i^-, s') = -C(s, Ch_i^+, s'), \forall i, \forall s, s' \in S$, where i denotes the speed level for charging or discharging.
- Case 2: The cost of discharging is the negation of the cost of charging plus a constant k. We again assume that the costs are known in advance. $C(s, Ch_i^-, s') = -C(s, Ch_i^+, s') + k, k \geq 0, \forall i, \forall s, s' \in S$, where i denotes the speed levels for charging and discharging. When $k = 0$, this case is the same as the previous case.
- Case 3: We assume that the cost of charging is known in advance, but the cost of discharging varies dynamically based on the actual level of demand. We assume that there is a known distribution for the demand level fluctuation based on the time of the day, with $P(d'|d, t)$ denoting the probability of demand d' at time t if the demand at time $t-1$ was d. $P(C(s, Ch_i^-, s') = r|t, d)$ denotes the probability of the discharge cost r for state s, given the demand d at time t.
- Case 4: We assume that the cost of charging is known in advance, but the cost of discharging varies dynamically based on the pricing distribution, which in turn depends on the current demand level. We assume that there is a known distribution for the demand level and for the pricing. For a price distribution p, $P(p|t, d)$ denotes the probability of the pricing distribution p at time t with demand level d, and $P(d'|d, t)$ denotes the probability of demand d' at time t if the demand at time $t - 1$ was d. Given the demand d and the pricing distribution p, $P(C(s, Ch_i^-, s') = r|t, d, p)$ denotes the probability of the discharge cost r for state s.

Generally SSPs are defined with non-negative costs as this would avoid any negative cost cycles. In a finite horizon SSP, negative costs cycles cannot be formed because it is not possible to transition to a state with a lower or equal time step from the current state. Since the EV charging problem described in this paper is modeled as a finite horizon SSP, the negative costs in the model would not lead to negative cost cycles.

Determinization of SSPs

A *determinization* yields a simplified variation of the SSP, with deterministic transition function that can be solved quickly using an off-the-shelf solver. Interest in determinization increased after the success of *FF-Replan* (Yoon, Fern, and Givan 2007) which won the 2004 IPPC, using the *Fast Forward* (FF) technique to generate fast deterministic plans (Hoffmann 2001). FF-Replan generates a deterministic version of the problem and solves it using FF. If an unexpected state is reached during plan execution, the process repeats with the unexpected state as the initial state, until a goal state is reached.

Following the success of FF-Replan, researchers have proposed various methods to improve determinization. Specifically, *Robust FF* (RFF) reduces the frequency of replanning by generating a plan for an envelope of states such that the probability of reaching a state outside the envelope is below some predefined threshold (Teichteil-Königsbuch, Kuter, and Infantes 2010). *HMDPP* generates plans with low probability of deviations using self-loop determinization and using a pattern database to avoid dead ends (Keyder and Geffner 2008). *FF-hindsight* uses hindsight optimization to approximate the value function of the MDP by sampling multiple deterministic futures that are solved using FF. These efforts resulted in a rich collection of determinization-based planning algorithms (Kolobov, Mausam, and Weld 2009; Yoon et al. 2008; 2010; Issakkimuthu et al. 2015; Keller and Eyerich 2011; 2012).

In the deterministic version of the SSP M, the start state s_0 and the goal states S_G are unaltered. Hence, the deterministic version M_d of the SSP M is denoted by the tuple $M_d = \langle S, A_d, T_d, C_d, s_0, S_G \rangle$, where, A_d is the finite set of actions (in this paper, $A_d = A$), $T_d : S \times A \rightarrow S$ denotes the deterministic transition function of M_d and $C_d : S \times A \rightarrow \mathbb{R}^+ \cup \{0\}$ specifies the cost function of the deterministic problem M_d. Conventional determinization techniques do not alter the cost function.

The optimal Q-value of M_d is computed as follows:

$$Q_d^*(s,a) = C_d(s,a) + V_d^*(T_d(s,a)) \quad \forall s \in S, a \in A_d. \tag{3}$$

In general, a determinization may introduce dead ends by ignoring an outcome that is crucial for goal reachability, making the goal unreachable in some states. However, it is possible to derive determinizations that preserve the goal reachability (for example, by using heuristics to devise the deterministic transition function that preserves goal reachability).

We define the loss of a determinization, l_d, as the *maximum* difference between the optimal Q-value of actions in the SSP M and the optimal Q-values in the determinized problem M_d.

Definition 1. *The* ***loss*** *associated with a determinization M_d of an SSP M is $l_d = \max\limits_{s,a} |Q^*(s,a) - Q_d^*(s,a)|$.*

This difference is treated as a loss as it could potentially cause the model to yield a policy that significantly deviates from the optimal policy. As this could affect the cost of the plan and goal reachability in many problems, it is beneficial to minimize the loss. However, conventional determinization techniques may have arbitrary non-zero loss, $l_d > 0$.

Lossless Determinization

In this section, we describe a simple technique to modify the cost function of M_d such that the loss is provably zero.

Definition 2. *A determinization is a* ***lossless determinization*** *(LLD) when the corresponding loss is zero: $l_d = 0$.*

Note that solving a lossless determinization of a given SSP guarantees that the selected actions are optimal for the SSP.

We can achieve zero loss, $l_d = 0$, by adjusting the cost of actions in each state. The costs are modified for every (s,a) pair to account for the values of the outcomes ignored by the determinization. Consequently, the optimal Q-values of M_d are equal to the optimal Q-values of M, reducing the loss to zero. The process of arriving at a lossless determinization by adjusting the cost of each action in every state is referred to as *Cost Adjustment for Lossless Determinization* (CALLD).

Algorithm 1 describes a cost modification technique that produces a lossless determinization. The input is the SSP and a deterministic transition function, and the output is the cost function for the determinized problem, C_d. Line 3 is the cost adjustment step, where $V^*(s')$ is the optimal value of the successor s' and $V^*(T_d(s,a))$ is the optimal value of the successor in the deterministic transition function in the SSP M. Since the cost adjustment in Algorithm 1 depends on the difference between outcome values, the costs yielded by CALLD algorithm may be negative. In general, in an infinite horizon SSP, negative costs may lead to negative cost cycles, affecting the validity of the SSP. Therefore, we discuss the necessary and sufficient conditions under which CALLD produces non-negative cost.

Algorithm 1: CALLD (M, T_d)

1 **foreach** $s \in S$ **do**
2   **foreach** $a \in A_s$ **do**
3         $C_d(s,a) \leftarrow \sum_{s'} \Big(T(s,a,s')V^*(s')\Big) + C(s,a) - V^*(T_d(s,a))$;
4   **end**
5 **end**
6 **return** C_d

Proposition 1. *The necessary and sufficient condition for having non-negative costs in a CALLD, $C_d(s,a) \geq 0$, is that the deterministic transition function chooses outcomes in M such that $Q^*(s,a) \geq V^*(T_d(s,a))$.*

Proof. We consider each one of the implication directions:

Case 1: $\Big(Q^*(s,a) \geq V^*(T_d(s,a))\Big) \implies \Big(C_d(s,a) \geq 0\Big)$
Assume $Q^*(s,a) \geq V^*(T_d(s,a))$. Using the definition of Q-values (Equation (2)), we get:

$$C(s,a) + \sum_{s'} \Big(T(s,a,s')V^*(s')\Big) - V^*(T_d(s,a)) \geq 0.$$

Substituting for $C_d(s,a)$ from Algorithm 1, we get $C_d(s,a) \geq 0$. Thus, $Q^*(s,a) \geq V^*(T_d(s,a))$ is a sufficient condition for non-negative cost in CALLD.

Case 2: $\Big(C_d(s,a) \geq 0\Big) \implies \Big(Q^*(s,a) \geq V^*(T_d(s,a))\Big)$
Assume $C_d(s,a) \geq 0$. Substituting for $C_d(s,a)$ from Algorithm 1,

$$C(s,a) + \sum_{s'} T(s,a,s')V^*(s') \geq V^*(T_d(s,a)).$$

Using the definition of Q-values, Equation (2), we get:

$$Q^*(s,a) \geq V^*(T_d(s,a)).$$

Hence, we conclude that the proposition holds. □

In the case of a finite horizon stochastic planning problem, negative cost cycles cannot be formed and therefore, satisfying the necessary and sufficient conditions for non-negative costs in a CALLD is non-mandatory.

Proposition 2. *CALLD produces a lossless determinization when a deterministic transition function that preserves the goal reachability is used.*

Proof. We need to show that given an SSP M and its determinization M_d that preserves the goal reachability, the optimal Q-values in M_d are equal to the Q-values in M, $Q^*_d(s,a) = Q^*(s,a)$, if the cost function $C(s,a)$ is modified to account for the outcomes ignored during determinization.

Substituting for $C_d(s,a)$ from Algorithm 1 in Equation (3) and using Equation (2) we get

$$Q^*_d(s,a) = Q^*(s,a) - V^*(T_d(s,a)) + V^*_d(T_d(s,a)).$$

Since the cost adjustment is performed for every (s,a) pair and the determinization preserves goal reachability, $V^*(T_d(s,a)) = V^*_d(T_d(s,a))$ and hence $l_d = 0$. □

Corollary 1. *There exists a lossless determinization for every SSP.*

Since it is possible to derive a goal reachability preserving determinization for every SSP and using CALLD would produce a lossless determinization (Proposition 2), it is possible to arrive at a lossless determinization (Definition 2) for any SSP.

Approximate Lossless Determinization

In many real-world problems, it is challenging to derive a complete cost adjusted lossless determinization of the problem without solving the SSP and this defeats the purpose of determinization. Therefore, we propose an approximation technique, referred to as *Approximate Lossless Determinization* (ALLD). An ALLD estimates the cost for each action in a state for a determinization of the SSP. We consider sampling and machine learning techniques for estimating the cost adjustment for a large SSP, which we will refer to as our target problem for simplicity. In this paper, the approximate cost adjustments for the target problem are learned from sampled small problems using a feature-based cost function.

Definition 3. *A **feature-based cost function** estimates the cost of an action in a state using the features of the state,* $C_d(s,a) = g(\vec{f}(s), a)$.

In a factored MDP, a state s is characterized by a set of features and these can be used to predict the cost of an action in the state. Let $\vec{f}(s) = \langle f_1(s), ..., f_n(s)\rangle$ be a set of features in a state s that significantly affect the cost of actions. Such features can be identified using machine learning techniques such as regression.

In order to estimate the feature-based approximate cost, sample problems are generated and solved. The costs adjustment values for the samples are computed in hindsight. The samples are obtained either from known small problem instances in the target domain or generated automatically by sampling states from the target problem. If the target problem has unavoidable dead ends, then sampling states may not be a good representative of the target problem. In such cases, smaller problem instances from the domain can be used. In this paper, smaller problems are created by multiple trials of depth limited random walk on the target problems and solved using LAO* (Hansen and Zilberstein 2001). The cost adjustments are computed for the samples using their exact solutions and the feature-based costs are learned. The learned values are projected onto the target problem using the feature-based cost function.

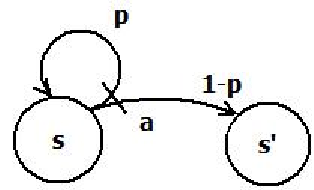

Figure 2: Example 1:Illustration

We also consider an extreme case, where the feature set characterizing each state is empty.

Definition 4. *A **state independent cost adjustment** assigns a* constant *cost adjustment per action, regardless of the state, resulting in a constant cost* $C_d(s,a) = g(a)$.

This simple form of generalization of the cost adjustment ignores the state altogether. In particular, PPDDL description of problems (Younes and Littman 2004) have a single description per action and hence having constant cost adjustment for actions in a problem instance can be extended to having constant cost adjustment for those actions in the various instances of problems in the domain.

For example, consider an action a and the set of states in which a is applicable. If the relative discrepancy between the values of the outcomes of a is the same in every state and the cost of a, $C(s,a)$, is the same in every state, then the cost adjustment can be trivially generalized with a state independent cost adjustment.

Although it is challenging to balance the trade-off between using state independent costs and solution quality, the following examples suggest that state independent cost adjustment could be effective.

Example 1. *Consider an SSP in which an action can achieve a successful outcome with probability* $1-p$ *or fail with probability* $p>0$. *When an action fails, the state remains unchanged. Let* s *denote a state of the SSP for which a successful execution of action* a *with cost* $C(s,a)$ *results in outcome state* s'. *Figure 2 is an illustration of this example.*

Proposition 3. *State independent cost adjustment produces zero loss for the class of problems identified in Example 1.*

Proof. In a goal reachability preserving determinization of the problem, the failure outcome would be ignored and the cost, $C_d(s,a)$, calculated by Algorithm 1 is,

$$C_d(s,a) = C(s,a) + \sum_{s'} \Big(T(s,a,s')V^*(s')\Big) - V^*(T_d(s,a)).$$

Since a fails with a probability p, we get

$$C(s,a) + \sum_{s'} \Big(T(s,a,s')V^*(s')\Big) = \frac{C(s,a)}{1-p} + V^*(T_d(s,a)).$$

Instance# ($\|S\|, \|A\|$)	%Charge (entry,exit)	Optimal Cost	Cost(Greedy)	Cost(MLO)	Cost(ALLD)	%DE (Greedy)	%DE(MLO)	%DE (ALLD)
P1 (909,7)	(80,20)	-7.60	-1.98 $\pm$ 0.07	-5.47 $\pm$ 0.02	-5.55 $\pm$ 0.01	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P2 (909,7)	(60,20)	-6.15	-1.93 $\pm$ 0.07	-4.42 $\pm$ 0.08	-4.52 $\pm$ 0.06	0 $\pm$ 0	21.40 $\pm$ 0.04	0 $\pm$ 0
P3 (909,7)	(30,40)	1.12	2.53 $\pm$ 0.07	1.86 $\pm$ 0.02	1.59 $\pm$ 0.01	0 $\pm$ 0	20.00 $\pm$ 0.20	16.00 $\pm$ 0.3
P4 (909,7)	(50,50)	-0.31	0 $\pm$ 0.01	0.09 $\pm$ 0.03	-0.62 $\pm$ 0.02	0 $\pm$ 0	20.00 $\pm$ 0.01	3.00 $\pm$ 0.02
P5 (909,7)	(30,60)	3.58	4.38 $\pm$ 0.05	4.12 $\pm$ 0.03	3.58 $\pm$ 0.03	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P6 (909,7)	(20,60)	4.12	5.58 $\pm$ 0.07	4.56 $\pm$ 0.03	4.33 $\pm$ 0.05	0 $\pm$ 0	23.05 $\pm$ 0.20	12.00 $\pm$ 0.3
P7 (909,7)	(40,90)	4.36	4.93 $\pm$ 0.06	4.86 $\pm$ 0.03	4.42 $\pm$ 0.05	0 $\pm$ 0	20.00 $\pm$ 0.02	15.00 $\pm$ 0.03

Table 1: Plan quality for seven instances of Electric Vehicle Charging- Cost Case 1

Instance# ($\|S\|, \|A\|$)	%Charge (entry,exit)	Optimal Cost	Cost(Greedy)	Cost(MLO)	Cost(ALLD)	%DE (Greedy)	%DE(MLO)	%DE (ALLD)
P1 (909,7)	(80,20)	-4.87	-1.55 $\pm$ 0.04	-2.72 $\pm$ 0.02	-3.17 $\pm$ 0.01	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P2 (909,7)	(60,20)	-3.97	-1.52 $\pm$ 0.05	-2.5 $\pm$ 0.02	-2.62 $\pm$ 0.02	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P3 (909,7)	(30,40)	1.16	3.06 $\pm$ 0.04	1.92 $\pm$ 0.06	1.54 $\pm$ 0.02	0 $\pm$ 0	1.00 $\pm$ 0.03	0.50 $\pm$ 0.03
P4 (909,7)	(50,50)	0.00	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P5 (909,7)	(30,60)	4.26	5.57 $\pm$ 0.08	4.63 $\pm$ 0.02	4.29 $\pm$ 0.02	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P6 (909,7)	(20,60)	4.44	5.54 $\pm$ 0.08	4.55 $\pm$ 0.09	4.46 $\pm$ 0.01	0 $\pm$ 0	3.00 $\pm$ 0.02	0.50 $\pm$ 0.02
P7 (909,7)	(40,90)	4.68	5.25 $\pm$ 0.03	4.85 $\pm$ 0.02	4.74 $\pm$ 0.02	0 $\pm$ 0	16.00 $\pm$ 0.02	1.10 $\pm$ 0.02

Table 2: Plan quality for seven instances of Electric Vehicle Charging- Cost Case 2

Substituting the above equation in the first equation, we get

$$\begin{aligned} C_d(s,a) &= \frac{C(s,a)}{1-p} + V^*(T_d(s,a)) - V^*(T_d(s,a)) \\ &= \frac{C(s,a)}{1-p}. \end{aligned} \tag{4}$$

Thus, the proposition illustrates a class of problems for which state independent cost (Equation 4) is perfectly accurate with *zero loss*. □

The above proposition highlights the potential benefits of ALLD that can achieve a perfectly accurate cost adjustment for problems such as the Blocksworld, that satisfies the required conditions. In the Blocksworld problem—an IPPC problem with stochastic actions— given an initial configuration of a collection of blocks, the blocks need to be rearranged to satisfy some goal conditions. Since the actions are stochastic, an action, for example, "pick block" may be successful or unsuccessful. If unsuccessful, the block slips and is dropped on the table and the action is repeated until it is successful. Since the relative discrepancy in the values of the outcomes is constant, a state independent cost adjustment that is accurate with zero loss is feasible. Consider the setting with unit cost actions that fail with a probability of 0.25. Our experiments show that regardless of the specific block, the state independent cost for this action is constant that matches the value of 1.33 obtained using Equation 4. This example illustrates the scope of generalized constant cost. However, not all domains satisfy this property. Identifying actions and domains that exhibit this property would alleviate the need for the preprocessing and help exploit the hidden structure in the given domain.

Experimental Results

The performance of ALLD is tested on four different cost function settings of EV. For each cost function, we consider seven different entry and exit level charges. In each cost function scenario, the charging costs are assumed to be known ahead of the decision process, and the costs associated with discharge actions may or may not be known ahead of time, depending on the cost function case. For the costs known in advance, we use the Time-of-Use (ToU) pricing (Eversource 2017). In case 3 of the cost function, we consider four demand levels– super off-peak, off-peak, mid-peak, and peak demand. In cost function case 4, we consider two pricing distributions – off-peak, and peak pricing levels, in addition to the four demand levels. The peak and the non-peak hours are based on real world peak hours and non-peak hours (Eversource 2017).

In all our experimental test cases, we consider an EV parked for a span of two hours and the duration of each timestep t is equivalent to 15 minutes in real time. The battery capacity and the three charge speed settings for the EV were selected based on Nissan Leaf EV configuration. We assume the discharge speeds to be the same as that of charge speeds. We also assume that the battery efficiency of the vehicle is not 100% and therefore, it may be required to buy more electricity from the grid than needed, and the electricity that reaches the grid during discharge would be lesser than the actual quantity discharged from the vehicle. We account for this battery inefficiency by adding a penalty of 15% to the current charging and discharging costs.

The goal in this domain is to devise a robust policy for EV charging such that the long-term operational cost of the vehicle is minimized, while being consistent with the

Instance# ($\|S\|, \|A\|$)	%Charge (entry,exit)	Optimal Cost	Cost(Greedy)	Cost(MLO)	Cost(ALLD)	%DE (Greedy)	%DE (MLO)	%DE (ALLD)
P1 (3636,7)	(80,20)	-4.60	-1.12 $\pm$ 0.02	-1.63 $\pm$ 0.04	-2.84 $\pm$ 0.03	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P2 (3636,7)	(60,20)	-2.60	-1.10 $\pm$ 0.01	-1.14 $\pm$ 0.02	-1.70 $\pm$ 0.01	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P3 (3636,7)	(30,40)	1.42	3.53 $\pm$ 0.03	1.62 $\pm$ 0.05	1.46 $\pm$ 0.02	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P4 (3636,7)	(50,50)	-0.94	0 $\pm$ 0	-0.02 $\pm$ 0.02	-0.32 $\pm$ 0.01	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P5 (3636,7)	(30,60)	2.49	4.20 $\pm$ 0.01	2.79 $\pm$ 0.02	2.49 $\pm$ 0.02	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P6 (3636,7)	(20,60)	3.28	4.52 $\pm$ 0.04	3.93 $\pm$ 0.02	3.72 $\pm$ 0.02	0 $\pm$ 0	5.00 $\pm$ 0.02	0 $\pm$ 0
P7 (3636,7)	(40,90)	3.47	4.86 $\pm$ 0.03	4.25 $\pm$ 0.04	3.70 $\pm$ 0.02	0 $\pm$ 0	15.00 $\pm$ 0.02	2.00 $\pm$ 0.02

Table 3: Plan quality for seven instances of Electric Vehicle Charging- Cost Case 3

Instance# ($\|S\|, \|A\|$)	%Charge (entry,exit)	Optimal Cost	Cost(Greedy)	Cost(MLO)	Cost(ALLD)	%DE (Greedy)	%DE (MLO)	%DE (ALLD)
P1 (7272,7)	(80,20)	-4.76	-3.22 $\pm$ 0.09	-3.30 $\pm$ 0.09	-3.52 $\pm$ 0.01	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P2 (7272,7)	(60,20)	-3.71	-3.08 $\pm$ 0.08	-3.21 $\pm$ 0.06	-3.38 $\pm$ 0.01	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P3 (7272,7)	(30,40)	1.10	1.78 $\pm$ 0.09	1.77 $\pm$ 0.03	1.34 $\pm$ 0.01	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P4 (7272,7)	(50,50)	-1.00	0 $\pm$ 0	-0.25 $\pm$ 0.04	-0.35 $\pm$ 0.02	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
P5 (7272,7)	(30,60)	2.90	3.79 $\pm$ 0.04	3.65 $\pm$ 0.03	3.07 $\pm$ 0.01	0 $\pm$ 0	10.00 $\pm$ 0.02	3.00 $\pm$ 0.02
P6 (7272,7)	(20,60)	3.29	5.09 $\pm$ 0.02	4.73 $\pm$ 0.03	3.91 $\pm$ 0.01	0 $\pm$ 0	20.00 $\pm$ 0.02	7.00 $\pm$ 0.02
P7 (7272,7)	(40,90)	3.59	4.84 $\pm$ 0.04	3.97 $\pm$ 0.02	3.59 $\pm$ 0.02	0 $\pm$ 0	12.80 $\pm$ 0.04	0 $\pm$ 0

Table 4: Plan quality for seven instances of Electric Vehicle Charging- Cost Case 4

owner's preferences. Any state from where the goal (exit level charge) cannot be reached in the remaining duration of the parking is treated as a dead end in the SSP.

A feature-based cost function is used to estimate the cost in ALLD in all our experiments. While ALLD requires a preprocessing step, estimating the costs is only required once per domain. The scalability of ALLD is preserved as we limit the size of the required sampled problems; in our experiments a depth of 4-8 was sufficient in most cases. The quality of the plan generated by ALLD is compared with:

- Quality of the plan generated by solving the *Most Likely Outcome* determinization of the SSP (MLO),
- Quality of the plan generated by using a greedy heuristic based on a naive human decision making.
- Optimal cost obtained by solving the SSP offline.

The value of the plan (average cost of reaching the goal) and fraction of dead end visits (%DE) are used as metrics to estimate the quality of the generated plan. Standard errors are reported for the value of the plan and the dead end visits based on 1000 trials per setting.

In MLO determinization and ALLD, the deterministic transition function chooses the most likely outcome for an action in a state. Since both the techniques share the determinisitic transition function, we use an optimal solver to solve the determinisitic problems, and to efficiently evaluate and compare the performance of ALLD. Therefore, the MLO determinization and ALLD are solved using the A* algorithm (Hart, Nilsson, and Raphael 1968), which is maximally efficient, and are complemented by replanning when necessary. It is assumed that the time taken for replanning is negligible.

Greedy heuristic We model a naive, and risk-averse human decision making as a simple greedy heuristic. Since most people prefer to ensure that the vehicle achieves the predefined exit charge over the monetary profits, we consider this as a risk-averse decision making. It is also considered naive and greedy because the decision is based only on the current state of the system. If the current charge of the vehicle is equal to the predefined exit charge for the vehicle, then the heuristic policy is to do NOP. If the current charge level is less than the exit charge, then the heuristic policy is to charge the vehicle at the maximum charge speed. If the current charge level is greater than the required exit charge, then the heuristic policy for that state is to discharge the electricity in medium speed as that would ensure profit for the vehicle without draining the battery quickly. Using these heuristic guidelines, a greedy policy can be devised for the EV charging.

Discussion The performance of ALLD in the different cost function scenarios is discussed below. The costs in the tables account for the expenses related to charging and profits from discharging.

In case 1, we assume that the cost of discharging is the negation of the cost of charging. The result of the 1000 trials are tabulated in table 1. In most cases, ALLD performed better than MLO and greedy approach, in terms of cost. In case 2, we assume that the cost of discharging is the negation of the cost of charging plus some non-zero constant which depends on the time and is known ahead of the decision process. The result of the 1000 trials are tabulated in table 2. In all our test cases, ALLD performed better than MLO and greedy approach, in terms of cost, with significantly lower dead end visits.

In case 3, we assume that the cost of charging is known in advance and is based on the Time-of-Use pricing. The discharging cost depends on the current demand level and we assume that the distribution is known. The result of the 1000 trials are tabulated in table 3. In most of our test cases, ALLD performed better than MLO and greedy approach, in terms of dead ends and cost. In case 4, we assume that the cost of charging is known in advance and is based on the Time-of-Use pricing. The discharging cost depends on the current demand level and the current pricing distribution. We assume that the demand and pricing distributions are known in advance. The result of the 1000 trials are tabulated in table 4. In all our test cases, ALLD performed better than MLO and greedy approach, in terms of cost, with significantly lower dead end visits.

For every test case in each case of the cost function discussed above, the greedy approach consistently avoids dead ends. This is an expected behavior as the greedy approach is conservative with respect to discharging electricity. However, the cost obtained by executing the greedy policy is much higher than the optimal, compared to the two determinization techniques. In cases where the entry charge is lower than the exit charge, the greedy policy is very expensive as it tries to charge as fast as possible to avoid dead end. However, this may be unnecessary if the parking duration is long enough. Also, the greedy policy does not consider the price variation which significantly affects the total cost.

Overall, while greedy approach is simple and easy, it is not effective when there is price variation with respect to time or in complicated settings such as the price depending on the current demand level. This reinforces the need for automated planning for EV charging under price uncertainty. ALLD always performs better than the greedy approach in terms of cost, and better than MLO determinization in terms of both cost and dead end visits, illustrating the potential of ALLD in achieving near-optimal policy for an EV operating under price uncertainty.

Conclusion and Future Work

Until recently, electric vehicles were primarily perceived as consumers of electricity from the smart grid. In the Vehicle-to-Grid (V2G) setting, electric vehicles can act as both consumers and producers of electricity. Using this feature, we aim to derive a charging policy for the EV that minimizes the long-term operational cost of the vehicle and that is consistent with the owner's preferences. Due to price uncertainty, this problem needs to be solved on-the-fly. Hence, we model this problem as a stochastic shortest path problem and employ determinization technique to solve it.

Since the policies yielded by conventional determinization techniques can significantly deviate from the optimal policy, we introduce the notion of lossless determinization that produces optimal action selection via determinization. We present cost adjustment for lossless determinization, an approach to achieve lossless determinization by adjusting the costs of actions in the deterministic problem. Since it is difficult to compute the exact cost adjustment without knowing the optimal values of the states, we propose approximation techniques to compute estimated costs. Our experiments show that ALLD can effectively use approximate costs to get better results than conventional determinization techniques.

The model presented in this paper aims to optimize the charging policies of an electric vehicle in a V2G setting. However, we do not consider the ancillary services like frequency regulation that an EV can offer. In future work, we plan to explore the role of determinization in general, and ALLD in particular, for other ancillary services of an EV in a V2G setting. While we assume that the duration of parking is known in advance, planning with stochastic parking duration is an interesting direction for the future work.

References

Donadee, J., and Ilic, M. D. 2014. Stochastic Optimization of Grid to Vehicle Frequency Regulation Capacity Bids. *IEEE Transactions on Smart Grid* 5(2):1061–1069.

Donadee, J.; Ilic, M.; and Karabasoglu, O. 2014. Optimal Autonomous Charging of Electric Vehicles with Stochastic Driver Behavior. In *Vehicle Power and Propulsion Conference (VPPC), 2014 IEEE*, 1–6. IEEE.

Eversource. 2017. Eversource Energy - Time of Use Rates. https://www.eversource.com/clp/vpp/vpp.aspx.

Guille, C., and Gross, G. 2009. A conceptual framework for the vehicle-to-grid (v2g) implementation. *Energy policy* 37(11):4379–4390.

Hansen, E. A., and Zilberstein, S. 2001. LAO*: A Heuristic Search Algorithm that Finds Solutions with Loops. *Artificial Intelligence* 129(1):35–62.

Hart, P. E.; Nilsson, N. J.; and Raphael, B. 1968. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics* 4(2):100–107.

Hoffmann, J. 2001. FF: The Fast-Forward Planning System. *AI Magazine* 22(3):57.

Issakkimuthu, M.; Fern, A.; Khardon, R.; Tadepalli, P.; and Xue, S. 2015. Hindsight Optimization for Probabilistic Planning with Factored Actions. In *Proceedings of the 25th International Conference on Automated Planning and Scheduling*, 120–128.

Keller, T., and Eyerich, P. 2011. A Polynomial All Outcome Determinization for Probabilistic Planning. In *Proceedings of the 21st International Conference on Automated Planning and Scheduling*, 331–334.

Keller, T., and Eyerich, P. 2012. PROST: Probabilistic planning based on UCT. In *Proceedings of the 22nd International Conference on Automated Planning and Scheduling*.

Kempton, W., and Letendre, S. E. 1997. Electric vehicles as a new power source for electric utilities. *Transportation Research Part D: Transport and Environment* 2(3):157–175.

Kempton, W., and Tomić, J. 2005. Vehicle-to-grid power fundamentals: Calculating capacity and net revenue. *Journal of power sources* 144(1):268–279.

Keyder, E., and Geffner, H. 2008. The HMDP Planner for Planning with Probabilities. In *Proceedings of the International Planning Competition (IPC 2008)*.

Kolobov, A.; Mausam; and Weld, D. S. 2009. ReTrASE: Integrating Paradigms for Approximate Probabilistic Planning. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, 1746–1753.

Ma, Y.; Houghton, T.; Cruden, A.; and Infield, D. 2012. Modeling the Benefits of Vehicle-to-Grid Technology to a Power System. *IEEE Transactions on power systems* 27(2):1012–1020.

Ruelens, F.; Vandael, S.; Leterme, W.; Claessens, B. J.; Hommelberg, M.; Holvoet, T.; and Belmans, R. 2012. Demand Side Management of Electric Vehicles with Uncertainty on Arrival and Departure Times. In *Innovative Smart Grid Technologies (ISGT Europe), 2012 3rd IEEE PES International Conference and Exhibition on*, 1–8. IEEE.

Shi, W., and Wong, V. W. 2011. Real-Time Vehicle-to-Grid Control Algorithm under Price Uncertainty. In *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, 261–266. IEEE.

Sortomme, E., and El-Sharkawi, M. A. 2011. Optimal charging strategies for unidirectional vehicle-to-grid. *IEEE Transactions on Smart Grid* 2(1):131–138.

Teichteil-Königsbuch, F.; Kuter, U.; and Infantes, G. 2010. Incremental Plan Aggregation for Generating Policies in MDPs. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, 1231–1238.

Vayá, M. G., and Andersson, G. 2012. Smart charging of plug-in vehicles under driving behaviour uncertainty. In *12th International Conference on Probabilistic Methods Applied to Power Systems*, 10–14.

Yoon, S. W.; Fern, A.; Givan, R.; and Kambhampati, S. 2008. Probabilistic Planning via Determinization in Hindsight. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, 1010–1016.

Yoon, S.; Ruml, W.; Benton, J.; and Do, M. B. 2010. Improving Determinization in Hindsight for On-line Probabilistic Planning. In *Proceedings of the 20th International Conference on Automated Planning and Scheduling*, 209–216.

Yoon, S. W.; Fern, A.; and Givan, R. 2007. FF-Replan: A Baseline for Probabilistic Planning. In *Proceedings of the 17th International Conference on Automated Planning and Scheduling*, 352–359.

Younes, H. L. S., and Littman, M. L. 2004. PPDDL1.0: The Language for the Probabilistic Part of IPC-4. In *Proceedings of the International Planning Competition*.