Stochastic Control with Distributionally Robust Constraints for Cyber-Physical Systems Vulnerable to Attacks

Nishanth Venkatesh[1], *Student Member, IEEE*, Aditya Dave[2], *Member, IEEE*,
Ioannis Faros[1], *Student Member, IEEE*, Andreas A. Malikopoulos[1,2], *Senior Member, IEEE*

***Abstract*— In this paper, we investigate the control of a cyber-physical system (CPS) while accounting for its vulnerability to external attacks. We formulate a constrained stochastic problem with a robust constraint to ensure robust operation against potential attacks. We seek to minimize the expected cost subject to a constraint limiting the worst-case expected damage an attacker can impose on the CPS. We present a dynamic programming decomposition to compute the optimal control strategy in this robust-constrained formulation and prove its recursive feasibility. We also illustrate the utility of our results by applying them to a numerical simulation.**

I. Introduction

Cyber-physical systems (CPSs) have enabled highly efficient control of physical processes by tightly coupling sensing, communication, and computational processing to generate real-time decisions with classical [1] and nonclassical information structures [2]. They span various important applications including, but not limited to, connected and automated vehicles [3], [4], Internet of Things [5], and social media platforms [6]. However, in each of these applications, the interplay between the cyber components and the physical world can make the system vulnerable to various security threats, e.g., control system malware [7] and staged attacks [8]. This has led to many studies on controlling CPSs while ensuring robustness and resilience to attacks [9], [10].

The common modeling framework for CPSs utilizes a stochastic formulation to account for uncertainties in the dynamics that arise within the evolution of the physical process. In this formulation, an agent is assumed to have access to a prior distribution for all uncertainties and must compute a control strategy to generate real-time control actions that minimize the total expected cost [11], [12]. In stochastic formulations [13], constraints on the state and actions are modeled as probabilistic constraints, which can be imposed either in expectation or with some probabilities [14]. Similarly, approaches like those reported in [15], [16] consider probabilistic constraints on the cumulative reward. However, the actual performance and constraint satisfaction of an optimal control strategy are very sensitive to changes in a mismatch between the assumed prior probability model and the actual model [17], [18]. Such a mismatch is bound to occur when a CPS is under attack from an adversary. Thus, it may not be appropriate to model safety-critical requirements on system behavior using probabilistic constraints in a stochastic formulation.

To accommodate the needs of safety-critical systems, several research efforts [19]–[22] have explored minimax formulations. Similar approaches [23], [24] consider non-stochastic formulations in which the agent does not have knowledge about the distributions of uncertainties and uses only the set of feasible values to compute optimal strategies that minimize the maximum costs. Though such approaches are suitable for applications under attack, such as cyber-security [25], and power systems [26], during regular operation of systems without attacks, they lead to outcomes that are overly conservative [27]. Consequently, there remains a need for alternative approaches to controlling vulnerable systems that avoid overly conservative decision-making during regular system operation and maintain a level of reliability when the system is occasionally attacked.

In this paper, we combine the superior performance of stochastic formulations in achieving an objective and the safety guarantees of worst-case formulations in minimizing vulnerabilities. To this end, we impose a distributionally robust constraint on a secondary objective that accounts for the vulnerabilities of a CPS to an attack. Concurrently, we aim to minimize the expected value of a primary cost for the best performance over a finite horizon. Our formulation generalizes the previous work of [28], which addressed the problem of minimizing an expected discounted cost subject to either an expected or a minimax constraint. By considering a distributionally robust constraint, our formulation allows for greater control over the trade-off between conservativeness and optimality by appropriately adjusting the size of the uncertainty set for probability distributions. In the extreme case that the set of feasible distributions is a singleton, we recover an expected value constraint. In contrast, if we expand the set to allow every possible distribution on the state space, we recover the non-stochastic worst-case constraint as a special case. Thus, by changing the set of feasible distributions, we can better select the level of conservativeness of our formulation.

Our main contributions in this paper are (1) the problem formulation of controlling a vulnerable CPS using a stochastic cost and distributionally robust constraint (Problem 1), (2) a dynamic programming (DP) decomposition for this problem, which computes the optimal strategy that ensures recursive feasibility of the constraint (Theorem 1), and (3) the

This research was supported by NSF under Grants CNS-2149520 and CMMI-2219761.

[1]Department of Systems Engineering, Cornell University, Ithaca, NY 14850 USA.

[2]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853 USA (email: ns942@cornell.edu; a.dave@cornell.edu; if74@cornell.edu; amaliko@cornell.edu).

illustration of the utility of our results by comparing them to both stochastic and worst-case approaches in numerical simulation (Section IV).

The remainder of the paper proceeds as follows. In Section II, we formulate the problem. In Section III, we present the DP decomposition. In Section IV, we demonstrate our results in a numerical example, and in Section V, we draw concluding remarks.

II. Model

We consider a CPS whose evolution is described by a finite Markov decision process (MDP), denoted by a tuple $(\mathcal{X}, \mathcal{U}, n, P, c, c_n)$, where $\mathcal{X}$ is a finite state space and $\mathcal{U}$ is a finite set of feasible actions available to an agent seeking to control the MDP. The system evolves over discrete time steps denoted by $t = 0, \ldots, n$, where $n \in \mathbb{N}$ is the finite time horizon. The state of the system and the control action of the agent at each t are denoted by the random variables X_t and U_t, respectively. The transition function at each t is denoted by $P_t : \mathcal{X} \times \mathcal{U} \times \mathcal{X} \to \Delta(\mathcal{X})$, where $\Delta(\mathcal{X})$ is the set of all probability distributions on the state space $\mathcal{X}$. Nominally, the transition function is given by $P_t = \bar{P}$ for all t, where $\bar{P} \in \Delta(\mathcal{X})$. For the realizations $x_t \in \mathcal{X}$ and $u_t \in \mathcal{U}$ of the state X_t and the control action U_t, the probability of transitioning to a state $x_{t+1} \in \mathcal{X}$ is $\mathbb{P}(X_{t+1} = x_{t+1} \,|\, x_t, u_t) = P_t(x_{t+1} \,|\, x_t, u_t)$. The agent selects the action using a control law $g_t : \mathcal{X} \to \mathcal{U}$ as $U_t = g_t(X_t)$, where g_t is chosen from the feasible set of control laws at time t, denoted as $\mathcal{G}_t$. The tuple of control laws denotes the control strategy of the agent $\boldsymbol{g} := (g_0, \ldots, g_{n-1})$, where $\boldsymbol{g} \in \mathcal{G}$ and $\mathcal{G} = \prod_{t=0}^{n-1} \mathcal{G}_t$. After selecting the action at each $t = 0, \ldots, n-1$, the agent incurs a cost $c(X_t, U_t)$ generated using the function $c : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$. Then, the performance of a strategy $\boldsymbol{g}$ is measured by the total expected cost beginning at an initial state $x_0 \in \mathcal{X}$:

$$\mathcal{J}_0(\boldsymbol{g}; x_0) = \mathbb{E}^{\boldsymbol{g}}\left[\sum_{t=0}^{n-1} c(X_t, U_t) + c_n(X_n) \,\Big|\, x_0\right], \tag{1}$$

where $c_n : \mathcal{X} \to \mathbb{R}$ is the terminal cost, and $\mathbb{E}^{\boldsymbol{g}}$ denotes the expectation on all the random variables with respect to the probability distributions generated by the choice of control strategy $\boldsymbol{g}$.

In the context of a CPS, the conventional approach of selecting a control strategy $\boldsymbol{g}$ to minimize the total expected cost (1) may not be adequate to ensure smooth operation, particularly when the CPS is vulnerable to attacks by an adversary. We consider that the presence or absence of an adversary during the system's operation is determined at the onset; however, this information is unknown to the agent. The adversary's influence on the system's dynamics results in a change in the transition probability at each $t = 0, \ldots, n-1$ from a known set $\mathcal{P} \subseteq \Delta(\mathcal{X})$. Thus, an attack may be reflected by the choice of the worst transition function from $\mathcal{P}$. We allow the adversary to attack the system with access to the realization of the state $x_t \in \mathcal{X}$ and action $u_t \in \mathcal{U}$. Note that the nominal transition function $\bar{P}$ belongs to the set $\mathcal{P}$ to allow for the case of no attack.

An agent that observes the presence or absence of an adversary can select either a purely robust or risk-neutral formulation, depending on the current situation. However, a risk-neutral formulation may involve an arbitrarily large risk for the agent and leave the CPS vulnerable during an attack. In contrast, a robust formulation may be too conservative for the majority of situations where no attack occurs. Thus, we impose a robust constraint to *limit* the worst-case damage possible during an attack while minimizing the expected total cost. To this end, the agent incurs a constraint penalty $d(X_t, U_t) \in \mathbb{R}$ at each $t = 0, \ldots, n-1$. The total expected worst-case penalty is given by

$$\mathcal{L}_0(\boldsymbol{g}; x_0) = \max_{P_{0:n-1} \in \mathcal{P}^n} \mathbb{E}^{\boldsymbol{g}}_{P_{0:n-1}}\left[\sum_{t=0}^{n-1} d(X_t, U_t) + d_n(X_n) \,\Big|\, x_0\right], \tag{2}$$

where $d_n : \mathcal{X} \to \mathbb{R}$ is the terminal penalty, $P_{0:n-1}$ is the collection of transition functions for $t = 0, \ldots, n-1$, each taking values in the set $\mathcal{P}$. Note that this penalty has a distributionally robust form where the attacker may select the worst transition function $P_t \in \mathcal{P}$ at each t. Furthermore, the choice of a particular function at any time t does not limit the functions available to the adversary at time $t+1$ in (2). The distributionally robust constraint is formulated by defining an upper bound $l_0 \in \mathbb{R}$, on the worst-case total expected penalty.

Remark 1. During an attack, the agent may prioritize a different property, e.g., safety, of the system rather than the total expected cost used in (1). Hence, the constraint penalty at each instance of time is considered to be distinct from the cost. However, if we seek to limit the influence of the adversary on the performance itself, the penalty in the constraint can be set equal to the cost at each t.

Next, we define the agent's constrained control problem.

Problem 1. The optimization problem is to compute the optimal control strategy $\boldsymbol{g}^* \in \mathcal{G}$, if one exists, subject to a constraint on (2), i.e.,

$$\min_{\boldsymbol{g} \in \mathcal{G}} \mathcal{J}_0(\boldsymbol{g}; x_0), \tag{3}$$

$$\text{s.t.} \quad \mathcal{L}_0(\boldsymbol{g}; x_0) \leq l_0, \tag{4}$$

for a given MDP $(\mathcal{X}, \mathcal{U}, n, P, c, c_n)$, penalty functions (d, d_n), set of transition functions $\mathcal{P}$, upper bound $l_0 \in \mathbb{R}$, and initial state $x_0 \in \mathcal{X}$.

We impose the following assumptions on our formulation:

Assumption 1. The costs and penalties at each instance of time are upper bounded by the finite maximum values $c^M \in \mathbb{R}$ and $d^M \in \mathbb{R}$, respectively. They are also lower bounded by the finite minimum values $c^m \in \mathbb{R}$ and $d^m \in \mathbb{R}$, respectively.

Assumption 1 ensures that the expected total cost (1) and robust total penalty (2) are finite for any value of $n \in \mathbb{N}$.

Assumption 2. The bound $l_0 \in \mathbb{R}$ is such that the set $\mathcal{G}_{l_0} := \{\boldsymbol{g} \in \mathcal{G} \,|\, \mathcal{L}_0(\boldsymbol{g}; x_0) \leq l_0\}$ is not empty.

Assumption 2 ensures that Problem 1 has a feasible solution and, thus, it is well-posed. Our goal is to efficiently compute an optimal solution to Problem 1 without violating the constraint. Next, we present a DP decomposition for the problem.

III. Dynamic Programming Decomposition

In this section, we present the value functions that constitute a DP decomposition to compute the optimal control strategy $\boldsymbol{g}^*$ for Problem 1. To show that the computed strategy satisfies the distributionally robust constraint (4), we need to prove its recursive feasibility for all $t = 1, \dots, n-1$. To achieve this, in Subsection III-A, we define the *penalty-to-go function* to express the application of the constraint only from any time t to the terminal time n. We then construct a set of upper bounds on the penalty-to-go function at any $t = 0, \dots, n-1$, such that these bounds admit a feasible solution, and present a methodology to compute these sets. We also introduce the notion of bound functions, which will be utilized it to ensure recursive feasibility. In Subsection III-B, we use bound functions within the proposed DP decomposition and prove its optimality.

A. Feasible bound for robust constraint

We begin by constructing the *penalty-to-go* function that maps each realization of the state $x_t \in \mathcal{X}$ at any $t = 0, \dots, n-1$ to an expected worst-case penalty to reach n using a sequence of control laws $g_{t:n-1} \in \prod_{\ell=t}^{n-1} \mathcal{G}_\ell$. Specifically, this penalty-to-go at each t is

$$\mathcal{L}_t(g_{t:n-1};x_t) = \max_{P_{t:n-1}\in\mathcal{P}^{n-t}} \mathbb{E}^{g_{t:n-1}}_{P_{t:n-1}}\left[\sum_{\ell=t}^{n-1} d(X_\ell,U_\ell) + d_n(X_n) \,\middle|\, x_t\right], \quad (5)$$

where the expectation on all the random variables is with respect to the distributions generated by the choice of control laws $g_{t:n-1} \in \prod_{\ell=t}^{n-1} \mathcal{G}_\ell$. Importantly, the control laws utilized prior to time t do not influence the penalty-to-go from time t. Additionally, note that the penalty-to-go from $t = 0$ is the total expected penalty in (2). Next, we construct a set of feasible upper bounds on the penalty-to-go function.

Definition 1. For all $t = 1, \dots, n-1$, the *set of feasible upper bounds* for a state $x_t \in \mathcal{X}$ is

$$\Lambda_t(x_t) := \left\{ l_t \in \mathbb{R} \mid \exists\; g_{t:n-1} \in \prod_{\ell=t}^{n-1} \mathcal{G}_\ell, \text{ s.t. } \mathcal{L}_t(g_{t:n-1}, x_t) \leq l_t \right\}, \quad (6)$$

with $\Lambda_n(x_n) := [d_n(x_n), d^M]$ at $t = n$ for each $x_n \in \mathcal{X}$ and $\Lambda_0(x_0) := \{l_0\}$ identically for all $x_0 \in \mathcal{X}$.

In Definition 1, the bound l_t acts only upon the penalty-to-go $\mathcal{L}_t(g_{t:n-1}; x_t)$. Thus, each bound $l_t \in \Lambda_t(x_t)$ ensures feasibility of only the control laws $g_{t:n-1} \in \prod_{\ell=t}^{n-1} \mathcal{G}_\ell$ for each $x_t \in \mathcal{X}$ and $t = 0, \dots, n-1$.

Next, to ensure recursive feasibility in our solution approach, our goal is to select a feasible bound on the penalty-to-go for all $t = 0, \dots, n-1$. These bounds should ensure that, starting with l_0 at $t = 0$, there exists at least one feasible sequence of control laws $g_{t:n-1} \in \prod_{\ell=t}^{n-1} \mathcal{G}_\ell$. We note that based on Assumption 2, such a sequence exists at $t = 0$.

To this end, we establish the notion of *bound functions* $\lambda_t : \mathcal{X} \to \mathbb{R}$ at each $t = 0, \dots, n$. The output of the bound function $\lambda_t(x_t)$ is a feasible bound from Definition 1 for all $x_t \in \mathcal{X}$ and all t. Then, for any bound $l_t \in \Lambda_t(x_t)$ and a control action $u_t \in \mathcal{U}$, the set of *recursively consistent bound functions* at time $t+1$ is

$$\begin{aligned} F_t(x_t, u_t, l_t) = \Big\{ \lambda_{t+1} \,\Big|\, & \lambda_{t+1}(x_{t+1}) \in \Lambda_{t+1}(x_{t+1}), \\ & \forall x_{t+1} \in \mathcal{X} \text{ and} \\ & \max_{P_t \in \mathcal{P}} \mathbb{E}_{P_t}[\lambda_{t+1}(X_{t+1}) \,|\, x_t, u_t] \leq l_t - d(x_t, u_t) \Big\}. \end{aligned} \quad (7)$$

The inequality in the conditioning of the set in (7) yields the allowable bound at time $t+1$ after considering the "consumption" of the bound l_t by the penalty $d(x_t, u_t)$ incurred at time t. This inequality is imposed upon the maximum expected value of $\lambda_{t+1}(X_{t+1})$ given the state x_t and action u_t to ensure recursive constraint satisfaction. Note that this maximization captures the distributionally robust form of transition functions in (5) and, thus, accounts for the possible influence of the attacker. Thus, given $l_t \in \Lambda_t(x_t)$ at time t, restricting attention to $\lambda_{t+1} \in F_t(x_t, u_t, l_t)$ ensures that any selected bound at time $t+1$ is feasible. Beginning with l_0 at $t = 0$ and applying this property for the set $F_t(x_0, u_0, l_0)$ for all $x_0 \in \mathcal{X}$ and $u_0 \in \mathcal{U}$ ensures recursive feasibility and constraint satisfaction for all $t = 0, \dots, n$. Due to the importance of the sets $\Lambda_t(x_t)$ in (7), it is essential to efficiently compute them before deriving an optimal strategy.

To begin, we observe that for any feasible $l_t \in \Lambda_t(x_t)$, there exists a sequence of control laws $g_{t:n-1}$ that satisfies the constraint $\mathcal{L}_t(g_{t:n-1}; x_t) \leq \hat{l}_t$ for all $l_t \leq \hat{l}_t \in \mathbb{R}$ and all $x_t \in \mathcal{X}$. Hence, it is sufficient to compute the smallest feasible bound $\lambda_t^m(x_t)$ for each $x_t \in \mathcal{X}$ and note that the set $\Lambda_t(x_t) \subseteq [\lambda_t^m(x_t), \infty)$. At the other extreme, without loss of generality, we can restrict the maxima of $\Lambda_t(x_t)$ to $l_t^M = \min\left\{l_0, \sum_{i=t}^n d_i^M\right\}$. This is because including bounds larger than l_t^M does not increase the set of feasible sequences of control laws $g_{0:t-1}$. Thus, the structural form of the set of feasible upper bounds is $\Lambda_t(x_t) = [\lambda_t^m(x_t), l_t^M]$ for all $t = 0, \dots, n-1$.

Next, we present a recursive approach to compute $\lambda_t^m(x_t)$ for all t and complete the construction of $\Lambda_t(x_t)$.

Lemma 1. *The lower bound $\lambda_t^m(x_t)$ of the set $\Lambda_t(x_t)$ for all $x_t \in \mathcal{X}$ and $t = 1, \dots, n-1$ is obtained by the following minimization problem*

$$\begin{aligned} \lambda_t^m(x_t) = \min_{u_t \in \mathcal{U}} \Big\{ & d(x_t, u_t) \\ & + \max_{P_t \in \mathcal{P}} \mathbb{E}_{P_t}\left[\lambda_{t+1}^m(X_{t+1}) \,\middle|\, x_t, u_t\right] \Big\}. \end{aligned} \quad (8)$$

Proof. The smallest feasible bound $\lambda_t^m(x_t)$ belongs to the set $\Lambda_t(x_t)$. From Definition 1, we can see that there exists a sequence $g_{t:n-1}$ for which the penalty-to-go is exactly equal to $\lambda_t^m(x_t)$ and any bound smaller than $\lambda_t^m(x_t)$ is infeasible. Thus we can compute $\lambda_t^m(x_t)$ as

$$\lambda_t^m(x_t) = \min_{g_{t:n-1} \in \prod_{\ell=t}^{n-1} \mathcal{G}_\ell} \mathcal{L}_t(g_{t:n-1}; x_t), \tag{9}$$

which becomes an instance of the standard distributionally robust DP problem. The objective is to minimize the penalty-to-go while being distributionally robust against the uncertainty in the transition function. Using the arguments presented in [20, Theorem 2.1] for deriving the optimal objective in such a problem, we can see how Lemma 1 computes the minimum value of $\mathcal{L}_t(g_{t:n-1}; x_t)$ at each t. This shows that, Lemma 1 can be used to recursively compute the smallest feasible bound $\lambda_t^m(x_t)$ for each $x_t \in \mathcal{X}$ and all t. □

B. Dynamic program for Problem 1

In this subsection, before presenting the DP decomposition, we begin by defining the cost-to-go in a manner similar to the penalty-to-go in Subsection III-A. For all $t = 0, \dots, n-1$, the cost-to-go from any $x_t \in \mathcal{X}$ is

$$\mathcal{J}_t(g_{t:n-1}; x_t) = \mathbb{E}^{g_{t:n-1}} \left[\sum_{\ell=t}^{n-1} c(X_\ell, U_\ell) + c_n(X_n) \,\middle|\, x_t \right], \tag{10}$$

where $\mathbb{E}^{g_{t:n-1}}$ denotes the expectation on all the random variables with respect to the distributions generated by the nominal transition function $\bar{P}$ and the choice of control laws $g_{t:n-1} \in \prod_{\ell=t}^{n-1} \mathcal{G}_\ell$. Note that the cost-to-go at time t is affected only by the sequence of control laws $g_{t:n-1}$ and the cost-to-go at $t = 0$ is equivalent to the performance measure (1) for any strategy $\boldsymbol{g}$.

Before we can construct a DP decomposition, we recall that Problem 1 also restricts the set of feasible strategies by constraining the penalty-to-go $\mathcal{L}_0(\boldsymbol{g}; x_0)$ with an upper bound l_0. Thus, as derived in Subsection III-A, we need to impose a constraint using the bound function $\lambda_t \in F_{t-1}(x_{t-1}, u_{t-1}, l_{t-1})$ for all $t = 1, \dots, n$ to ensure recursive feasibility in our solution approach. To this end, at each $t = 0, \dots, n-1$, we expand our state-space $\mathcal{X}$ by appending a set of possible bounds, $\mathbb{R}$. Thus, the value functions of our DP decomposition are functions of $(X_t, L_t) \in \mathcal{X} \times \mathbb{R}$, where the random variable $L_t = \lambda_t(X_t)$. The realizations of the random variable L_t are denoted by l_t. Furthermore, at each $x_t \in \mathcal{X}$, the control law $g_t \in \mathcal{G}_t$ at each $t = 0, \dots, n-1$ selects a control action $U_t \in \mathcal{U}$ using the expanded state space as $U_t = g_t(X_t, L_t)$.

Remark 2. We note that expanding the state space from X_t to (X_t, L_t) expands the domain of control laws compared to the standard Markovian control law for regular MDPs. However, the result in Subsection III-A is still valid for control laws with this larger domain because the functions introduced in III-A depend only on the realization $x_t \in \mathcal{X}$ of X_t and are independent of the realization $l_t = \lambda_t(x_t)$ of L_t.

For all $t = 0, \dots, n-1$, the value function for all $x_t \in \mathcal{X}$ and $l_t \in \Lambda(x_t)$ corresponding to the sequence of control laws $g_{t:n-1}$ is given by

$$V_t^{g_{t:n-1}}(x_t, l_t) = \begin{cases} \mathcal{J}_t(g_{t:n-1}; x_t) & \text{if } \mathcal{L}_t(g_{t:n-1}; x_t) \leq l_t, \\ \kappa & \text{otherwise,} \end{cases} \tag{11}$$

where $\kappa \in \mathbb{R}$ is a large constant that satisfies $\kappa > n \cdot c^M$ and indicates constraint violation by $g_{t:n-1}$. Eventually, when we minimize over the set of strategies, the presence κ will help us exclude infeasible solutions. At the terminal time n, where no actions are allowed, the value function is simply $V_n(x_n, l_n) = c(x_n)$. Then, the optimal value functions for all $x_t \in \mathcal{X}$, $l_t = \lambda_t(x_t)$ and all $t = 0, \dots, n-1$ are

$$V_t(x_t, l_t) = \min_{g_{t:n-1} \in \prod_{\ell=t}^{n-1} \mathcal{G}_\ell} V_t^{g_{t:n-1}}(x_t, l_t). \tag{12}$$

Theorem 1. *At each $t = 0, \dots, n-1$, for all $x_t \in \mathcal{X}$ and $l_t = \lambda_t(x_t)$, the optimal value function can be recursively computed using the following DP decomposition:*

$$V_t(x_t, l_t) = \min_{\substack{u_t \in \mathcal{U}, \\ \lambda_{t+1} \in F_t(x_t, u_t, l_t)}} \Big\{ c(x_t, u_t) + \mathbb{E}\Big[V_{t+1}(X_{t+1}, \lambda_{t+1}(X_{t+1})) \,|\, x_t, u_t \Big] \Big\}, \tag{13}$$

where, at the terminal time $t = n$, the optimal value function is simply given by $V_n(x_n, l_n) = c(x_n)$.

Proof. We prove that the DP decomposition presented in Theorem 1 computes the optimal value function recursively using mathematical induction. At the terminal time, the value function is given by $V_n(x_n, l_n) = c(x_n)$. Suppose that the optimal value function V_{t+1} at time $t+1$ can be computed according to (13). It is enough to show that (13) can be used to compute $V_t(x_t, l_t)$ at time t. We need first to show that the left-hand side in (13) is lower bounded by the right-hand side and vice-versa. As a result, the left-hand side of (13) will be both upper and lower bounded by the expression on the right-hand side. Hence, we conclude that in (13), the left-hand side is equal to the right-hand side. Details of the mathematical arguments are provided in Appendix A of [29]. □

Remark 3. We showed that the DP decomposition presented in (13) computes the optimal value function at each $t = 0, \dots, n-1$. Using Theorem 1, we can compute the sequence of optimal control laws $g^*_{0:n-1} \in \prod_{\ell=0}^{n-1} \mathcal{G}_\ell$ which yields the optimal value function $V_0(x_0, l_0)$ at time $t = 0$.

Remark 4. At any $t = 0, \dots, n-1$, and for all $x_t \in \mathcal{X}$ and a feasible bound l_t, Theorem 1 states that the optimal control action is $u_t^* = g_t^*(x_t, l_t)$, i.e., the minimizing argument in (13). Subsequently, the optimal bound function $\lambda^*_{t+1}(\cdot)$ is computed as a function of the state x_t, bound l_t, and optimal action u_t^*. Since the optimal bound function is computed at the preceding time step, during implementation, λ_t^* is available at the onset of time t and the agent ensures that $l_t = \lambda_t^*(x_t)$. Hence, to solve Problem 1, the control action at all t is selected as $u_t^* = g_t^*(x_t, \lambda_t^*(x_t))$. This shows that the optimal control strategy can be selected using $x_t \in \mathcal{X}$ during implementation.

IV. Numerical Example

In this Section, we illustrate the efficiency of the effectiveness of our approach using a numerical example. We consider a reach-avoid problem where an agent seeks to navigate to a designated cell in a 4×4 grid world while avoiding a different cell in the grid. At each $t = 0, \dots, n$, the agent's position X_t takes values in the set of grid cells $\mathcal{X} = \{(0,0), (0,1), \dots, (3,2), (3,3)\}$. The action U_t denotes the agent's direction of movement and takes values in the set $\mathcal{U} = \{(-1,0), (1,0), (0,0), (0,1), (0,-1)\}$. Under normal system operation, the agent has a small chance of movement failure by "slipping." This nominal transition function is modeled by considering that at each t, the agent moves in the direction selected by the action U_t with probability 0.8 and may slip by moving in either the clockwise or anticlockwise direction to U_t with probabilities of 0.1 each. The agent does not slip when selecting the action $(0,0)$, i.e., when deciding not to move. Thus, starting at a randomly selected initial state $x_0 \in \mathcal{X}$, the agent's dynamics for all for all $t = 0, \dots, n-1$ are

$$P(x_{t+1} \mid x_t, u_t) = \begin{cases} 0.8 & \text{if } x_{t+1} = x_t + u_t, \\ 0.1 & \text{if } x_{t+1} = x_t + u_t^{\text{cl}}, \\ 0.1 & \text{if } x_{t+1} = x_t - u_t^{\text{cl}}, \end{cases} \tag{14}$$

where if $u_t = (u_t^1, u_t^2)$, then $u_t^{\text{cl}} = (-u_t^2, u_t^1)$ is the clockwise rotation of u_t. If the agent's position X_t is at one of the four corners of the grid or along the edge of the grid, then the agent may only slip in the available directions and not move off the grid. Thus, if only one direction is available, they move in the direction of the selected action with a probability of 0.8 and move in the available direction with a probability of 0.2.

The goal of the agent is to reach the destination cell $(3,2)$ marked by D, while avoiding a "trap" cell $(2,1)$, marked by X in Fig. 1 and Fig. 2. Thus, after a time horizon of $n = 10$ time steps, the agent incurs a terminal cost given by the distance from the position X_{10} and the target $(3,2)$, i.e., $c_n(X_n) = \eta\big(X_n, (3,2)\big)$ where $\eta(\cdot,\cdot)$ denotes the Manhattan distance. The agent incurs no interim costs, i.e., $c(X_t, U_t) := 0$ for all $t = 0, \dots, n-1$. Furthermore, the agent incurs a penalty of 1 unit at any instance of time if their position coincides with the trap $(2,1)$, i.e., $d(X_t, U_t) = \mathbb{I}\big[X_t = (2,1)\big]$ for all $t = 0, \dots, n$. An adversary, if present, attacks the reliability of the agent's actuator. Thus, under an attack, the probability of slipping may increase. We incorporate vulnerability to attacks by defining, on the tuple of actual movements $(u_t, u_t^{\text{cl}}, -u_t^{\text{cl}})$ for a given action $u_t \in \mathcal{U}$, the set of possible probability distributions:

$$\begin{aligned} \mathcal{P} := \big\{&(0.7, 0.3, 0), (0.7, 0.2, 0.1), \dots, (0.7, 0, 0.3), \\ &(0.8, 0.2, 0), (0.8, 0.1, 0.1), (0.8, 0, 0.2)\big\}. \end{aligned} \tag{15}$$

We run 5000 simulations for two initial positions, $(1,0)$ and $(0,1)$ with an upper bound $l_0 = 2.5$, for which the heat map of the path selected by the agent is given in Fig. 1 and Fig. 2 respectively. In each simulation, the system is vulnerable to attack, and the transition probability is randomly picked from the set $\mathcal{P}$ as given in (15) to emulate the attack. For the purpose of demonstration, the computed optimal control strategy is implemented in a receding horizon manner for 200 time steps. For each initial condition, we compare three cases to show how our approach provides more control over the trade-off between conservativeness and optimality. In the first case, while we compute the control strategy, we consider the set of probability distributions $\mathcal{P}$ as given in (15), which yields the distributionally robust control strategy. In this case, the agent visits the "trap" cell 497 and 250 times, as shown in Fig. 1a and Fig. 2a, respectively. For the second case, during the computation of the control strategy, we expand the set $\mathcal{P}$ to include every possible distribution in $\Delta(\mathcal{X})$ to yield a conservative strategy. As a result, in Fig. 1b and Fig. 2b, the number of times the agent moves into the "trap" cell are 35 and 10, respectively. Lastly, we compute a stochastic strategy by considering that the set of probability distributions $\mathcal{P}$ is a singleton, with only the nominal transition function. Accordingly, in Fig. 1c and Fig. 2c, the agent moves 764 and 363 times into the "trap" cell, respectively. We observe that when the agent utilizes the distributionally robust control strategy, it visits the trap cell more often than the conservative strategy. However, it reaches the target cell in fewer moves than the conservative strategy. On the other hand, it reaches the destination as quickly as the stochastic strategy, with fewer visits to the "trap," essentially being more robust.

V. Conclusion

In this paper, we proposed the problem of controlling a CPS, which is vulnerable to attack as a distributionally robust stochastic cost minimization problem. For this problem, we presented DP decomposition to compute the optimal control strategy, which ensures the recursive feasibility of the distributionally robust constraint. Finally, we illustrated the utility of our solution approach using a numerical example. Future work should consider using these results in tandem with fast computation techniques for applications with large state space like human-robot collaboration tasks, power grids, and connected and automated vehicles. In such applications, it is essential to avoid over-conservatism while maintaining resilience against any vulnerabilities.

References

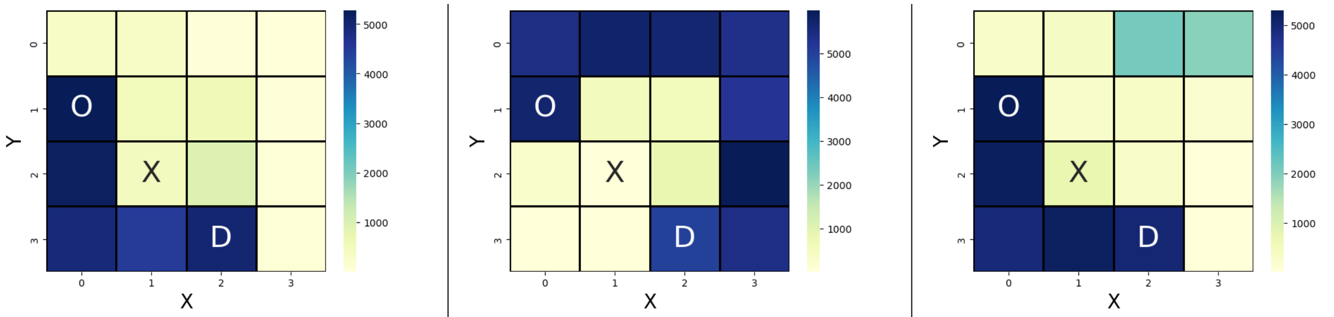
[1] K.-D. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, 2012.

[2] A. A. Malikopoulos, "On team decision problems with nonclassical information structures," *IEEE Transactions on Automatic Control*, vol. 68, no. 7, pp. 3915–3930, 2023.

[3] A. A. Malikopoulos, L. E. Beaver, and I. V. Chremos, "Optimal time trajectory and coordination for connected and automated vehicles," *Automatica*, vol. 125, no. 109469, 2021.

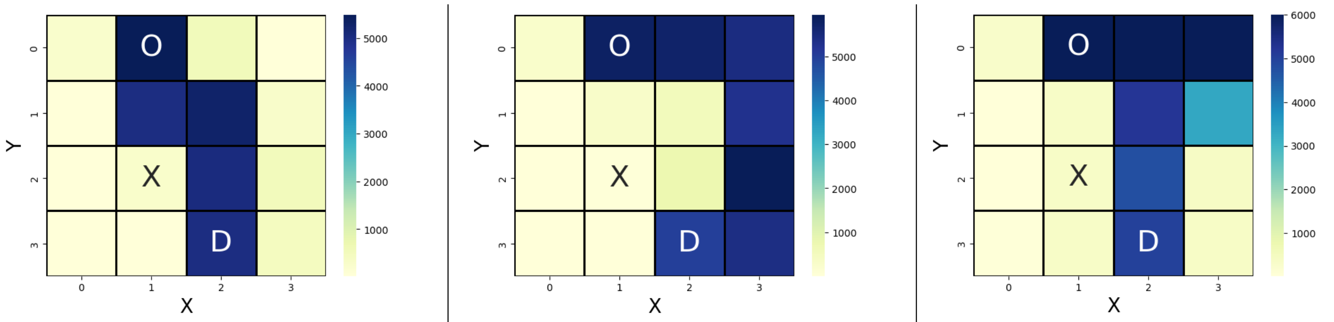
[4] N. Venkatesh, V.-A. Le, A. Dave, and A. A. Malikopoulos, "Connected and automated vehicles in mixed-traffic: Learning human driver behavior for effective on-ramp merging," in *Proceedings of the 62nd IEEE Conference on Decision and Control (CDC)*, 2023 (to appear, arXiv:2304.00397).

[5] J. A. Ansere, G. Han, L. Liu, Y. Peng, and M. Kamal, "Optimal resource allocation in energy-efficient internet-of-things networks with imperfect csi," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5401–5411, 2020.



(a) (b) (c)

Fig. 1: For the initial state $(1, 0)$, the strategy implemented in : (a) distributionally robust (b) conservative (c) stochastic



(a) (b) (c)

Fig. 2: For the initial state $(0, 1)$, the strategy implemented in : (a) distributionally robust (b) conservative (c) stochastic

[6] A. Dave, I. V. Chremos, and A. A. Malikopoulos, “Social media and misleading information in a democracy: A mechanism design approach,” *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2633–2639, 2022.

[7] M. Baezner and P. Robin, “Stuxnet,” tech. rep., ETH Zurich, 2017.

[8] M. Serror, S. Hack, M. Henze, M. Schuba, and K. Wehrle, “Challenges and opportunities in securing the industrial internet of things,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 2985–2996, 2020.

[9] M. Ghiasi, T. Niknam, Z. Wang, M. Mehrandezh, M. Dehghani, and N. Ghadimi, “A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past, present and future,” *Electric Power Systems Research*, vol. 215, p. 108975, 2023.

[10] L. Zhang, K. Sridhar, M. Liu, P. Lu, X. Chen, F. Kong, O. Sokolsky, and I. Lee, “Real-time data-predictive attack-recovery for complex cyber-physical systems,” in *2023 IEEE 29th Real-Time and Embedded Technology and Applications Symposium (RTAS)*, pp. 209–222, 2023.

[11] A. Dave, N. Venkatesh, and A. A. Malikopoulos, “Decentralized control of two agents with nested accessible information,” in *2022 American Control Conference (ACC)*, pp. 3423–3430, IEEE, 2022.

[12] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[13] A. A. Malikopoulos, “A duality framework for stochastic optimal control of complex systems,” *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2756–2765, 2016.

[14] V. Varagapriya and V. V. Singh, “Chance-constrained formulation of mdps under total reward criteria: an application to advertisement model,” in *2023 European Control Conference (ECC)*, pp. 1–6, IEEE, 2023.

[15] E. Altman, *Constrained Markov decision processes*. Routledge, 2021.

[16] S. Ermon, C. Gomes, B. Selman, and A. Vladimirsky, “Probabilistic planning with non-linear utility functions and worst-case guarantees,” in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pp. 965–972, 2012.

[17] A. A. Malikopoulos, “Separation of learning and control for cyber-physical systems,” *Automatica*, vol. 151, no. 110912, 2023.

[18] S. Mannor, D. Simester, P. Sun, and J. N. Tsitsiklis, “Bias and variance approximation in value function estimates,” *Management Science*, vol. 53, no. 2, pp. 308–322, 2007.

[19] D. Bertsekas and I. Rhodes, “Sufficiently informative functions and the minimax feedback control of uncertain dynamic systems,” *IEEE Transactions on Automatic Control*, vol. 18, no. 2, pp. 117–124, 1973.

[20] G. N. Iyengar, “Robust dynamic programming,” *Mathematics of Operations Research*, vol. 30, no. 2, pp. 257–280, 2005.

[21] M. Gagrani and A. Nayyar, “Decentralized minimax control problems with partial history sharing,” in *2017 American Control Conference (ACC)*, pp. 3373–3379, IEEE, 2017.

[22] Y. Shoukry, J. Araujo, P. Tabuada, M. Srivastava, and K. H. Johansson, “Minimax control for cyber-physical systems under network packet scheduling attacks,” in *Proceedings of the 2nd ACM international conference on High confidence networked systems*, pp. 93–100, 2013.

[23] A. Dave, N. Venkatesh, and A. A. Malikopoulos, “On decentralized minimax control with nested subsystems,” in *2022 American Control Conference (ACC)*, pp. 3437–3444, IEEE, 2022.

[24] A. Dave, N. Venkatesh, and A. A. Malikopoulos, “Approximate Information States for Worst-Case Control and Learning in Uncertain Systems,” *arXiv:2301.05089 (in review)*, 2023.

[25] M. Rasouli, E. Miehling, and D. Teneketzis, “A scalable decomposition method for the dynamic defense of cyber networks,” in *Game Theory for Security and Risk Management*, pp. 75–98, Springer, 2018.

[26] Q. Zhu and T. Başar, “Robust and resilient control design for cyber-physical systems with an application to power systems,” in *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 4066–4071, IEEE, 2011.

[27] S. P. Coraluppi and S. I. Marcus, “Mixed risk-neutral/minimax control of discrete-time, finite-state markov decision processes,” *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 528–532, 2000.

[28] R. C. Chen and G. L. Blankenship, “Dynamic programming equations for discounted constrained stochastic control,” *IEEE transactions on automatic control*, vol. 49, no. 5, pp. 699–709, 2004.

[29] N. Venkatesh, A. Dave, I. Faros, and A. A. Malikopoulos, “Stochastic control with distributionally robust constraints for cyber-physical systems vulnerable to attacks,” *arXiv preprint arXiv:2311.03666*, 2023.