1 Problem Setting

Overall, the classical system administrator problem admits the following setting:

- Each computer has three states: inoperative ($s_n = 0$), semi-operative ($s_n = 1$), and fully operative ($s_n = 2$).
- The initial distribution q is set to be a uniform distribution over $\mathcal{S}$, i.e.,

$$q(s) = \prod_{n \in [N]} q(s_n), \quad \text{where} \quad q_n(s_n) = \frac{1}{3}, \, \forall s_n \in \{0,1,2\}, \, \forall n \in [N].$$

- The reward function is given by

$$r(s,a) = \sum_{n=1}^{N} r_n(s_n), \quad \text{where} \quad r_n(s_n) = s_n, \, \forall s_n \in \{0,1,2\}, \, \forall n \in [N].$$

- For each computer n, if the system administrator does not reboot it ($a_n = 0$), then its transition kernel is given by

$$p_n(s'_n \mid s, a) = f(s'_n; s_n, s_j, s_k) \tag{1}$$

where s_j and s_k denote all the predecessors of computer n in the topology. Note that in all these network topologies shown in Figure 4, each computer has at most two predecessors. If a computer has just one (zero) predecessor, we let $s_j = s_n$ (and $s_k = s_n$). The function f is given by

$$f(s'_n; s_n, s_j, s_k) = \begin{cases} 0.95 \cdot \left(\frac{3}{2}\right)^{4s_n + s_j + s_k - 12} & \text{if } s'_n = 2, \\ 1 - 0.95 \cdot \left(\frac{3}{2}\right)^{4s_n + s_j + s_k - 12} - 0.95 \cdot \left(\frac{3}{2}\right)^{-4s_n - s_j - s_k} & \text{if } s'_n = 1, \\ 0.95 \cdot \left(\frac{3}{2}\right)^{-4s_n - s_j - s_k} & \text{if } s'_n = 0. \end{cases}$$

- The discount factor $\gamma = 0.95$.

2 Inexact Cutting Plane Scheme

The cutting plane scheme to solve problem (3) is implemented in the following inexact way.

Algorithm 1 Inexact Cutting Plane Scheme

Input: A constraints set $\mathcal{C}$, some parameters $T, L, \epsilon, \delta, \tau$;
for $t = 1, 2, \ldots, T$ **do**
    for $l = 1, 2, \ldots, L$ **do**
        Randomly generate a state-action pair $(s, a) \in \mathcal{S} \times \mathcal{A}$.
        Apply the coordinate-wise descent approach to find a more constraint-violated pair $(\tilde{s}, \tilde{a})$.
        Add $(\tilde{s}, \tilde{a})$ to the constraints set $\mathcal{C}$.
    end for
    Solve the master problem to update the weight w^t and get the objective value v^t.
    if $v^t \leqslant (1+\epsilon)v^{t-1}$ **then**
        Solve the subproblem as an MILP with early termination $ET(\tau_{\min}, \tau_{\max}, \delta \cdot v^t)$.
        if a constraint (s, a) with a sufficiently large violation is found **then**
            Add (s, a) into set $\mathcal{C}$
            Solve the master problem and update the weight w^t.
        else
            Stop the whole algorithm.
        end if
    end if
end for

The early termination condition $ET(\tau_{\min}, \tau_{\max}, \delta \cdot v^t)$ is defined as the solution time of the solver being bigger than $\tau_{\min}$ seconds and one of the following holds:

1) identify a constraint with violation larger than $\delta \cdot v^t$;

2) conclude that all the constraint violation is less than $\delta \cdot v^t$;

3) exceed the time limit of $\tau_{\max}$ seconds.

Let N denote the instance size of the system administrator problem, then these parameters we choose are summarized in the following table.

T	L	ϵ	δ	$\tau_{\min}$	$\tau_{\max}$
$20N$	$10N$	10^{-4}	10^{-3}	10	$100 + 3N$

3 Additional Results for Robust FMDPs

In this part, we present the numerical results on the 3-Legs network topology with $N = 24$ computers.

3.1 Data-Driven Experiment

The setting is the same as described in Section 5.2, except for the radius of the ambiguity sets $\epsilon \in \{0, 0.05, 0.1, 0.15, 0.2, 0.25\}$. Figure 1 compares the out-of-sample performances of the nominal and the robust policy with the performance of the clairvoyant policy that has access to the true transition kernel. The figure shows that the robust FMDP policies outperform their nominal counterparts out-of-sample by a statistically significant margin.

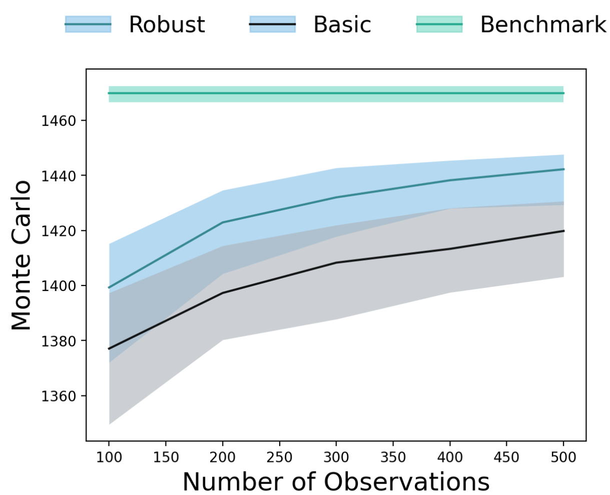


Figure 1. Out-of-sample expected total rewards for the nominal (gray, 'Basic') and robust (blue, 'Robust') FMDP policies based on historical policies of different length (abscissa). The clairvoyant policy is shown in green. Bold lines represent median performances, and shaded regions correspond to the 25%-75% quantile ranges.

3.2 Non-Data-Driven Experiment

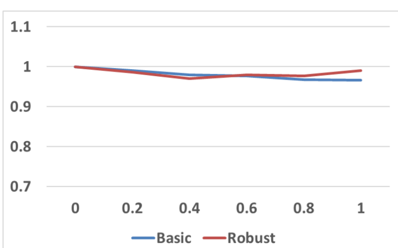
We present the results for a non-data-driven experiment where we solve both the nominal and the robust model and then evaluate both models under a true distribution $\hat{p}$ that is a mixture between the nominal transition probability p defined in (1) and a specially chosen transition probability p^{ϵ}

which satisfies $\|p - p^{\epsilon}\|_{\infty} \leqslant \epsilon$. Let $r \in [0, 1]$ denote the mixture ratio and

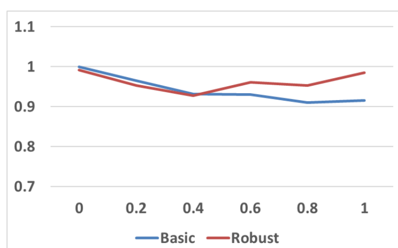
$$\hat{p} = (1 - r)p + rp^{\epsilon}.$$

However, we only have the information about the nominal distribution p. So the nominal model is a non-robust FMDP using the nominal distribution p. The robust model considers the ambiguity set whose center is the nominal distribution p and the radius is equal to ϵ.

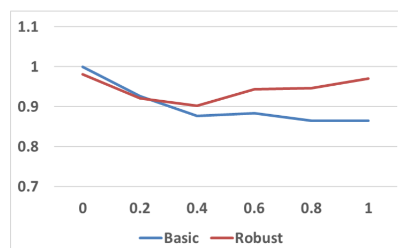
Figure 2 summarizes the results for $\epsilon = 0.1, 0.2, 0.3, 0.4, 0.5$, where the x-axis is the mixture ratio $r \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ (from nominal, left, to worst-case, right), and the y-axis is the performance of the nominal and the robust policy for the true distribution $\hat{p}$, where the performance is the Monte Carlo value divided by the Monte Carlo value of the full-informtion policy, i.e., if we knew the true distribution $\hat{p}$, we could solve the non-robust FMDP via cutting plane method and get a policy. As the figure shows, the nominal policy performs better on the left side of the graph, whereas the robust policy performs better on the right side of the graph. Besides, as the radius ϵ increases, the advantage of the robust formulation becomes more obvious.



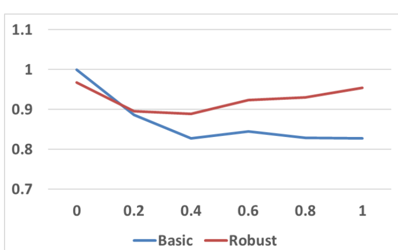
(a) $\epsilon = 0.1$



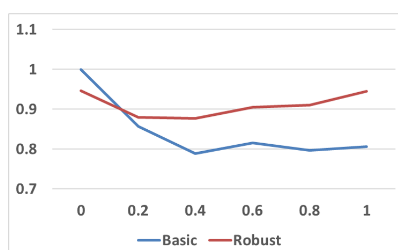
(b) $\epsilon = 0.2$



(c) $\epsilon = 0.3$



(d) $\epsilon = 0.4$



(e) $\epsilon = 0.5$

Figure 2. Comparison of Monte Carlo simulation results between the non-robust formulation and the robust formulation as the mixture ratio varies. The setting is 3 Legs topology and 24 computers. The radius of the ambiguity set is set as $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ in sub-figure (a) - (e).