Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

ADMISSION CONTROL IN THE PRESENCE OF ARRIVAL FORECASTS WITH BLOCKING-BASED POLICY OPTIMIZATION

Karthyek Murthy

Singapore University of Technology and Design
8 Somapah Rd
Singapore, SG 487372, SINGAPORE

Divya Padmanabhan
Satyanath Bhat

Indian Institute of Technology Goa
Farmagudi, Ponda
Goa, 403401, INDIA

ABSTRACT

This paper presents a simulation based policy optimization scheme for performing queuing admission control in the presence of noisy arrival forecasts. The forecast models considered for arrival times go beyond the no noise and no show forecast models treated in the literature and incorporate realistic features such as decreasing accuracy profile for jobs arriving farther in future. Assuming access to forecasts for arrivals within a look-ahead window, the paper proposes optimization over a policy class which approximates combinations of threshold and blocking type policies in the literature. While threshold policies tend to be optimal for admission control problems without forecasts, blocking-based policies have been effective in settings where exact arrival data is known. Exact knowledge on future arrivals is however unrealistic and a key novelty is the use of robust optimization to compute blocking policy statistics. Numerical experiments demonstrate good reductions in waiting costs achievable by incorporating forecast data.

1 INTRODUCTION

Forecasting models are employed in several operations management applications to improve decisions made in the presence of uncertainty. The range of applications which employ forecasts for this purpose has significantly broadened in the recent years due to the increasing accessibility to data and algorithmic advances in prediction modeling. Improving service quality to reduce wait times is a prominent avenue of exploration in this respect, as this arises naturally in applications such as traffic flow management, regulating overcrowding in hospital emergency departments, etc. Processing capacity cannot be altered quickly in these settings as a response to being overwhelmed and tactical redirection of arrivals may be necessary. Studying admission control in queues offer a good abstraction of these decision-making situations and our interest in this paper is to examine how noisy forecasts can be incorporated to decide dynamically on accepting or redirecting job arrivals.

Queuing admission control problems enjoy a rich literature: Under conventional assumptions on jobs arriving according to a renewal process and independent job sizes, the sequential decision-making task at hand is approached typically by formulating as a Markov or semi-Markov decision process instance; see, for example, (Naor 1969; Johansen and Stidham 1980) and surveys (Stidham and Prabhu 1974; Stidham 1985) and references therein. The costs involved are factored in by assuming a fixed one-time cost per job rejection/redirection and a holding cost per unit time a job is held in the system. In instances where the system manager's knowledge is restricted to the current queue length and statistics pertaining to the interarrival and job size distributions, the optimal policies tend to have a threshold structure: the system manager admits the next job when the present queue length (or workload) is below a critical level, redirects the job when the present workload is above the critical level, and admit/reject with a suitable probability when the present workload matches the critical level; see, for example, (Beutler and Ross 1986).

If the system manager additionally has access to forecasts for realizations of future arrival times and job sizes, then it is natural to expect better admission decisions. However the extent of the reduction on waiting and rejection costs have not been known until the breakthrough results in (Spencer et al. 2014). Assuming knowledge of exact arrival times and sizes of jobs arriving within a finite future window, Spencer et al. (2014) presents an intuitive policy which utilizes the future arrival information to anticipate workload buildups and make job redirections accordingly. With sufficiently long look-ahead window of exact information on future arrivals, Spencer et al. (2014) demonstrate that the expected queue length and workload in the queue can be controlled to remain bounded even in heavy-traffic conditions where admitted workload rate is close to the processing rate of jobs by the server.

Such substantial reduction is achieved by deploying a so called *blocking policy* which, roughly speaking, peeks into the baseline workload offered by all jobs arriving within the look-ahead window and rejects/redirects jobs arriving at inopportune times. These inopportune times are characterized by arrival time points at which the baseline workload level climbs up and does not revert back within the look-ahead window. Rejecting these blocking arrivals involves only polynomial time computation and does not require an exhaustive search over policy space as would be required in an Markov decision process formulation.

Deploying a blocking-type policy of Spencer et al. (2014) is not straightforward in practice, however, due to its requirement of the knowledge of exact future arrival information. Forecasts on future arrivals, if available, are often noisy (see (Sun et al. 2009)) and the accuracy of forecasts typically improve over time. In other words, forecasts of arrivals nearer in the future are typically more accurate than forecasts of arrivals farther in the future. Considering the problem of crowding in emergency departments, Xu and Chan (2016) show that a variant of blocking policies can cope with a certain amount of "no-show" noise. They show that combining the blocking policy in (Spencer et al. 2014) with a thresholding on current workload can lead to a superior performance in all traffic conditions. As mentioned earlier, the advantage with these policies is that they do not require an exhaustive search over a policy space and yet offer good performance in M/M/1 setting. However, a straightforward application of blocking type policies to different noise models in forecasts do not offer the same impressive performance and adapting them suitably to general forecasts and complex queuing environments appears non-trivial.

Motivated by these studies, we develop an admission control formulation featuring arrival forecasts which incorporate the following realistic features: Forecasts become available for future arrivals within a finite look-ahead window. Further, the forecasts are taken to be noisy versions of the unknown actual arrival times which get better with more updates arriving over time. This leads to the realistic proposition of nearer arrivals being more accurate than arrivals farther in the future, as a consequence of having accumulated more updates over time. We formulate a sequential decision problem for admission control under general forecasts in terms of the usually considered combination of waiting and rejection costs.

To tackle the challenges introduced by the large state-space comprising all workload and forecast information, we proceed with a simulation based policy optimization scheme to search for an effective policy within a parameterized softmax policy class. The proposed policy class is rich enough to approximate the universe of policies which possess threshold or blocking structures described above, or a suitable combination of them. Since the proposed policy class includes combinations of policies with threshold and blocking structures, the proactive admission policies in (Xu and Chan 2016) also get approximated well by the proposed parameterized softmax policy class. Searching for an effective policy for control in queuing and stochastic processing networks by means of policy gradients has been considered previously in (Bhatnagar and Lakshmanan 2012; Dai and Gluzman 2022).

A key novelty in our paper is the use of a robust optimization formulation for arriving at statistics required for incorporating a blocking-type policy in the absence of exact information on future arrivals. This enables fast computation of blocking-based feature vector components in place of expensive simulation of numerous sample paths. Using robust optimization to arrive at blocking features with a worst-case perspective is appealing conceptually as well, as it allows to factor-in the effects of rouge realizations that can happen within the look-ahead window. Due to reflection at origin, queuing workloads do not lend themselves immediately to inexpensive computation. However, in our case, approximating a blocking policy requires only minimum workload over a look-ahead window and this computation can be performed with a computationally inexpensive recursion.

The rest of the paper is organized as follows: Upon introducing the queueing and arrival forecasts models in Section 2, we formulate the admission control task as a sequential decision problem in Section 3. The parameterized policy class considered and the simulation-based policy optimization procedure are explained in Section 4. Numerical experiments which demonstrate the benefits of using a combination of blocking and threshold policies via the proposed softmax policy class are given in Section 5

2 A QUEUEING MODEL WITH ARRIVAL FORECASTS

We consider a queuing workload process $(W_n : n \geq 0)$ whose evolution is modeled by,

$$W_n = (W_{n-1} + su_n - 1)^+ , \quad n \geq 1, \tag{1}$$

where s is the deterministic workload brought by an arriving job and u_n is the number of jobs admitted at time n. While the admitted arrivals enter the waiting room at the time of admission, the rejected / redirected arrivals leave the system immediately. The quantity W_n captures the total unfinished work in the waiting room and the server. The system features a non-idling single-server processing one unit of workload per unit time.

Arrivals are indexed by labels $a \in \mathcal{A}$. Each arrival a has a scheduled arrival time s_a and its actual time of arrival T_a may deviate from the scheduled arrival time due to uncertainty. Let σ_a^2 denote the variance in the actual job arrival time T_a given the initial schedules $\mathcal{S} = (s_a : a \in \mathcal{A})$.

At every time-step $n \geq 1$, the decision maker facing arrival uncertainty is equipped with a forecast list which includes the latest forecasts f_n of arrival times for all jobs originally scheduled to arrive before time $n + \omega$. For an arrival $a \in \mathcal{A}$ scheduled to arrive at or after time $n + \omega$, the forecast arrival time f_{na} merely equals the scheduled arrival time s_a. The forecast time f_{na} and the actual arrival time T_a coincide if and only if the arrival a happens not later than time n. Thus the forecast vector $f_n = (f_{na} : a \in \mathcal{A})$ comprises the best prediction we have about arrival times $(T_a : a \in \mathcal{A})$ before decision-making at time n and the window ω is a positive integer determining the scheduled arrivals for whom the decision-maker is equipped to receive forecasts.

Instead of assuming perfect knowledge of the future arrival times, the forecast list f_n is taken to comprise only unbiased noisy predictions which get better over time. Specifically, if we let $E_n[\cdot]$ and $\text{Var}_n[\cdot]$ denote the conditional expectation and variance operators with respect to the filtration available while decision-making at time n, we assume the forecasts $(f_n : n \geq 0)$ satisfy Assumption 1 below.

Assumption 1 For any $a \in \mathcal{A}$ and $n \geq 0$, the forecasts $(f_{na} : n \geq 0)$ satisfy

$$E_n[T_a] = f_{na}, \quad \text{Var}_n[T_a] = \begin{cases} \sigma_a^2 & \text{if } n + \omega \leq s_a, \\ 0 & \text{if } n \geq f_{na}, \end{cases}, \quad \text{and} \quad E[\text{Var}_n[T_a]] = v_a(n), \tag{2}$$

where the function $n \mapsto v_a(n)$ is decreasing in n.

Observe that $v_a(n) = \sigma_a^2$ for any $a \in \mathcal{A}$ with schedule s_a situated beyond the forecast window (that is, if $s_a \geq n + \omega$). Indeed, this follows from the above system description that non-zero forecast updates happen at time n only for arrivals which have not yet arrived and are scheduled to arrive in the interval $(n, n + \omega)$. Moreover, since the forecasts are unbiased, their mean-square error is given by

$$\text{mse}_a(n) := E\left[(T_a - f_{na})^2\right] = v_a(n), \qquad \text{for } a \in \mathcal{A}.$$

Example 1 (Random walk forecast model) Suppose the forecasts evolve as in $f_n = f_{n-1} + Z_{n-1}$, where $\{Z_n : n \geq 0\}$ are zero mean independent $\mathbb{R}^{\mathcal{A}}$-valued random vectors modeling the evolution of forecasts.

In particular, one may view Z_{n-1} as the forecast update term capturing information revealed during interval $[n-1, n)$ about the unknown arrival times in $(T_a : a \in \mathcal{A})$. As the forecast $f_n = s_a$ for any time instant n preceding the forecast window for flight a beginning at $s_a - \omega$, we have

$$f_{na} = s_a + \sum_{i=\lfloor s_a-\omega \rfloor+1}^{n\wedge\tau_a} Z_{i-1,a},$$

where $\tau_a = \inf\{n \geq 0 : f_{na} \leq n\}$ and $T_a = f_{\tau_a,a}$ as a consequence. In case where the schedules s_a are integers and updates Z_i are supported on $\{-1, 0, 1\}$, we have from the definition of the hitting time τ_a that $T_a = f_{\tau_a} = \tau_a$. In particular, if every non-zero update is assumed to possess equal variance, then we have

$$\text{Var}_n[Z_{na}] = \begin{cases} \frac{\sigma_a^2}{\omega} & \text{if } n \in (s_a-\omega, f_{na}) \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad \text{Var}_n(T_a) = \begin{cases} \frac{\sigma_a^2}{\omega}(f_{na}-n)^+ & \text{if } n > s_a-\omega \\ \sigma_a^2 & \text{otherwise.} \end{cases} \tag{3}$$

Indeed, these observations follow from an application of Wald's identity to the random walk f_{na} to obtain $E[\tau_a] = s_a$, $\sigma_a^2 = E[(T_a - s_a)^2] = E[\sum_{i=s_a-\omega}^{\tau_a-1} Z_{ia}]$ and $\text{Var}_n[\tau] = E_n[\tau_a - n]\mathbf{1}_{\{f_{na}>n\}}$. Moreover, at any decision-making instance n, the total variance in any arrival time $\text{Var}[T_a] = \sigma_a^2$ gets split as,

$$\begin{aligned} \sigma_a^2 = \text{Var}[T_a] &= E\left[\text{Var}_n\left[T_a\right]\right] + \text{Var}\left[E_n\left[T_a\right]\right] \\ &= v_n(a) + \text{Var}[f_{na}]. \end{aligned}$$

Since $\text{Var}[f_{na}] = \text{Var}[\sum_{i=s_a-\omega}^{n\wedge\tau_a-1} Z_{ia}] = (E[n\wedge T_a] - s_a + \omega)\sigma_a^2/\omega$, we obtain

$$v_n(a) = \frac{\sigma_a^2}{\omega}\left(s_a - E[n\wedge T_a]\right),$$

which is decreasing in n, thus satisfying Assumption 1. □

In general, a forecasts dynamics of the form

$$f_n = g(f_{n-1}, Z_{n-1}) \tag{4}$$

captures a generic forecast update equation with Z_{n-1} informing the random update made available during time $[n-1, n)$. Besides the special case of the random walk model considered in Example 1, it is natural for a number of forecast models to satisfy the general conditions stipulated in Assumption 1. Further, to keep the focus on the impact of future arrival forecasts on admission control, we consider deterministic service times. Since the key driver of the Lindley recursions defining workload or storage processes, in general, involves only the difference between interarrival and service times, the approaches considered in the paper pertaining to forecasts of interarrival times are transferable to settings with random service times.

3 A SEQUENTIAL DECISION-MAKING FORMULATION

Together with the initial schedules $\mathcal{S}$, the tuple $X_n := (W_n, f_n)$ describes the state of the system at time n. Let

$$\mathcal{H}_n = \{\mathcal{S}, (W_k : k \leq n-1), (f_k : k \leq n), (u_k : k \leq n-1)\}$$

denote the history available while performing admission control at time n. We take $\mathcal{H}_0 = \{\mathcal{S}, W_0, f_0\}$. As rejected customers leave the system, the jobs which are eligible for admission at a given time $n \geq 1$ is given by the following set of arrivals,

$$A_n := \left\{a \in \mathcal{A} : f_{na} < n,\ f_{(n-1)a} \geq n-1\right\}, \tag{5}$$

which have occurred since the previous decision epoch $n-1$. Consequently,

$$|A_n| = \sum_{a\in\mathcal{A}} \left[\mathbf{1}_{[0,n)}(f_{na}) - \mathbf{1}_{[0,n-1)}(f_{(n-1)a})\right] = \sum_{a\in\mathcal{A}} \mathbf{1}_{\{f_{na}\leq n,\ f_{(n-1)a}>n-1\}},$$

is the number of arrivals which arrive during the interval $[n-1,n)$. Thus given the history $\mathcal{H}_n$ at time n, the set of permissible actions is the collection, $\{0,1\}^{A_n}$, of all functions from A_n to $\{0,1\}$. An action $u_n \in \{0,1\}^{A_n}$ is interpreted as

"admit arrival $a \in A_n$ if and only if $u_n(a) = 1$."

Admission control formulations often proceed by explicitly modeling the waiting costs and rejection costs to capture the tradeoff between the two. Keeping an arrival in the system incurs a cost of $c_s = 1$ unit per unit time, whereas rejecting an arriving customer incurs a one-time cost of $c_r \in (0,\infty)$. The assumption $c_s = 1$ is without loss of generality, as c_r captures the relative cost of rejection.

A policy π comprises a sequence of maps $\pi = (\pi_n : n \geq 1)$ in which the component π_n maps any given history realization $\mathcal{H}_n$ to a probability distribution on $\{0,1\}^{|A_n|}$. Equipped with the history $\mathcal{H}_n$, a decision-maker employing a policy π at time n selects an action by sampling from the probability distribution $\pi_n(\mathcal{H}_n)$. Let $E_\pi[\cdot]$ denote taking expectation with the probability distribution induced by selecting actions $u_n \sim \pi_n(\mathcal{H}_n)$ for any $n \geq 1$. A goal of the decision-maker is to select a policy π from the policy class Π which minimizes the expected discounted cost,

$$E_\pi\left[\sum_{n=0}^{N} \gamma^n (W_n + c_r(|A_n| - 1^T u_n)\right], \tag{6}$$

where N is the horizon of the control problem (which could possibly be infinite), $\gamma \in (0,1)$ is a discount factor, and u_n denotes the action selected from the probability distribution $\pi_n(\mathcal{H}_n)$. Since $u_n \in \{0,1\}^{A_n}$, the component $1^T u_n$ is the number of customers admitted and $|A_n| - 1^T u_n$ is the number of customers rejected. Thus the above expected cost formulation includes a combination of admitted workload cost borne by the system as a consequence of admissions and the rejection costs incurred due to rejections.

4 DESCRIPTION OF THE POLICY CLASS

4.1 Softmax Parameterization of Policies

The state-space comprising the set of all possible workload and forecast combinations is high-dimensional and exact solution schemes will suffer from the curse of dimensionality that is common for MDPs with large state-spaces. To overcome this curse of dimensionality, we consider a softmax policy class described as follows. At time $n \geq 1$, the decision-maker uses the history $\mathcal{H}_n$ to employ a softmax policy to independently admit each arrival $a \in A_n$ with probability,

$$\Pr(\text{admit } a \mid \mathcal{H}_n) = \frac{1}{1+\exp(-\theta^T \xi_n)},$$

where ξ_n is a suitable $\mathbb{R}^d$-valued feature vector constructed from the history $\mathcal{H}_n$ and θ is a suitable parameter vector from a set $\Theta \subseteq \mathbb{R}^d$. An advantage of parameterizing policies according to the softmax over available actions is that the approximate policy can approach a deterministic policy. Randomized policies are also helpful in promoting exploration across states which appear identical under a feature mapping or function approximation (see Sutton and Barto 2018). Since the admission decisions are independent and identically distributed for each $a \in A_n$ given $\mathcal{H}_n$, we have

$$1^T u_n \sim \text{Binomial}\left(|A_n|,\ p_\theta(\xi_n)\right), \quad \text{where} \quad p_\theta(\xi) := \frac{1}{1+\exp(-\theta^T \xi)},$$

and the notation $X \sim \text{Binomial}(n, p)$ denotes X distributed as a binomial variable with n trials each succeeding with probability p.

Our approach towards approximating policy class, instead of the more common approach of approximating value functions, is guided by the aim to leverage the insights from the rich literature on queueing control. We discuss briefly some of relevant optimal policy structures in admission control problems in the following section.

4.2 On Effective Policy Structures for Admission Control with and without Forecasts

Threshold policies serve as prominent examples due their optimality in an array of queueing control settings. Indeed, in the absence of forecasts, a natural variant of the admission control problem we consider admits an optimal threshold policy (Stidham 1985). Threshold policies are easy to implement and explainable as they admit an arrival based on whether the current workload is below a threshold. However, in the presence of exact future information, the performance of threshold policies which only utilize the current workload information has been shown to be significantly bettered by blocking policies proposed in Spencer et al. 2014.

To understand blocking policies, suppose that at time n we have the exact knowledge of arrival times and service times in an M/M/1 queue within the future window $[n, n+\omega)$. Let $\tilde{Q}$ denote the queue length process which occurs in the event all arrivals are unconditionally admitted. Equipped with this baseline queue length process $\tilde{Q}$, the blocking policy introduced in Spencer et al. 2014 proceeds by rejecting an arrival occurring at time t if

$$\min_{0 \leq t \leq \omega} \tilde{Q}(n+t) \geq \tilde{Q}(t^+),$$

where $\tilde{Q}(t^+)$ is the right-limit of $\tilde{Q}$ at t. Such arrivals are labeled *ω-blocking arrivals*, as the baseline queue length process $\tilde{Q}$ does not return to $\tilde{Q}$ at time t within the next ω units of time. As these arrivals happen at the beginning of an inopportune period where more arrivals are likely to happen than can be handled, they tend to delay all subsequent arrivals upon admission. Interestingly, a careful incorporation of information on future arrival and service times has been shown in Spencer et al. 2014 to result only in expected queue length of magnitude $O_p(1)$ even in heavy-traffic situations where the admitted arrival rate is close to the service rate. Please refer Xu and Chan 2016 for further explanation on the motivation and effectiveness of the ω-blocking policies under perfect knowledge of arrivals.

If equipped with information required for threshold and blocking policies, observe that the softmax policy parameterization introduced in Section 4.1 is rich enough to approximate these optimal policy structures by appropriate selection of feature vectors ξ_n. We discuss this further in the following section on selection of features in ξ_n given $\mathcal{H}_n$.

4.3 Using Robust Optimization to Inform Blocking Arrivals

While effective, the ω-blocking policies require the exact knowledge of future arrival times within time $n+\omega$ at time n itself. However in practice, it is common that available forecasts are only noisy predictions of true arrival times. Moreover, as in Section 2, it would be natural to have that the nearer arrival times are known with higher accuracy than later arrivals. Therefore the blocking policies are not directly applicable in the presence of predictions with considerable noise.

To work around this difficulty, suppose that $(\tilde{W}_k^n : k \geq n)$ denotes the baseline workload process resulting given history $\mathcal{H}_n$ and all arrivals are admitted from decision epoch n onward. Since the history $\mathcal{H}_n$ at time n is fixed and all arrivals are admitted from time n onward, we take $\tilde{W}_{n-1}^n = W_{n-1}$ as given in the history $\mathcal{H}_n$. Then a natural alternative to arrive at a policy inspired by blocking policy in our setting could be to identify ω-blocking arrivals as arrivals in time instances n for which

$$\rho\left(\min_{0\,\leq\,k\leq\omega} \tilde{W}^n_{n+k}\right) \geq \tilde{W}^n_{n-1}, \tag{7}$$

where $\rho(\cdot)$ is a risk measure such as mean, median, or value at risk of the conditional distribution of $\min_{0\,\leq\,k\leq\omega} \tilde{W}^n_{n+k}$ given forecasts in $\mathcal{H}_n$. Since this check involving baseline workload process $\tilde{W}^n_{n+k}$ computed at time n (hence indexed by superscript n) would be needed only at time n, we drop the superscript n for ease of notation in all instances where there is no scope for confusion.

For a chosen risk measure $\rho(\cdot)$, the quantity $\rho(\min_{0\,\leq\,k\leq\omega} \tilde{W}_{n+k})$ is rarely available exactly and Monte Carlo simulation is a natural way to facilitate its computation. An estimation of this type involving numerous sample paths at every decision-making epoch is however a computationally intensive endeavour to be expended on a feature vector construction sub-task. To lighten the burden on computation while retaining the information contained in forecasts, we consider the following worst-case realization of $\min_{0\,\leq\,k\leq\omega} \tilde{W}_{n+k}$.

For $n \geq 0$, $a \in \mathcal{A}$, and $\Gamma \geq 0$, let

$$\mathcal{U}^a_n(\Gamma) = \left\{ t : |t - f_{na}| \leq \Gamma \text{Var}^{1/2}_n[T_a] \right\}$$

be an interval uncertainty set featuring potential arrival times of arrival a given the history $\mathcal{H}_n$. The uncertainty set is centered at the conditional expectation of T_a and its radius is determined by the conditional variance $\text{Var}_n[T_a] = \text{Var}[T_a \mid \mathcal{H}_n]$. This leads to the following box uncertainty set for the arrival times $T = (T_a : a \in \mathcal{A})$:

$$\mathcal{U}_n(\Gamma) = \prod_{a\in\mathcal{A}} \mathcal{U}^a_n(\Gamma).$$

Here $\Gamma > 0$ can be treated as an additional parameter which can be used to tune the extent of uncertainty in the conditional distributions of arrival times given forecasts. The parameter Γ can be chosen sufficiently large such that the interval $\mathcal{U}^a_n(\Gamma)$ contains T_a with high probability. As long as the variances in forecasts remain finite, one can use Chebyshev's inequality, for example, to help accomplish this.

Inspired by the computational tractability which comes often with robust optimization formulations, we consider the following worst-case minimum value of the baseline workload in the next ω period: Given $\mathcal{H}_n$ and any $k \geq 0$, the quantity

$$\sup_{T\,\in\,\mathcal{U}_n(\Gamma)} \min_{0\leq j\leq k} \tilde{W}_{n+j} \tag{8}$$

denotes the worst-case minimum workload that can happen during the interval $[n, n+k]$ if the realizations of arrival times $T = (T_a : a \in \mathcal{A})$ lie within the uncertainty set $\mathcal{U}_n(\Gamma)$. In this baseline workload computation, recall that the workload process gets determined by arrival times as described in Section 2. As all arrivals from decision epoch n are set to admitted in the computation of baseline workload process, we have

$$\tilde{W}_{n-1} = W_{n-1}, \qquad \tilde{W}_{n+j} = \left(\tilde{W}_{n+j-1} + s\sum_{a\in\mathcal{A}} \mathbb{I}(T_a \in [n+j-1, n-j) - 1\right)^+, \tag{9}$$

defining the sequence $\{\tilde{W}_{n+j} : j \geq 0\}$ over which the worst-case computation over all realizations of $T = (T_a : a \in \mathcal{A})$ is done in (8). Facilitating the comparison (7) via (8) is driven by computational considerations, as we shall see in the following section that the worst-case minimum workload admits tractable computation given $\mathcal{H}_n$. It also serves as a robust optimization based analogue for the expected minimum baseline workload within the look-ahead window.

4.4 Tractable Computation of Worst-case Minimum Workload in the Baseline Process

In order to demonstrate tractable computation of the worst-case minimum workload in (8), let us suppose that the conditional variances in arrival times T_a satisfy Assumption 2 below.

Assumption 2 For every $a \in \mathcal{A}$, the conditional variance $\text{Var}_n[T_a] = \text{Var}[T_a \mid \mathcal{H}_n] = \nu_a^2(f_a - n)$, for some increasing function $\nu_a : \mathbb{R} \to \mathbb{R}_+$.

Assumption 2 stipulates that the conditional variance $\text{Var}_n[T_a]$ is determined only by how far the arrival a has been presently forecast to arrive in the future. Farther the arrival time, larger is the remaining conditional variance. In the context of the random walk forecast model in Example 1, we have from (3) that Assumption 2 is satisfied with

$$\nu_a(x) = \sigma_a \sqrt{\frac{x^+}{\omega}}.$$

Let us introduce the functions $l_a : \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$ and $u_a : \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$, for $a \in \mathcal{A}$, as follows:

$$l_a(x, y, \Gamma) = x - \Gamma \left[\nu_a(x)\mathbf{1}_{\{y < \omega\}} + \sigma_a \mathbf{1}_{\{y \geq \omega\}}\right] \text{ and}$$
$$u_a(x, y, \Gamma) = x + \Gamma \left[\nu_a(x)\mathbf{1}_{\{y < \omega\}} + \sigma_a \mathbf{1}_{\{y \geq \omega\}}\right].$$

Proposition 1 Suppose the forecast model satisfies Assumptions 1 and 2. Then given $W_{n-1} = w$, $f_{n-1} = f'$, $f_n = f$, scheduled arrival times $\mathcal{S} = (s_a : a \in \mathcal{A})$, $\Gamma \geq 0$, and integers $n, k \geq 0$, we have

$$\sup_{T \in \mathcal{U}_n(\Gamma)} \min_{0 \leq j \leq k} \tilde{W}_{n+j} = m_k(w,\, f - n,\, f' - (n-1),\, \mathcal{S} - n,\, \Gamma),$$

where the functions $\{m_k : k \geq 0\}$ are defined as follows: for any $x, y, z \in \mathbb{R}^{\mathcal{A}}$,

$$m_0(w, x, y, z, \Gamma) = (w + s|\mathcal{N}_0| - 1)^+ \quad \text{and} \tag{10}$$
$$m_k(w, x, y, z, \Gamma) = \min\left\{m_{k-1}(w, x, y, z, \Gamma),\ (w + s|\mathcal{N}_k| - (k+1))^+\right\}, \tag{11}$$

for $k \geq 1$, $\mathcal{N}_0 = \sum_{a \in \mathcal{A}} \mathbf{1}_{\{x_a < 0,\, y_a \geq 0\}}$ and

$$\mathcal{N}_k := \mathcal{N}_{k-1} \cup \{a \in \mathcal{A} \setminus \mathcal{N}_{k-1} : (k-1) \vee l_a(x_a, y_a, \Gamma)\} \ \leq \ k \wedge u_a(x_a, y_a, \Gamma)\}. \tag{12}$$

Proposition 1 provides a characterization in which the worst-case realizations of the minimum baseline workload in (8) can be computed recursively via the sets $\mathcal{N}_0, \ldots, \mathcal{N}_k$. For any $j \leq k$, the set $\mathcal{N}_j$ is the collection of all arrivals who could have potentially arrived during the period $[n-1, n+j)$ and can be handled with the computationally inexpensive recursion in (12).

Proof of Proposition 1. The expression for $m_0(\cdot)$ in (10) follows from (9) and by observing that, given history $\mathcal{H}_n$, the set $\mathcal{N}_0$ equals the collection of arrivals A_n in (5). If we let $m_{-1}(\cdot) = -\infty$, then $m_0(\cdot)$ can be seen as satisfying (11) with $k = 0$.

To verify (11) with $k = i$ for some $i \geq 1$, suppose that (11) is satisfied for $k = i - 1$. We consider two cases based on the value of $m_{k-1}(w, x, y, z, \Gamma)$. If $m_{k-1}(w, x, y, z, \Gamma) = 0$, then since $\min_{j \leq i} \tilde{W}_{n+j}$ is non-increasing in i, we have (11) satisfied since $\min_{j \leq i} \tilde{W}_{n+j} = 0 = \min\{0, (w + s|\mathcal{N}_i| - i - 1)^+\}$.

If $m_{i-1}(w, x, y, z, \Gamma) > 0$, there exists a realization for $T = (T_a : a \in \mathcal{A})$ in the uncertainty set $\mathcal{U}_n(\Gamma)$ such that $\tilde{W}_n, \tilde{W}_{n+1}, \ldots, \tilde{W}_{n+i-1}$ are all (strictly) positive. In that case, we have from (9) and the definition of $\mathcal{N}_j$ that

$$\begin{aligned} \tilde{W}_{n+j} &= \tilde{W}_{n-1} + \sup_{T \in \mathcal{U}_n(\Gamma)} \sum_{a \in \mathcal{A}} s\mathbb{I}(T_a \in [n-1, n-j)) - (j+1) \\ &= w + s|\mathcal{N}_j| - (j+1), \end{aligned} \tag{13}$$

for every $j = 0, \ldots, i-1$ and some realization in $\mathcal{U}_n(\Gamma)$. The latter equation follows by observing that $T_a \in \mathcal{U}_n(\Gamma)$ satisfies $T_a \in [n-1, n+j-1)$ if and only if $a \in \mathcal{N}_j$. Indeed this can be seen by noting the argument $x_a = f_{na} - n$ in $\nu(x_a)$, which makes $\nu_a(x_a) = \text{Var}_n[T_a]$ as per Assumption 2.

Moreover, since the minimum workload $\min_{j \leq i-1} \tilde{W}_{n+j}$ is zero for all realizations for which (13) is not true, one can restrict the supremum in (8) to the set of realizations in $\mathcal{U}_n(\Gamma)$ for which (13) holds for every $j = 0, \ldots, i-1$. Then it follows from (9) and the definition of the collection $\mathcal{N}_i$ that

$$\begin{aligned}\tilde{W}_{n+i} &= \big(\tilde{W}_{n+i-1} + s \sum_{a \in \mathcal{A}} \mathbb{I}(T_a \in [n+i-1, n-i) - 1\big)^+ \\ &= \big(\tilde{W}_{n+i-1} + s|\mathcal{N}_i \setminus \mathcal{N}_{i-1}| - 1\big)^+ \\ &= \big(w + s|\mathcal{N}_{i-1}| - i + s|\mathcal{N}_i \setminus \mathcal{N}_{i-1}| - 1\big)^+,\end{aligned}$$

where the last equation follows from (13) with $j = i-1$. Observing that $|\mathcal{N}_{i-1}| + |\mathcal{N}_i \setminus \mathcal{N}_{i-1}| = |\mathcal{N}_i|$ completes the verification for the chosen k. Thus, by the principle of mathematical induction, the conclusions in proposition 1 stand verified.

□

4.5 Feature Map Incorporating Worst-case Blocking-based Features

Motivated by existing effective policy structures described in Section 4.2, we take the feature vector $\xi_n = (\xi_{n1}, \ldots, \xi_{n5})$ mapping the history $\mathcal{H}_n$ to $\mathbb{R}^5$ as follows:

$\xi_{n1} = W_{n-1}$ is the most recent workload information,

$\xi_{n2} = m_\omega\left(W_{n-1}, f_n - n, f_{n-1} - (n-1), \mathcal{S}_a - n, 0\right)$ is the minimum baseline workload over the period $[n, n+\omega]$ if the forecasts are taken to be exact,

$\xi_{n3} = m_\omega\left(W_{n-1}, f_n - n, f_{n-1} - (n-1), \mathcal{S}_a - n, \Gamma\right)$ is the worst-case realization of the minimum baseline workload over the period $[n, n+\omega]$ if $T \in \mathcal{U}_n(\Gamma)$,

$\xi_{n4} = \sum_{a \in \mathcal{A}} \mathbf{1}_{\{f_n - n < 0,\ f_{n-1} - (n-1) \geq 0\}}$ is the number of arrivals in the current decision epoch, and

$\xi_{n5} = 1$ is the component for intercept/offset term.

Together with the softmax policy parameterization considered in Section 4.1, these feature elements have sufficient information to approximate threshold policy families, blocking policy variants, and as well a combination of them. A consequence of Assumption 2 is that the feature vector depends on the forecasts and schedules of future arrivals only via how far away they are in the future.

4.6 Simulation based Policy Optimization via Policy Gradients

Equipped with the parameterized policy class and the precise description of the feature vectors, Algorithm 1 provides a description of a basic policy gradient method one can employ to arrive at an effective, but potentially local optimal policy parameter while minimizing (6).

It is often useful to subtract a baseline function from the costs $C_n^{(i)}$ before plugging it in (14) in order to reduce the variance in the gradient estimate $\hat{g}$. The baseline function can, in turn, be steered to approximate the advantage difference in value function and be parameterized in terms of the feature vector ξ and cross-combinations of its elements $\{\xi_i \xi_j : i, j = 1, \ldots, 4\}$. Then the cost observations $C_n^{(i)}$ serve in updating the baseline function parameterization in each outer iteration, for example, by means of minimizing least-squares (see, for example, Chapter 13 in Sutton and Barto 2018).

Algorithm 1: A basic policy gradient scheme for searching for an optimal policy parameter θ

Input: Initial policy parameter $\theta \leftarrow \theta_0$, step-size α, relative rejection cost c_r, discount factor $\gamma \in (0,1)$, uncertainty set parameter $\Gamma \geq 0$, and time step N for simulation of trajectories

for iteration $= 1, 2, \ldots,$ **do**

Simulate a set of m trajectories $\{\tau^{(i)} : i \leq m\}$ for N time steps by admitting arrivals independently with probability $p_\theta(\xi_n)$ at each time step $n \leq N$

for each time step $n = 1, \ldots, N$ of each trajectory $\tau^{(i)}$ **do**

Compute the cost $C_n^{(i)} = \sum_{k=n}^{N} \gamma^{k-n} \left[w_k^{(i)} + c_r(a_k^{(i)} - u_k^{(i)}) \right]$

Compute the gradient $g_n^{(i)} = \xi_n^{(i)} \left[u_n^{(i)} - a_n^{(i)} \left(1 + \exp(-\theta^T \xi_n^{(i)}) \right)^{-1} \right]$

end **for**

Compute the policy gradient estimate

$$\hat{g} = \frac{1}{m} \sum_{i=1}^{m} \sum_{n=1}^{N} \gamma^{n-1} C_n^{(i)} g_n^{(i)} \tag{14}$$

Update policy parameter $\theta \leftarrow \theta - \alpha \hat{g}$

end **for**

return θ

5 NUMERICAL EXPERIMENTS

We now report the findings of our numerical experiments comparing the performance of the proposed policy with that of the threshold policy. The simulations are run over a horizon N of 150 time steps with a look-ahead $\omega = 10$ time units. Each arriving job is taken to require $s = 0.25$ time units of service. The total number of arrivals within the simulation window is taken to be Poisson distributed with mean $1.2N/s$, thus representing a over-loaded system. The unconditioned variance in each arrival time is set to be $\sigma^2 = (3N/|\mathcal{A}|)^2$.

The policy gradient procedure in Algorithm 1 is run for several possible values of rejection cost c_r and the uncertainty set parameter choices $\Gamma = 2$ and 3. The resulting iterates are combined with Polyak-Ruppert iterate averaging scheme. Figure 1 reports the (rejection rate, average workload) pairs observed upon evaluating the policy output by Algorithm 1 for every choice of rejection cost c_r and Γ. The focus in Figure 1 is on regions wherein the rejection rate as well as average workloads are non-trivial. To gain an understanding on how well the output policies improve upon the baseline threshold policy, we report the (rejection rate, average workload) obtained for various choices of threshold between 0 and 15. Each of the threshold policy choice is, in turn, optimal (within the family of threshold policies) for the respective rejection rate.

We observe that for rejection rates below 0.3, our policy optimization procedure ends up with an average workload much lower than the threshold policy that attains the same rejection rate. Incorporating information on future arrivals, as explained in (Spencer et al. 2014; Xu and Chan 2016), allows admission control policies to proactively reject arrivals in periods with potential for workload buildup in the near-future. This anticipatory admission control is an explanation for why policies incorporating forecast information can lead to lower workloads for a fixed rejection rate. When allowing less-ambitious rejection rates exceeding than 0.3, the traffic load experienced by the server is lesser than 0.9 and the proposed policies do not offer significant advantage over average workloads than the threshold policy.

In Figure 2 we plot the average trajectory cost in each iteration of the outer loop in a sample run of the policy optimization procedure in Algorithm 1. We observe that the trajectory costs come down over

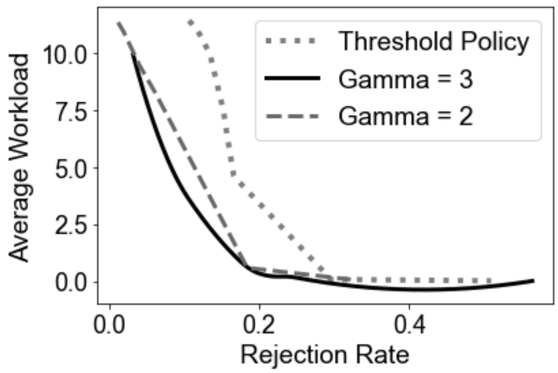

Figure 1: Comparison with threshold policies.

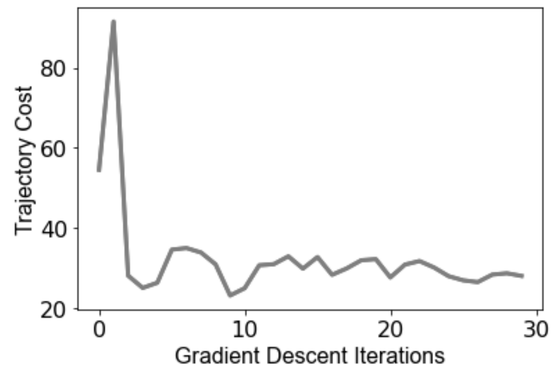

Figure 2: Decrease in trajectory cost over iterations.

time and then stabilise, indicating that a policy that works towards minimizing the overall trajectory cost is arrived at iteratively by Algorithm 1.

In Figure 3, we show a sample trajectory of the system on which admission control is performed using the proposed policy (obtained from Algorithm 1 with $\Gamma = 3$) and a threshold policy. The threshold policy was tuned to reject at least as many arrivals as our proposed policy. As the policy admit/reject arrivals differently, the workload trajectories evolve differently when the two policies are employed (despite the same number of total rejections). As can be seen in the figure, the workload attained with the proposed policy has smaller peak workloads and leads to more balanced workload over time.

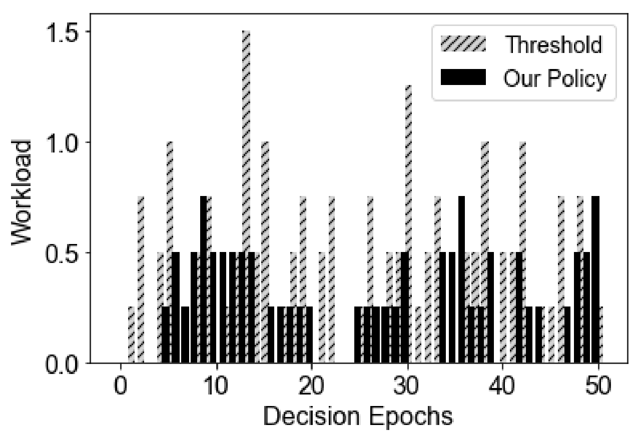

Figure 3: Comparison of sample workload trajectories.

6 CONCLUSIONS

We have considered queuing admission control wherein noisy forecasts on future arrival times become available periodically. The resulting high-dimensional Markov decision problem is handled by optimizing over a parameterized family of policies involving features motivated from conventional threshold policies and blocking-based policies. Computation of the relevant feature maps (as a function of the state variable) is shown to simplify if the future arrival times are assumed to lie in a suitable uncertainty set and robust optimization is invoked for computing the respective worst-case estimates. Experimental results demonstrate the superior performance of the proposed policy optimization over threshold policies which ignore forecast information.

Examining the scope and benefits of including a richer set of features, learnt possibly via an algorithmic representation learning procedure involving kernels or deep neural networks, is an interesting direction to consider. Obtaining a clear understanding of how optimal delay reduction is affected by the magnitude of the forecast noise and the length of look-ahead window remains an important question. Extensions to settings with lead times (where accept / redirect decisions are to be made a fixed number of time units prior to the arrival) become pertinent when considering applications from traffic and hospital emergency department congestion management.

ACKNOWLEDGEMENTS

Support from Civil Aviation Authority of Singapore via grant RTASI1904 is gratefully acknowledged.

REFERENCES

Beutler, F. J., and K. W. Ross. 1986. "Time-Average Optimal Constrained Semi-Markov Decision Processes". *Advances in Applied Probability* 18(2):341–359.

Bhatnagar, S., and K. Lakshmanan. 2012, jan. "An Online Actor–Critic Algorithm with Function Approximation for Constrained Markov Decision Processes". *Journal of Optimization Theory and Applications* 153(3):688–708.

Dai, J. G., and M. Gluzman. 2022. "Queueing Network Controls via Deep Reinforcement Learning". *Stochastic Systems* 12(1):30–67.

Johansen, S. G., and S. Stidham. 1980. "Control of Arrivals to a Stochastic Input-Output System". *Advances in Applied Probability* 12(4):972–999.

Naor, P. 1969. "The Regulation of Queue Size by Levying Tolls". *Econometrica* 37(1):15–24.

Spencer, J., M. Sudan, and K. Xu. 2014. "Queuing with future information". *The Annals of Applied Probability* 24(5):2091 – 2142.

Stidham, S. 1985. "Optimal control of admission to a queueing system". *IEEE Transactions on Automatic Control* 30(8):705–713.

Stidham, S., and N. U. Prabhu. 1974. "Optimal Control of Queueing Systems". In *Mathematical Methods in Queueing Theory*, edited by A. B. Clarke, 263–294. Berlin, Heidelberg: Springer Berlin Heidelberg.

Sun, Y., B. H. Heng, Y. T. Seow, and E. Seow. 2009, jan. "Forecasting daily attendances at an emergency department to aid resource planning". *BMC Emergency Medicine* 9(1).

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

Xu, K., and C. W. Chan. 2016. "Using Future Information to Reduce Waiting Times in the Emergency Department via Diversion". *Manufacturing & Service Operations Management* 18(3):314–331.

AUTHOR BIOGRAPHIES

KARTHYEK MURTHY is an Assistant Professor in Singapore University of Technology and Design. His research interests revolve around building models and methods for incorporating competing considerations such as risk, robustness, and fairness in data-driven optimization problems affected by uncertainty. His e-mail address is karthyek_murthy@sutd.edu.sg.

DIVYA PADMANABHAN is an Assistant Professor at Indian Institute of Technology, Goa. Her research interests are in building robust models for uncertain optimization problems with applications in healthcare, operations research and learning. Her email address is divya@iitgoa.ac.in.

SATYANATH BHAT is an Assistant Professor at Indian Institute of Technology, Goa. His research interests lie in incentive compatible learning and game theory. His e-mail address is satyanath@iitgoa.ac.in.