



LookAhead: the optimal non-decreasing index policy for a time-varying holding cost problem

Keerthana Gurushankar¹  · Zhouzi Li¹ · Mor Harchol-Balter¹ · Alan Scheller-Wolf¹

Received: 2 May 2025 / Revised: 27 November 2025 / Accepted: 19 December 2025
© The Author(s) 2026

Abstract

In practice, the cost of delaying a job can grow as the job waits. Such behavior is modeled by the time-varying holding cost (TVHC) problem, where each job's instantaneous holding cost increases with its current age (a job's age is the time since it arrived). The goal of the TVHC problem is to find a scheduling policy that minimizes the time-average total holding cost across all jobs. However, no optimality results are known for the TVHC problem outside of the asymptotic regime. In this paper, we study a simple yet still challenging special case: A two-class M/M/1 queue in which class 1 jobs incur a non-decreasing, time-varying holding cost and class 2 jobs incur a constant holding cost. Our main contribution is deriving the first optimal (non-decreasing) index policy for this special case of the TVHC problem. Our optimal policy, called LookAhead, stems from the following idea: Rather than considering each job's *current* holding cost when making scheduling decisions, we should look at their cost some X time into the future, where this X is intuitively called the “lookahead amount.” This paper derives that optimal lookahead amount.

Keywords Dynamic holding costs · $c\mu$ rule · Queueing theory · Restless multi-armed bandits · Index policy

Mathematics Subject Classification 60K25 · 68M20 · 90B36 · 91B32

K. Gurushankar and Z. Li have contributed equally to this work.

✉ Keerthana Gurushankar
kgurusha@andrew.cmu.edu

✉ Zhouzi Li
zhouzil@andrew.cmu.edu

Mor Harchol-Balter
harchol@cs.cmu.edu

Alan Scheller-Wolf
awolf@andrew.cmu.edu

¹ Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

1 Introduction

Every day, systems around us must decide which task to do next. A clinic decides which patient to see, a data center chooses which request to serve, a factory picks which order to process. In many of these settings, the cost of delaying a job increases as it waits. In a hospital emergency department, triage protocols must account for the fact that a patient with a moderate condition may become critical if not treated in time. In a cloud platform delivering live video or interactive applications, delays past a few hundred milliseconds can cause perceptible lag. In warehouses, orders with perishable goods may incur increasing holding costs due to spoilage risk or expiry windows. In all these systems, the scheduler must balance limited service capacity against growing costs of delay, and must often tradeoff between completing a short but low-urgency job and an expensive job that may take longer.

All these problems can be modeled as a stream of jobs where each job has some instantaneous holding cost (cost per second that the job is not complete). A job's holding cost may be constant or vary over time. At every moment, the service provider has to pay a total holding cost across all jobs in the system. The goal of the service provider is to minimize the time-average holding cost across all jobs.

More formally, consider the problem of scheduling jobs in a single-server queue to minimize the time-average holding cost. This is a classical objective in queueing theory and operations research. Optimal scheduling is well understood when jobs incur constant instantaneous holding cost while in system. For example, when job sizes are exponential, the well-known $c\mu$ rule is optimal [1, 2]. For many extensions such as general job size distributions, optimal scheduling is given by the Gittins index policy [3, 4].

However, when the instantaneous holding cost varies over time, no optimal policy is known. To capture cases where holding costs increase as a job is waiting, we define a job's *age* as the time that the job has spent in the system, and we let the instantaneous holding cost of a job be a function of its age. This *Time-Varying Holding Cost (TVHC)* regime was first studied in the seminal paper by Van Mieghem [5]. That paper introduces the generalized $c\mu$ rule to minimize time-average holding cost. However, the rule turns out to only be optimal in a special asymptotic regime, and so far no optimality results are known outside of asymptotic regimes (see Sect. 2).

Note that our definition of *age*, based on time in system, is in contrast to *attained service* which is sometimes called age in other literature [6, 7]. The key difficulty with the fact that holding costs increase with age (not just attained service) is that the problem intrinsically becomes a *restless multi-armed bandit* (R-MAB) problem. These problems are notoriously difficult; it is well-known that the optimal policy for such problems is usually intractable [8].

1.1 Our problem

In this paper, we study a simple setting of the TVHC problem (see Figs. 1 and 2) which still captures the key difficulty of restlessness: A 2-class M/M/1 queue in which one class of jobs has a non-decreasing holding cost function that depends on the job's age

Fig. 1 A 2-class M/M/1 queue with age-based holding costs

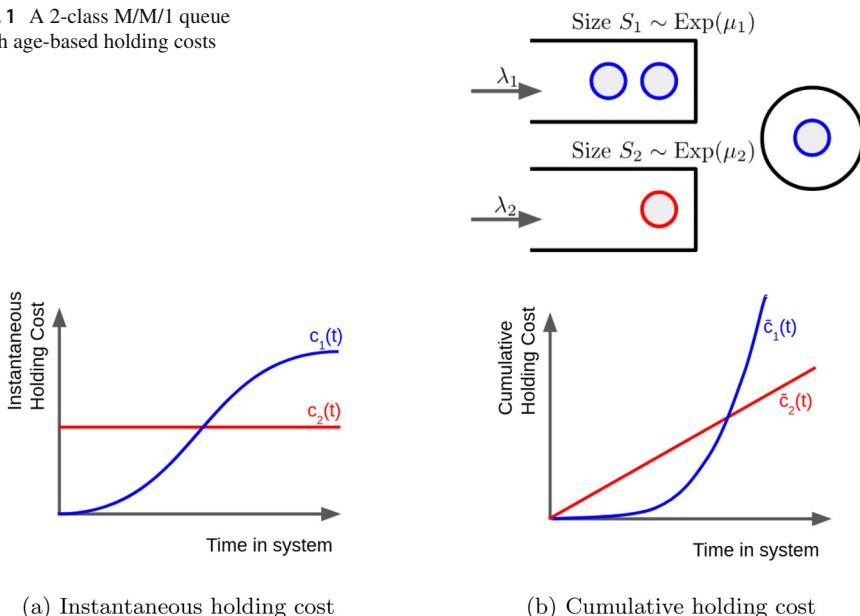


Fig. 2 (Instantaneous) holding cost and cumulative holding cost

(time in system), while the other class has a constant instantaneous holding cost. Jobs have exponential service times with class-dependent rates. Even this simple sounding problem is entirely open. Our model captures many settings in which one class has a dominant delay sensitivity. For instance, user facing jobs in computer systems serving mixed workloads alongside background processes may have important and complex latency requirements. Likewise in business operations, corrective maintenance tasks may be much more urgent than preventative tasks. Our problem models many such settings using a general non-decreasing holding cost function for one class, and a constant holding cost for the other.

In general, the optimal policy for our problem may be arbitrarily complex. It could be viewed as the solution to a Markov Decision Process (MDP) that needs to track every combination of jobs with every age. One way of simplifying the problem is to limit the search space to *index policies*. In an index policy, every job is assigned a scalar-valued *index* as a function of its state (class and age),¹ and the scheduler always serves the job with highest index. The vast majority of scheduling policies studied in the queueing literature are index policies (see e.g. [4, 6, 9–14]). Index policies are also known to be optimal in many settings, such as the Gittins index for rested bandits. However, in our setting with age-dependent holding costs, even the optimal policy among the class of index policies is unknown.

We observe that when a class's holding cost increases with age and its job sizes are exponential with unknown remaining service time, the optimal policy must serve

¹ If job sizes are not exponentially distributed, then the index can also be a function of the job's attained service, in addition to its age.

jobs within that class in *first-come-first-serve* (FCFS) order [15, Lemma 3.1]. This motivates us to restrict our search to the class of index policies which enforce FCFS within each class: Equivalently, each class has a non-decreasing index function and ties between jobs within a class are broken using FCFS. We refer to such index policies as the set of *non-decreasing index policies* in our paper, and our goal is to find the optimal policy within that set.

1.2 Toward solving our problem

We now turn to the question of finding the optimal non-decreasing index policy. Intuitively, we want something like the *cμrule*, where the index of a job with holding cost c_i and completion rate μ_i is $c_i \cdot \mu_i$. While the *cμ* rule is optimal for the case of constant holding costs, it is not defined for our setting. When holding costs are not constant, the *generalized cμrule* [5] assigns index $c_i(t) \cdot \mu_i$ to a job of class i and age t . Unfortunately, the generalized *cμ* rule is only asymptotically optimal for the TVHC problem.

It is easy to see what goes wrong in the generalized *cμ* rule: If we look at Fig. 2 when $\mu_1 = \mu_2$, a job of class 1 only gets priority over class 2 at the *instant* that the blue and red curves cross. Would it not make more sense to prioritize class 1 jobs a little earlier, so that we can get those jobs done *before* their holding cost gets really high? That is exactly the intuition behind our proposed index policy, which we call *LookAhead*. Under *LookAhead*, the index of class 1 jobs depends not on their holding cost at time t , but rather their holding cost in the future, specifically X time into the future. The “lookahead amount,” X , is defined in Theorem 1.1, below, and intuition for the amount is given later in the paper.

Theorem 1.1 (*LookAhead*) *Suppose we have a 2-class M/M/1 queue, where class i jobs arrive at rate λ_i with service rate μ_i for $i \in \{1, 2\}$. Class 1 jobs incur instantaneous holding cost at rate $c_1(t)$ when they have age t , and class 2 jobs incur holding cost at constant rate c_2 while in system. The optimal policy among the class of non-decreasing index policies is given by the following index functions:*

$$V_1(t) = \mu_1 \mathbb{E}[c_1(t + X)], \quad \text{where } X \sim \text{Exp}(\mu_1 - \lambda_1), \quad \text{and } V_2(t) = \mu_2 c_2.$$

Ultimately, our paper makes the following novel contributions:

- (i) We derive the first optimal scheduling result for a TVHC problem.
- (ii) We derive the first analysis of response time tail for an age-based scheduling policy (see Lemma 3.3).
- (iii) Our index policy coincides with the Whittle index policy for the same R-MAB as we define [15]. The Whittle index policy is typically a heuristic with only asymptotic optimality guarantees [16–18]. Thus we show a surprising case of optimality for a Whittle index policy.

Thus we establish index optimal scheduling for a specific two-class TVHC setting. Nonetheless, the amortized *cμ* interpretation presented in Sect. 3.5 as well as our

recent work on the Whittle index for TVHC [15] suggest principled heuristic policies for broader TVHC problems. In non-preemptive regimes, the same index function can be applied at departure instants. An important question which is outside the scope of this paper is to characterize the gap between index optimal scheduling and truly optimal scheduling. Because the globally optimal policy is defined over a high-dimensional continuous time state space, even empirical evaluation of this gap poses significant challenges.

2 Related work

In this section, we review related work on scheduling that informs our study of time-varying holding costs (TVHC). We group prior work into four areas:

First, we review classic scheduling policies that prioritize jobs based on service received so far (Sect. 2.1). These have been well-studied but cannot analyze age-based scheduling. Next, we discuss age-based scheduling policies (Sect. 2.2). These are closer to our setting but only consider a few simple policies, and focus on response time analysis rather than optimal scheduling. We then describe the generalized $c\mu$ rule (Sect. 2.3), the only prior work offering any optimality result for TVHC; though only in the diffusion limit. Finally, we examine Whittle index-based heuristics (Sect. 2.4) for TVHC.

2.1 Prioritizing jobs based on attained service

The past half century of queueing literature has devoted enormous energy to studying the response time of various scheduling policies in the M/G/1 queue, where the scheduling policy makes decisions based on the *service that a job has received so far* [10, Chapters 28-33]. Examples include *Shortest Remaining Processing Time (SRPT)* and *Least Attained Service (LAS)* [11]. Recently (2018), Scully et al. [6] proposed a unified analysis framework (SOAP) for a very broad class that includes almost all known scheduling policies, in which a job's priority depends on its own class, size and the service that it has received so far.

2.2 Prioritizing jobs based on age

Much less research has gone into the analysis of scheduling policies where a job's priority is based on the time it has spent in the system. The notable exception is *accumulating priority scheduling* [19–21]. Under this model, a job's priority grows linearly with the time it spends in the system (starting at priority 0). Different classes may accumulate priority at different rates, allowing an older slow-accumulating job to eventually overtake a newer fast-accumulating one.

However, jobs in accumulating priority models are not associated with a holding cost function, and the goal is not to minimize holding cost, but rather to analyze response time. Additionally, accumulating priority policies only allow analysis in settings where two jobs that arrive at the same time will never flip relative priority [21,

[22]. That said, we do borrow some of their Poisson process transformation lemmas, in our own technical arguments (such as adapting their [19, Lemma 4.2] for our Lemma 3.2).

2.3 Asymptotically optimal scheduling for the TVHC problem: generalized $c\mu$

The TVHC problem we consider was introduced in van Mieghem's seminal paper [5]. This paper provided a scheduling policy known as the *generalized $c\mu$ rule* for minimizing age-based convex holding costs. The generalized $c\mu$ rule indexes jobs by the product of their instantaneous holding cost and instantaneous service rate. This policy was shown to be asymptotically optimal for minimizing time-average total holding cost in the diffusion limit (i.e., when arrival and service rates scale as $n\lambda$ and $n\mu$ as $n \rightarrow \infty$) under heavy traffic for multi-class M/G/1 queues with convex holding costs as a function of delay [5, Proposition 8].

The generalized $c\mu$ rule is, to our knowledge, the only work to provide *any* optimality guarantees for TVHC, and only in the asymptotic scaling regimes. In contrast, our work develops an index policy that is provably optimal in finite time-scales.

Interestingly, the generalized $c\mu$ rule corresponds to our policy in the special case where the optimal lookahead parameter $X \rightarrow 0$. This happens precisely in the diffusion limit, thus our result is consistent with van Mieghem's Proposition 8, and hence is also asymptotically optimal. But under finite time-scales, we demonstrate (see Sect. 5) that our LookAhead policy can significantly outperform the generalized $c\mu$ rule.

2.4 Heuristic scheduling for TVHC: whittle

Many scheduling problems can be viewed as instances of *multi-armed bandit* (MAB) problems. In a standard Markov MAB model, each arm corresponds to a Markov process with internal states that evolve according to a transition model. The agent selects one arm to activate at each time step, accruing a cost determined by the states of all arms, and seeks to minimize the long-run average cost. In scheduling, each job or job class corresponds to an arm, and the decision is which job to serve at a given time.

The MAB framework has led to powerful results in optimal scheduling, such as the *Gittins index policy*, which is optimal for preemptive scheduling in an M/G/1 queue with unknown job size distributions [4]. These results fall under the category of *rested bandits*, where the state of an arm (i.e., a job) evolves only when the arm is pulled—equivalent to a job being served.

However, when a job's state (e.g., age) evolves even while it is not being served, the problem becomes a *restless bandit*. This setting is significantly harder: It is known to be PSPACE-hard in general, and optimal policies are difficult to characterize [8]. Several papers have proposed Whittle index-based policies for various scheduling problems [13–15, 23]. However, works in this literature only provide heuristics with only asymptotic guarantees.

In particular, Li et al. [15] provide a Whittle index policy for TVHC. This work develops a heuristic index policy based on job age and demonstrates strong empirical

performance. In fact, the policy that we analyze in this paper coincides with the one proposed in [15]. However, [15] does not establish optimality. In contrast, we provide a rigorous proof that this policy is the optimal index policy for one type of two-class TVHC problem.

We also borrow several key structural lemmas from [15], particularly in our analysis of the cost dynamics and structural monotonicity properties of the value function, which are used in our main proof arguments. To the best of our knowledge, ours is the first work to prove the optimality of an exact index policy for a dynamic, delay-based holding cost problem.

For a broader overview of bandit models and index policies in scheduling, we refer readers to [15, 24].

3 A shorter proof using an additional assumption

In this section, we present a shorter proof of our result using one additional assumption. Intuitively, since class 2 jobs have constant holding cost, it is reasonable to assume that the index of class 2 jobs should be constant. This is the crux of Assumption 3.1 (explained more rigorously below). Based on this assumption, we will now derive the optimal non-decreasing index policy. A longer and more involved proof without this assumption is provided in Sect. 4.

We structure this section as follows: In Sect. 3.1, we define Assumption 3.1. In Sect. 3.2, we discuss the structure imposed on the scheduling problem when class 1 has an arbitrary non-decreasing index function and class 2 has a constant index. We call the class of policies defined by this structure “Overtake.” In Sect. 3.3, we analyze the response time tail for any Overtake policy. In Sect. 3.4, we use the response time tail analysis to analytically solve the optimization problem of minimizing time-average holding cost among Overtake policies.

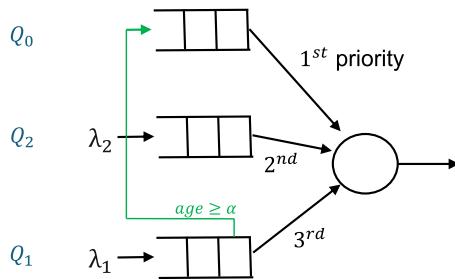
3.1 Discussion of the additional assumption

Recall that the goal of this paper is to find the optimal non-decreasing index policy. This section makes an assumption that the index for class 2 jobs should not only be non-decreasing, but furthermore be a constant.

Assumption 3.1 There exists an optimal non-decreasing index policy where the index for any class 2 job is a constant regardless of the job’s age.

Although this assumption is well-motivated, it is not obvious. For example, one may argue that the age of a class 2 job reveals some information about the number of class 2 jobs in the system, which may affect the decision. The longer proof in Sect. 4 addresses these concerns and justifies our result.

Using Assumption 3.1 allows us to restrict our optimization problem to the class of Overtake policies, which provides for a simpler proof.

Fig. 3 Overtake(α) policy

3.2 Policy class: Overtake

For any non-decreasing index policy where the class 2 index is a constant, let V_1, V_2 be the index functions of class 1 and class 2 jobs, respectively. Define α to be the youngest age where $V_1(\alpha) = v = V_2(\cdot)$, where v represents the constant index of any class 2 job. Then the index policy is equivalent to a policy which has three levels of priority: (Q_0) class 1 jobs with age $\geq \alpha$; (Q_2) all class 2 jobs; and (Q_1) class 1 jobs with age $< \alpha$. Within each level of priority, jobs are served in FCFS order. We call this policy Overtake(α). Pictorially, it can be illustrated in Fig. 3, where priority is enforced as: $Q_0 > Q_2 > Q_1$.

In this way, any non-decreasing index policy with constant class 2 index is equivalent to some Overtake policy. Thus, by Assumption 3.1, the optimal non-decreasing index policy is equivalent to some Overtake policy. Therefore, to find an optimal non-decreasing index policy, it suffices to optimize time-average total holding cost within the class of Overtake policies.

3.3 Characterizing response time tail

In this section, we derive the tail of the response time of class 1 jobs under Overtake(α). The key insight is the following lemma: The busy period of Q_0 can be seen as that of an M/M/1 queue. Note that this does not imply that Q_0 is an M/M/1 queue, because the inter-arrival times of jobs into Q_0 during a Q_0 -idle period are not exponentially distributed.

Lemma 3.2 *Let $T^{M/M/1}$ denote the response time of an M/M/1 queue with arrival rate λ_1 and job size distribution $\text{Exp}(\mu_1)$ under an FCFS policy. Then jobs that enter Q_0 will stay in Q_0 for a time distributed as $T^{M/M/1}$.*

Proof We track the dynamics of the oldest age in Q_0 and in a standard M/M/1 queue. Note that for any sample path, the oldest age trajectory uniquely determines the response time of each job (since each oldest age drop is equivalent to a job completion). Thus, it suffices to prove that the stochastic behaviors of the oldest age in Q_0 and the M/M/1 queue are the same during a Q_0 busy period.

Define the oldest age in an M/M/1 queue to be A . If $A < 0$, this means there is no job in the M/M/1 queue, and the next job arrives $(-A)$ time later. In this case, A grows with rate 1. Otherwise if $A > 0$, during each δ time step, with probability

$1 - \mu_1 \delta + o(\delta)$, no job is finished and the oldest age grows with rate 1. With probability $\mu_1 \delta + o(\delta)$, the oldest job is finished.² ³ In this case, since the inter-arrival time is $\text{Exp}(\lambda_1)$, we know that the oldest age drops by a random variable $\text{Exp}(\lambda_1)$. If the oldest age drops below 0, the busy period ends.

Similarly, in Q_0 , if there exists jobs in Q_0 , define the time the oldest job in Q_0 has spent in Q_0 to be A_0 . By definition of the Overtake(α) policy, A_0 is equal to the age of the oldest job in Q_0 minus α . Thus, a busy period in Q_0 starts with $A_0 = 0$, and then the oldest age A_0 follows the same dynamics as A in an M/M/1 queue (when $A > 0$), until A_0 drops below 0, which means that the Q_0 busy period ends. Therefore, A_0 has the same dynamics as A when they are larger than 0, which means the job completion process in Q_0 is the same as that in an M/M/1 queue during any Q_0 busy period. Thus, the time a job spends in Q_0 given it has entered Q_0 is equal to the response time in an M/M/1 queue. \square

The following lemma characterizes the tail probability of a class 1 job's response time. Using Lemma 3.2, we are able to express a class 1 job's response time in terms of the response time under a strict priority queueing system where class 2 has preemptive priority over class 1.

Lemma 3.3 (*Tail Probability of Class 1 jobs*) *Let T_1^α denote the response time of class 1 jobs under the Overtake(α) policy. Let $T_1^{\text{P-Prio}(2;1)}$ denote the response time of class 1 jobs in a preemptive priority system where class 2 jobs have strict priority over class 1 jobs. Then the tail probability of T_1^α is as follows:*

$$\mathbb{P}[T_1^\alpha > t] = \begin{cases} \mathbb{P}\left[T_1^{\text{P-Prio}(2;1)} > t\right] & \text{if } t \leq \alpha, \\ \mathbb{P}\left[T_1^{\text{P-Prio}(2;1)} > \alpha\right] \cdot e^{-(t-\alpha)(\mu_1 - \lambda_1)} & \text{otherwise.} \end{cases} \quad (1)$$

Proof We start by considering the case where $t \leq \alpha$. We will prove that

$$\mathbb{P}[T_1^\alpha \leq t] = \mathbb{P}\left[T_1^{\text{P-Prio}(2;1)} \leq t\right].$$

Let $t \leq \alpha$. For any tagged class 1 job whose response time is $\leq t$, let W^α denote the work that runs ahead of the tagged class 1 job in the Overtake(α) system. Because our tagged class 1 job does not enter Q_0 , we can express W^α as a sum of two components: (i) all the class 1 jobs which arrive before our job, and (ii) all the class 2 jobs which arrive before our tagged class 1 job completes. Therefore W^α is the same as $W^{\text{P-Prio}(2;1)}$, the work that runs ahead of our tagged class 1 job in a P-Prio(2; 1) system. Therefore, the response time of our tagged class 1 job is the same under Overtake(α) and under P-Prio(2; 1), assuming that these response times are $\leq \alpha$. We have thus shown that

$$\mathbb{P}[T_1^\alpha \leq t] \leq \mathbb{P}\left[T_1^{\text{P-Prio}(2;1)} \leq t\right].$$

² For the remainder of this paper, we use δ to refer to a small time step and omit the $o(\delta)$ terms in δ -time step discrete optimizations of our continuous time system.

³ An alternative way to describe the dynamics of A is $dA = dt - I \cdot dN_{\mu_1}(t)$, where $I \sim \text{Exp}(\lambda_1)$, and N_{μ_1} is a Poisson counting process with rate μ_1 .

Moreover, the response time of a class 1 job is smaller in any Overtake policy than in P-Prio(2; 1). Thus

$$\mathbb{P}[T_1^\alpha \leq t] \geq \mathbb{P}[T_1^{\text{P-Prio}(2;1)} \leq t].$$

These two facts show that

$$\mathbb{P}[T_1^\alpha > t] = \mathbb{P}[T_1^{\text{P-Prio}(2;1)} > t], \quad \text{for } t \leq \alpha.$$

For $t > \alpha$, we can condition on the class 1 job reaching age α (thus entering Q_0):

$$\begin{aligned} \mathbb{P}[T_1^\alpha > t] &= \mathbb{P}[T_1^\alpha > t \mid T_1^\alpha > \alpha] \cdot \mathbb{P}[T_1^\alpha > \alpha] \\ &= \mathbb{P}[T_0^\alpha > t - \alpha \mid T_1^\alpha > \alpha] \cdot \mathbb{P}[T_1^{\text{P-Prio}(2;1)} > \alpha], \end{aligned}$$

where T_0^α is the time that class 1 job stays in Q_0 .

By Lemma 3.2, we know that T_0^α is the same as the response time in an M/M/1 queue. It is well-known that the tail probability of response time in this M/M/1 queue is

$$\mathbb{P}[T^{M/M/1} > t] = e^{-(\mu_1 - \lambda_1)t}.$$

Thus we have that

$$\mathbb{P}[T_0^\alpha > t - \alpha \mid T_1^\alpha > \alpha] = \Pr[T^{M/M/1} > t - \alpha] = e^{-(\mu_1 - \lambda_1)(t - \alpha)}. \quad (2)$$

This allows us to express the tail probability when $t > \alpha$ as:

$$\begin{aligned} \mathbb{P}[T_1^\alpha > t] &= \mathbb{P}[T_0^\alpha > t - \alpha \mid T_1^\alpha > \alpha] \cdot \mathbb{P}[T_1^{\text{P-Prio}(2;1)} > \alpha] \\ &= \mathbb{P}[T_1^{\text{P-Prio}(2;1)} > \alpha] \cdot e^{-(\mu_1 - \lambda_1)(t - \alpha)}. \end{aligned}$$

□

3.4 Optimization problem

In this section, we derive the optimal Overtake policy. The proof is organized in the following order: First, using the conservation law on mean response times [25], we transform our optimization problem into terms purely dependent on T_1^α in Lemma 3.5. We then use our tail analysis from Lemma 3.3 to analytically solve the resulting optimization problem: Two lemmas are derived in Lemmas 3.6 and 3.7 and finally we prove Theorem 3.8. The optimal non-decreasing index policy is a direct corollary of Theorem 3.8, which is stated in Corollary 3.9.

We start by introducing a definition to simplify our objective.

Definition 3.4 (\widehat{Cost}) Let $\bar{c}_1(t) = \int_0^t c_1(s) ds$ refer to the cumulative holding cost incurred by a class 1 job when it has reached age t . We define $\widehat{Cost} := \mathbb{E}[\bar{c}_1(T_1^\alpha)] - c_2 \frac{\mu_2}{\mu_1} \mathbb{E}[T_1^\alpha]$.

Now we apply the conservation law to the objective.

Lemma 3.5 *Minimizing time-average total holding cost is equivalent to minimizing \widehat{Cost} .*

Proof Since preemption is allowed, any reasonable policy is work conserving. Thus, we can apply the conservation law [25]:

$$\rho_1 \mathbb{E}[T_1^\alpha] + \rho_2 \mathbb{E}[T_2^\alpha] = W, \quad (3)$$

where W is the time-average amount of work in an M/G/1 system with both class 1 jobs and class 2 jobs, which can be expressed as

$$W = \frac{1}{1 - \rho} \left(\frac{\lambda_1}{\mu_1^2} + \frac{\lambda_2}{\mu_2^2} \right).$$

From (3), we have that

$$\mathbb{E}[T_2^\alpha] = \frac{W}{\rho_2} - \frac{\rho_1}{\rho_2} \mathbb{E}[T_1^\alpha].$$

This means the time-average total holding cost is equal to

$$\begin{aligned} \mathbb{E}[Cost] &= \lambda_1 \mathbb{E}[\bar{c}_1(T_1^\alpha)] + \lambda_2 c_2 \mathbb{E}[T_2^\alpha] \\ &= \lambda_1 \mathbb{E}[\bar{c}_1(T_1^\alpha)] + \lambda_2 c_2 \left(\frac{W}{\rho_2} - \frac{\rho_1}{\rho_2} \mathbb{E}[T_1^\alpha] \right) \\ &= \lambda_1 \widehat{Cost} + \lambda_2 c_2 \frac{W}{\rho_2}. \end{aligned}$$

This means minimizing the time-average total holding cost is equivalent to minimizing \widehat{Cost} . \square

We give a basic formula for exponential random variables.

Lemma 3.6 *For any smooth function f and exponential variable X , we have that*

$$\mathbb{E}[f(x_0 + X)] - f(x_0) = \mathbb{E}[X] \mathbb{E}[f'(x_0 + X)].$$

$$f(x_0) - \mathbb{E}[f(x_0 - X)] = \mathbb{E}[X] \mathbb{E}[f'(x_0 - X)].$$

Proof Suppose $X \sim \text{Exp}(\theta)$. Then we have that

$$\mathbb{E}[f(x_0 + X)] - f(x_0) = \int_0^\infty f'(x_0 + t) \mathbb{P}[X > t] dt$$

$$\begin{aligned}
&= \int_0^\infty f'(x_0 + t) e^{-\theta t} dt \\
&= \frac{1}{\theta} \int_0^\infty f'(x_0 + t) \theta e^{-\theta t} dt \\
&= \mathbb{E}[X] \mathbb{E}[f'(x_0 + X)].
\end{aligned}$$

The other equation is derived similarly. \square

Lemma 3.7 For any differentiable function $c(\cdot)$,

$$\frac{d\mathbb{E}[c(T_1^\alpha)]}{d\alpha} = \mathbb{E}[c'(\alpha + \text{Exp}(\mu_1 - \lambda_1))] \frac{d\mathbb{E}[T_1^\alpha]}{d\alpha}.$$

Proof We combine Lemmas 3.6 and 3.3.

$$\begin{aligned}
\mathbb{E}[c(T_1^\alpha)] &= c(0) + \int_0^\infty c'(t) \Pr[T_1^\alpha > t] dt \\
&= c(0) + \int_0^\alpha c'(t) \Pr[T_1^\alpha > t] dt + \int_\alpha^\infty c'(t) \Pr[T_1^\alpha > t] dt \\
&= c(0) + \int_0^\alpha c'(t) \Pr[T_1^{\text{P-Prio}(2:1)} > t] dt \\
&\quad + \Pr[T_1^{\text{P-Prio}(2:1)} > \alpha] \int_0^\infty c'(\alpha + s) e^{-(\mu_1 - \lambda_1)s} ds \quad (\text{by Lemma 3.3}) \\
&= c(0) + \int_0^\alpha c'(t) \Pr[T_1^{\text{P-Prio}(2:1)} > t] dt \\
&\quad + \Pr[T_1^{\text{P-Prio}(2:1)} > \alpha] \frac{1}{\mu_1 - \lambda_1} \mathbb{E}[c'(\alpha + \text{Exp}(\mu_1 - \lambda_1))] \\
&= c(0) + \int_0^\alpha c'(t) \Pr[T_1^{\text{P-Prio}(2:1)} > t] dt + \Pr[T_1^{\text{P-Prio}(2:1)} \\
&\quad > \alpha] (\mathbb{E}[c(\alpha + \text{Exp}(\mu_1 - \lambda_1))] - c(\alpha)) \quad (\text{by Lemma 3.6})
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{d\mathbb{E}[c(T_1^\alpha)]}{d\alpha} &= c'(\alpha) \Pr[T_1^{\text{P-Prio}(2:1)} > \alpha] + \frac{d\Pr[T_1^{\text{P-Prio}(2:1)} > \alpha]}{d\alpha} \\
&\quad (\mathbb{E}[c(\alpha + \text{Exp}(\mu_1 - \lambda_1))] - c(\alpha)) + \Pr[T_1^{\text{P-Prio}(2:1)} > \alpha] \\
&\quad (\mathbb{E}[c'(\alpha + \text{Exp}(\mu_1 - \lambda_1))] - c'(\alpha)) \\
&\quad (\text{Dominated Convergence Theorem}) \\
&= \frac{d\Pr[T_1^{\text{P-Prio}(2:1)} > \alpha]}{d\alpha} (\mathbb{E}[c(\alpha + \text{Exp}(\mu_1 - \lambda_1))] - c(\alpha)) \\
&\quad + \Pr[T_1^{\text{P-Prio}(2:1)} > \alpha] \mathbb{E}[c'(\alpha + \text{Exp}(\mu_1 - \lambda_1))] \\
&= \left(\Pr[T_1^{\text{P-Prio}(2:1)} > \alpha] + \frac{d\Pr[T_1^{\text{P-Prio}(2:1)} > \alpha]}{d\alpha} \frac{1}{\mu_1 - \lambda_1} \right)
\end{aligned}$$

$$\mathbb{E}[c'(\alpha + \text{Exp}(\mu_1 - \lambda_1))]. \quad (\text{by Lemma 3.6})$$

Note that this equation holds for any function c . By substituting the function $c(t) = t$, we have that

$$\frac{d\mathbb{E}[T_1^\alpha]}{d\alpha} = \Pr[T_1^{\text{P-Prio}(2:1)} > \alpha] + \frac{d\Pr[T_1^{\text{P-Prio}(2:1)} > \alpha]}{d\alpha} \frac{1}{\mu_1 - \lambda_1}.$$

Thus, we have that

$$\frac{d\mathbb{E}[c(T_1^\alpha)]}{d\alpha} = \frac{d\mathbb{E}[T_1^\alpha]}{d\alpha} \cdot \mathbb{E}[c'(\alpha + \text{Exp}(\mu_1 - \lambda_1))].$$

□

Finally, we characterize the optimal Overtake policy.

Theorem 3.8 (*Optimal Overtake policy*) *The optimal Overtake policy, which we call Overtake(α^*), is characterized as follows:*

- If $\mathbb{E}[c_1(\text{Exp}(\mu_1 - \lambda_1))] > \frac{c_2\mu_2}{\mu_1}$, the optimal α^* is 0, which means that giving full priority to class 1 jobs is optimal.
- If $\lim_{t \rightarrow \infty} c_1(t) < \frac{c_2\mu_2}{\mu_1}$, the optimal α^* goes to infinity, which means that giving full priority to class 2 jobs is optimal.
- Otherwise, the optimal α^* satisfies

$$\mathbb{E}[c_1(\alpha^* + \text{Exp}(\mu_1 - \lambda_1))] = \frac{c_2\mu_2}{\mu_1}. \quad (4)$$

Further, the time-average total holding cost is convex in overtaking age.

Proof Define $c(t) = \mu_1\bar{c}_1(t) - \mu_2c_2t$. Thus we have that for policy Overtake(α), $\widehat{Cost} = \mathbb{E}[c(T_1^\alpha)]$. By Lemma 3.7, we know

$$\frac{d\widehat{Cost}}{d\alpha} = (\mu_1\mathbb{E}[c_1(\alpha + \text{Exp}(\mu_1 - \lambda_1))] - \mu_2c_2) \frac{d\mathbb{E}[T_1^\alpha]}{d\alpha}.$$

Since increasing α makes class 1 response times strictly larger, we have that $\frac{d\mathbb{E}[T_1^\alpha]}{d\alpha} > 0$. Moreover, since c_1 is non-decreasing, $\mu_1\mathbb{E}[c_1(\alpha + \text{Exp}(\mu_1 - \lambda_1))] - \mu_2c_2$ is non-decreasing in α .

Now if $\frac{d\widehat{Cost}}{d\alpha} > 0$ for all $\alpha \geq 0$, then $\alpha^* = 0$ is optimal. Likewise, if $\frac{d\widehat{Cost}}{d\alpha} < 0$ for all $\alpha \geq 0$, the optimal α goes to infinity, which means full priority to class 2 jobs is optimal. Otherwise $\frac{d\widehat{Cost}}{d\alpha}$ crosses 0 exactly once at α^* defined in Equation (4). This means the policy Overtake(α^*) is the optimal Overtake policy. □

Finally, the optimal non-decreasing index policy follows directly from Assumption 3.1 and Theorem 3.8. In particular, we call this policy LookAhead policy.

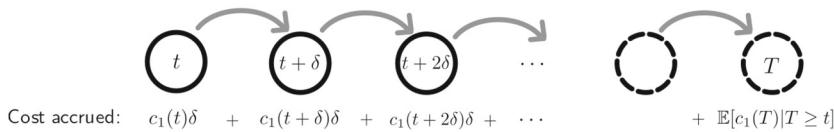


Fig. 4 Expected additional cost accrued by a class 1 job which has reached age t

Corollary 3.9 *Given Assumption 3.1, the optimal policy, called LookAhead, among the class of non-decreasing index policies is given by the following index functions:*

$$V_1(t) = \mu_1 \mathbb{E}[c_1(t + X)], \quad \text{where } X \sim \text{Exp}(\mu_1 - \lambda_1), \quad \text{and } V_2(t) = \mu_2 c_2.$$

Proof Given Assumption 3.1, the optimal non-decreasing index policy has a constant class 2 index. Therefore, it is equivalent to some Overtake policy. Moreover, the index policy defined by index functions $V_1(t) = \mu_1 \mathbb{E}[c_1(t + X)]$ and $V_2(t) = \mu_2 c_2$ is equivalent to $\text{Overtake}(\alpha^*)$, the optimal Overtake policy. Thus we have the proof. \square

3.5 An intuitive interpretation of the LookAhead policy via amortized cost

Our policy stated in Corollary 3.9 can be seen as an *amortized $c\mu$ rule*.

A class 1 job with age t will accrue an expected remaining holding cost of

$$\text{expected remaining holding cost} = \mathbb{E}[\bar{c}_1(T_1)|T_1 \geq t] - \bar{c}_1(t).$$

It will accrue this cost over its expected remaining time in system, where

$$\text{expected remaining time in system} = \mathbb{E}[T_1|T_1 \geq t] - t.$$

Suppose the job were to incur its expected remaining holding cost over its expected remaining time in system at a constant instantaneous holding cost rate, then the *effective* expected cost rate (see Fig. 4) would be

$$c_1^{\text{eff}}(t) = \frac{\mathbb{E}[\bar{c}_1(T_1)|T_1 \geq t] - \bar{c}_1(t)}{\mathbb{E}[T_1|T_1 \geq t] - t}.$$

If we defined c_2^{eff} similarly, we would have $c_2^{\text{eff}}(t) = c_2$ since $c_2(t)$ is constant. We could then apply the $c\mu$ rule to the *constant* holding costs c_1^{eff} and c_2^{eff} by always running the job with the highest $\mu_i c_i^{\text{eff}}(t)$. Note that this policy definition is recursive since the response time distribution depends on the policy, and we are stipulating a policy which depends on the response time.

Interestingly, it turns out that our LookAhead policy is an amortized $c\mu$ rule policy for the problem setting in our paper. To see this, note that since $c_2(t)$ is constant, we

will always have $c_2^{\text{eff}}(t) = c_2$. Further, if $c_1(\cdot)$ is increasing,

$$c_1^{\text{eff}}(t) = \frac{\mathbb{E}[\int_t^{T_1|T_1 \geq t} c_1(s) ds]}{\mathbb{E}[\int_t^{T_1|T_1 \geq t} 1 ds]}$$

is also increasing. As discussed in Sect. 3.2 and Fig. 3, this means the policy reduces to an Overtake(α) policy for α such that $\mu_1 c_1^{\text{eff}}(\alpha) = \mu_2 c_2$. Then as discussed in Lemma 3.2, class 1 jobs enter an M/M/1 queue past their overtaking age and experience response time $[T_1|T_1 \geq \alpha] = \alpha + X$, where $X \sim \text{Exp}(\mu_1 - \lambda_1)$. Thus we have

$$\mu_1 c_1^{\text{eff}}(\alpha) = \mu_1 \cdot \frac{\mathbb{E}[\bar{c}_1(\alpha + X) - \bar{c}_1(\alpha)]}{\mathbb{E}[X]} = \mu_1 \mathbb{E}[c_1(\alpha + X)] \text{ (by Lemma 3.6)}$$

Therefore the amortized $c\mu$ rule's overtaking age α is exactly the optimal overtaking age α^* derived in Theorem 3.8. Equivalently, we can say that our LookAhead policy always serves the job with highest amortized $c\mu$ index. This view gives us intuition for *why* the optimal LookAhead amount is X , where X is the response time of an M/M/1 queue with only class 1 jobs.

4 A longer proof without assumption 3.1

In Sect. 3, we proved that our policy is optimal under Assumption 3.1. That is, our policy is optimal among those policies which have a non-decreasing index function for class 1 and a constant index for class 2. In this section, we relax Assumption 3.1. Namely, we show that our policy LookAhead (which is equivalently Overtake(α^*)) is optimal among policies where both index functions, V_1 and V_2 , are non-decreasing. Recall that Overtake(α^*) is defined in Theorem 3.8. Specifically, in this section we only consider the interesting case where α^* satisfies (4). The proof of optimality for the case when $\alpha^* = 0$ is a straightforward modification of our proof. For the case when $\alpha^* \rightarrow \infty$, an interchange argument can be used to prove that P-Prio(2;1) is optimal.

4.1 Translating to R-MAB

We begin by translating our holding cost problem to an R-MAB problem provided by [15, Theorem 1]. As stated in Lemma 3.5, our holding cost minimization problem is equivalent to a holding cost minimization problem where the holding cost of class 1 is $c(t) = \mu_1 c_1(t) - c_2 \mu_2$, and class 2 jobs have zero holding cost. Applying [15, Theorem 1] to the new holding costs yields the following bandit.

Theorem 4.1 (Theorem 1, [15]) *For any set of non-decreasing index functions $\{V_i(\cdot)\}_{i=1}^k$, the corresponding index policies (breaking ties by FCFS) in the TVHC problem incur the same cost as in the following R-MAB problem:*

- The R-MAB has 2 arms, each representing a class. There are two actions for each arm (active or passive), where arm i is active means the oldest class i job is served, and passive means the job served at this moment is of the other type.
- *States:* At any time, we can represent the state of the R-MAB system by the age of the oldest job of each class. Namely, we can represent the system state space as $\mathcal{S} = \mathbb{R}^2$, where (t_1, t_2) represents the state where the oldest class i job has age t_i for $i \in \{1, 2\}$. If there is no type i job in the system, t_i is negative, which means the next type i arrival happens $(-t_i)$ time later. Note that for arm i , the action active is only allowed when t_i is positive.
- *Transition Probability:* If the action for arm i is passive, the i th arm state grows with rate 1. Otherwise if arm i is active, the i th arm state may drop an $\text{Exp}(\lambda_i)$ amount according to a Poisson process (when completions happen). Mathematically, the transition function is

$$\text{passive: } dt_i = dt, \quad \text{active: } dt_i = dt - I \cdot dN_{\mu_i}(t),$$

where $I \sim \text{Exp}(\lambda_i)$, and N_{μ_i} is a Poisson counting process with rate μ_i .

This transition function can be interpreted in the TVHC problem as follows: A passive action means that the oldest class i job is not in service and its age grows with rate 1. If the action is active, then in the next dt time period, the oldest class i job is served. There is a probability of $\mu_i dt$ that the job is completed and leaves the system, in which case the oldest class i job in the system becomes the previously second-oldest class i job. Since the inter-arrival time follows the distribution $\text{Exp}(\lambda_i)$, the age of the oldest job drops by an $\text{Exp}(\lambda_i)$ amount.

- *Cost Function $r(t_1)$:* for any system state (t_1, t_2) , define the cost incurred in the system to be

$$r(t_1) := c(t_1) + \mathbb{E}\left[\sum_{j=1}^N c(Y_j)\right], \quad (5)$$

where $Y_j = \sum_{m=1}^j I_m$, $I_m \sim \text{Exp}(\lambda_1)$, and N is the random variable denoting the smallest stopping time such that $Y_{N+1} \geq t_1$. Observe that the second term of (5) represents the expected holding cost of all class 1 jobs present, except for the oldest class 1 job.

Note that the cost function $r(t_1)$ depends only on the class 1 state, t_1 . This is consistent with our reformulation of the problem where class 1 has holding cost $c(t)$ and class 2 has zero holding cost.

- *Objective:* The objective is to minimize the long-run expected cost.

4.2 Bellman criteria for optimality

Note that the length of a busy period of a queueing system is identical for any work conserving policy. Thus, since our objective is to minimize the long-run cost incurred in a renewal system, by the Renewal Reward Theorem, it suffices to minimize the cost incurred in each busy period.

A policy π specifies which arm to pull at each state. Specifically, if π pulls arm 1 at state (t_1, t_2) , we define $\pi(t_1, t_2) = 1$. Let $V_\pi(t_1, t_2)$ be the cost incurred starting from state (t_1, t_2) , until the end of the first busy period (i.e., the first time $t_1, t_2 < 0$), under policy π . Define the *Q-value* of a policy π , an action a and a state s to be the total cost incurred until the end of the busy period, given that we start in state s , do action a for the next δ time, and act according to policy π thereafter, until the end of the busy period. Then the Q-values of policy π for pulling arm 1 or 2 in state (t_1, t_2) can be expressed as follows:

$$\begin{aligned} Q((t_1, t_2), |1, \delta, \pi|) &= r(t_1)\delta + (1 - \mu_1\delta)V_\pi(t_1 + \delta, t_2 + \delta) \\ &\quad + \mu_1\delta\mathbb{E}[V_\pi(t_1 + \delta - I_1, t_2 + \delta)] + o(\delta), \\ Q((t_1, t_2), |2, \delta, \pi|) &= r(t_1)\delta + (1 - \mu_2\delta)V_\pi(t_1 + \delta, t_2 + \delta) \\ &\quad + \mu_2\delta\mathbb{E}[V_\pi(t_1 + \delta, t_2 + \delta - I_2)] + o(\delta), \end{aligned} \quad (6)$$

where $I_i \sim \text{Exp}(\lambda_i)$ and $|i, \delta, \pi|$ refers to the policy which pulls arm i for the first δ time, then follows policy π .

By Bellman optimality, a policy π^* is optimal if and only if the following Bellman criteria is satisfied for any $t_1, t_2 \geq 0$ [26]:

$$\begin{aligned} \text{if } \pi^*(t_1, t_2) = 1, \text{ then } \lim_{\delta \rightarrow 0} \frac{Q((t_1, t_2), |1, \pi^*|) - Q((t_1, t_2), |2, \pi^*|)}{\delta} &\leq 0, \\ \text{if } \pi^*(t_1, t_2) = 2, \text{ then } \lim_{\delta \rightarrow 0} \frac{Q((t_1, t_2), |1, \pi^*|) - Q((t_1, t_2), |2, \pi^*|)}{\delta} &\geq 0. \end{aligned}$$

By (6), this is equivalent to

$$\begin{aligned} \text{if } \pi^*(t_1, t_2) = 1, \text{ then } \mu_1(V_{\pi^*}(t_1, t_2) - E[V_{\pi^*}(t_1 - I_1, t_2)]) &\geq \mu_2(V_{\pi^*}(t_1, t_2) - E[V_{\pi^*}(t_1, t_2 - I_2)]), \\ \text{if } \pi^*(t_1, t_2) = 2, \text{ then } \mu_2(V_{\pi^*}(t_1, t_2) - E[V_{\pi^*}(t_1 - I_1, t_2)]) &\leq \mu_2(V_{\pi^*}(t_1, t_2) - E[V_{\pi^*}(t_1, t_2 - I_2)]). \end{aligned} \quad (7)$$

In the following subsections, we will show that our policy Overtake(α^*) satisfies the above Bellman optimality criteria for the R-MAB problem.

Overtake(α^*) prefers arm 1 over arm 2 in state (t_1, t_2) iff $t_1 \geq \alpha^*$. For ease of notation, we omit expectations and for example write $V(t_1 - I_1, t_2 - I_2)$ to refer to the expectation $\mathbb{E}_{I_1, I_2} [V_{\text{Overtake}(\alpha^*)}(t_1 - I_1, t_2 - I_2)]$. Under Overtake(α^*), (7) transform into the following claims: For all $t_2 \geq 0$,

- Claim 1. If $t_1 \geq \alpha^*$, $\mu_1(V(t_1, t_2) - V(t_1 - I_1, t_2)) \geq \mu_2(V(t_1, t_2) - V(t_1, t_2 - I_2))$.
- Claim 2. If $0 \leq t_1 < \alpha^*$, $\mu_1(V(t_1, t_2) - V(t_1 - I_1, t_2)) \leq \mu_2(V(t_1, t_2) - V(t_1, t_2 - I_2))$.

The following definition provides notation for a busy period started by some initial work. We will use this notation throughout the rest of this section.

Definition 4.2 (Busy period with initial work) We denote the length of a busy period started by W_0 initial work, and with Poisson arrivals of rate λ and job-size distribution S , as $\text{BP}[W_0; \lambda, S]$.

We also provide notation characterizing the 2-class M/M/1 queue, which we will use throughout the rest of this section.

Definition 4.3 (λ, S) Let the total mean arrival rate to the two class M/M/1 queue be denoted as

$$\lambda = \lambda_1 + \lambda_2.$$

Let the job size distribution of class i be $S_i \sim \text{Exp}(\mu_i)$ for $i \in \{1, 2\}$. The overall job size distribution is denoted

$$S = \begin{cases} S_1, & \text{with probability } \frac{\lambda_1}{\lambda}, \\ S_2, & \text{with probability } \frac{\lambda_2}{\lambda}. \end{cases}$$

4.3 Verifying claim 1

The goal of this section is to verify Claim 1 (Proposition 4.16). We start by defining the notation for the total class 1 work arriving during a t period of time.

Definition 4.4 Define $\widehat{W}(t)$ to be the total class 1 work arriving during a t period of time. Mathematically,

$$\widehat{W}(t) := \sum_{i=1}^N S_1^{(i)}, \text{ where } S_1^{(i)} \sim \text{Exp}(\mu_1), N \sim \text{Pois}(\lambda_1 t).$$

We define a random process A . Intuitively, A represents the dynamics of the oldest age in an M/M/1 queue with arrival rate λ_1 and job size distribution $\text{Exp}(\mu_1)$.

Definition 4.5 (A) Given an initial state $A(0)$, we define the following random process A : At each time t ,

- if $A(t) < 0$, then $dA(t) = dt$;
- if $A(t) \geq 0$, then $dA(t) = dt - I \cdot dN_{\mu_1}(t)$, where $I \sim \text{Exp}(\lambda_1)$ and N_{μ_1} is a Poisson counting process with rate μ_1 .

Now we introduce a lemma characterizing the value function of Overtake(α^*).

Lemma 4.6 For any $t_1 \geq \alpha^*$,

$$V(t_1, t_2) = \mathbb{E} \left[\int_0^{\text{BP}[S_1 + S_2 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \middle| A(0) = t_1 - \alpha^* \right] + V(\alpha^* - I_1, t_2 - I_2) \quad (8)$$

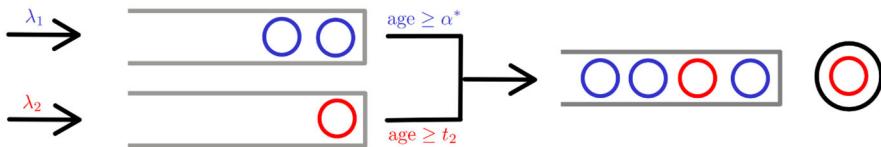


Fig. 5 Only jobs in the active queue (on the right) are visible to the server

$$V(t_1, t_2 - I_2) = \mathbb{E} \left[\int_0^{\text{BP}[S_1 + \hat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* \right] + V(\alpha^* - I_1, t_2 - I_2) \quad (9)$$

$$V(t_1 - I_1, t_2) = \mathbb{E} \left[\int_0^{\text{BP}[S_2 + \hat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* - I_1 \right] + V(\alpha^* - I_1, t_2 - I_2) \quad (10)$$

Proof We view the system through “glasses” where we only see class 1 jobs of age $\geq \alpha^*$ and class 2 jobs of age $\geq t_2$, as in Fig. 5. We call these jobs *active*.

In state (t_1, t_2) , the system starts with one visible job of each class, and Poisson arrivals behind these oldest jobs. Under the policy Overtake(α^*), we will serve only active jobs as long as we see them, and we will serve them P-Privo(1; 2) for this time. Further, while we are only serving active jobs, we will see Poisson arrivals of both classes to the active queue.

Since the initial active work is the oldest job of each class (of age, respectively, t_1 and t_2), and also any younger class 1 jobs with age larger than α^* , the total initial active work is $S_1 + S_2 + \hat{W}(t_1 - \alpha^*)$. Thus we will serve active jobs for a total busy period $\text{BP}[S_1 + S_2 + \hat{W}(t_1 - \alpha^*); \lambda, S]$. The busy period ends at a departure, when the next oldest job of each class is younger than α^* and t_2 , respectively: The ages of the oldest jobs in class 1 and class 2 are, respectively, distributed as $\alpha^* - I_1$ and $t_2 - I_2$ at the end of the busy period. This means at the end of the busy period, the system state is distributed as $(\alpha^* - I_1, t_2 - I_2)$.

Similarly, in states $(t_1, t_2 - I_2)$ and $(t_1 - I_1, t_2)$, we serve jobs for busy periods $\text{BP}[S_1 + \hat{W}(t_1 - \alpha^*); \lambda, S]$ and $\text{BP}[S_2 + \hat{W}(t_1 - \alpha^*); \lambda, S]$, respectively. At the end of the busy period, we reach the same state distribution $(\alpha^* - I_1, t_2 - I_2)$. \square

Thus, using this lemma, to prove Proposition 4.16 it suffices to characterize the expectation terms in Lemma 4.6. The approach we adopt is: We strategically introduce a discount factor β into the expectation terms, and use Lemmas from [15] to characterize the discounted terms. Then we take the limit $\beta \rightarrow 0$ to obtain the expressions for the expectation terms in Lemma 4.6.

4.3.1 Introducing the discount factor

In order to characterize the expectation terms in Lemma 4.6, we introduce a discount factor β and define the following discounted version of our expectations.

Definition 4.7 For any $t_1 \geq \alpha^*$, define

$$\Omega_\beta(t_1, W) = \mathbb{E} \left[\int_0^{\text{BP}[W; \lambda; S]} r(\alpha^* + A(t)) e^{-\beta t} dt \mid A(0) = t_1 - \alpha^* \right]. \quad (11)$$

Note that the expectation terms in Lemma 4.6 can be captured by Ω_β by taking the limit $\beta \rightarrow 0$. We use the following term as an intermediate step to get the expression for Ω_β .

Definition 4.8 For any $t_1 \geq \alpha^*$, define

$$U_\beta(t_1) = \mathbb{E} \left[\int_0^\infty r(\alpha^* + A(t)) e^{-\beta t} dt \mid A(0) = t_1 - \alpha^* \right]. \quad (12)$$

The following subsection manipulates several lemmas from [15] to get an expression of U_β , and then in Sect. 4.3.3 we return to Ω_β and finally the expectation terms in Lemma 4.6.

4.3.2 Characterizing U_β

In this subsection, lemmas from [15] are used to characterize U_β . We adapt the following definitions from [15].

Definition 4.9 (Definitions in [15]) For any variable Y , define $\tilde{Y}(x) := \mathbb{E}[e^{-xY}]$ to be its Laplace transform. Define

$$\gamma_1 := \mathbb{E}[e^{-\beta \text{BP}[S_1; \lambda_1; S_1]}] = \widetilde{\text{BP}}[S_1; \lambda_1; S_1](\beta), \quad (13)$$

$$\gamma_2 := \mathbb{E}[e^{-\beta I_1}] = \widetilde{I}_1(\beta), \quad (14)$$

$$X \sim \text{Exp} \left(\frac{\beta}{1 - \gamma_1} \right), \quad (15)$$

$$\text{Cost}(t, \beta) := \mathbb{E} \left[\int_0^{\text{BP}[S_1; \lambda_1; S_1]} r(t + A(t)) \cdot \beta e^{-\beta t} dt \mid A(0) = 0 \right], \quad (16)$$

$$\Gamma(t, \beta) := \mathbb{E} \left[\int_0^{I_1} r(t - I_1 + t) \cdot \beta e^{-\beta t} dt \right], \quad (17)$$

$$\Phi(t, \alpha^*) := \mathbb{E} \left[\sum_{i=1}^M \gamma_1^{i-1} \text{Cost}(y_i, \beta) \right], \quad (18)$$

where $y_1 = t - \text{Exp}(\lambda_1)$, $y_{i+1} = y_i - \text{Exp}(\lambda)$, and M is the random variable such that $y_M \geq \alpha^*$, $y_{M+1} < \alpha^*$, with $M \sim \text{Pois}(\lambda_1(t - \alpha^*))$.

The following lemma from [15] gives the closed form of these terms.

Lemma 4.10 (Lemmas in [15]) For the terms defined in Definition 4.9, we have the following expressions:

$$\bar{Cost}(t, \beta) = (1 - \gamma_1) \mathbb{E}[r(t + X)], \quad \text{where } X \sim \text{Exp}\left(\frac{\beta}{1 - \gamma_1}\right), \quad (19)$$

$$\Gamma(t, \beta) = (1 - \gamma_2) \mathbb{E}[r(t - I_1)], \quad (20)$$

$$\Phi(t, \alpha^*) = \int_{\alpha^*}^t e^{-\lambda_1(1-\gamma_1)(t-s)} \lambda_1 \bar{Cost}(s, \beta) ds, \quad (21)$$

$$\mathbb{E}[\gamma_1^M] = PGF(\gamma_1) = e^{-\lambda(t-\alpha^*)(1-\gamma_1)}. \quad (22)$$

Proof See Lemma 4.17, 4.18 and Case 2 of Lemma E.1 in [15]. \square

We can use the terms defined in Definition 4.9 to get the characterization of U_β .

Lemma 4.11 For any $t_1 \geq \alpha^*$,

$$U_\beta(\alpha^* - I_1) = \frac{1}{1 - \gamma_1 \gamma_2} \left(\frac{1}{\beta} \Gamma(\alpha^*, \beta) + \gamma_2 \frac{1}{\beta} \bar{Cost}(\alpha^*, \beta) \right) \quad (23)$$

$$U_\beta(t_1) = \frac{1}{\beta} (\bar{Cost}(t, \beta) + \gamma_1 \Phi(t_1, \alpha^*)) + \mathbb{E}[\gamma_1^{M+1}] U_\beta(\alpha^* - I_1). \quad (24)$$

$$U_\beta(t_1 - I_1) = \frac{1}{\beta} \Phi(t_1, \alpha^*) + \mathbb{E}[\gamma_1^M] U_\beta(\alpha^* - I_1). \quad (25)$$

Proof We start by proving (23). Starting at state $\alpha^* - I_1$, there is first an I_1 period of time before $A(t)$ returns to 0. During this period of time, $\frac{1}{\beta} \Gamma(\alpha^*, \beta)$ cost is incurred. Then starting at $A = 0$, by definition, a cost of $\frac{1}{\beta} \bar{Cost}(\alpha^*, \beta)$ is incurred during a $BP(S_1; \lambda_1; S_1)$ time period. Finally, A returns to the state $\alpha^* - I_1$ (because each drop is $\text{Exp}(\lambda_1)$). Given that γ_1, γ_2 are, respectively, the discounting factor after time period I_1 and $BP(S_1; \lambda_1; S_1)$, we have that

$$U_\beta(\alpha^* - I_1) = \frac{1}{\beta} \Gamma(\alpha^*, \beta) + \gamma_2 \frac{1}{\beta} \bar{Cost}(\alpha^*, \beta) + \gamma_1 \gamma_2 U_\beta(\alpha^* - I_1).$$

This proves the first equation.

To prove (24), we use a similar argument as in Lemma E.1 in [15]: Starting from $A = t_1$, the policy stays active until the state drops below t_1 . During this time ($BP(S_1; \lambda_1; S_1)$), a cost of $\bar{Cost}(t_1, \beta)$ is incurred. After the state drops below t_1 , the amount it is below t_1 follows an exponential distribution with rate λ_1 . Suppose it is $y_1 \sim t_1 - \text{Exp}(\lambda_1)$. If y_1 is still larger than α^* , the policy stays active until the state drops below y_1 , incurring another $\bar{Cost}(y_1, \beta)$ cost. This process continues until the state drops below α^* . Note that each time a $\bar{Cost}(y_i, \beta)$ is incurred, the state drops by an exponential amount, thus the total number of iterations follow a Poisson distribution $\text{Pois}(\lambda_1(t_1 - \alpha^*))$. After the final iteration, the state is $\text{Exp}(\lambda_1)$ below α^* . This process yields (24).

Equation (25) can be proved by a similar argument. \square

Finally, we take the limit $\beta \rightarrow 0$ on the expressions. We first give the following well-known lemma characterizing the limit of a Laplace transform.

Lemma 4.12 *Given a random variable Y , we have that*

$$\lim_{\beta \rightarrow 0} \frac{1 - \tilde{Y}(\beta)}{\beta} = \mathbb{E}[Y].$$

Proof The proof follows immediately from L'Hospital's Rule and the property of the Laplace transform:

$$\lim_{\beta \rightarrow 0} \frac{1 - \tilde{Y}(\beta)}{\beta} = -\tilde{Y}'(0) = \mathbb{E}[Y].$$

Thus we can take the limit $\beta \rightarrow 0$ on the random variables.

Lemma 4.13 *When $\beta \rightarrow 0$, we have that*

$$\lim_{\beta \rightarrow 0} \frac{\bar{Cost}(t, \beta)}{\beta} = \frac{1}{\mu_1 - \lambda_1} \mathbb{E}[r(t + \text{Exp}(\mu_1 - \lambda_1))], \quad (26)$$

$$\lim_{\beta \rightarrow 0} \frac{\Gamma(t, \beta)}{\beta} = \frac{1}{\lambda_1} \mathbb{E}[r(t - I_1)], \quad (27)$$

$$\lim_{\beta \rightarrow 0} \frac{\Phi(t, \alpha^*)}{\beta} = \frac{\lambda_1}{\mu_1 - \lambda_1} \int_{\alpha^*}^t \mathbb{E}[r(s + \text{Exp}(\mu_1 - \lambda_1))] ds, \quad (28)$$

$$\lim_{\beta \rightarrow 0} \frac{1 - \mathbb{E}[\gamma_1^M]}{\beta} = \lambda_1(t - \alpha^*) \frac{1}{\mu_1 - \lambda_1}, \quad (29)$$

$$\lim_{\beta \rightarrow 0} \beta U_\beta(\alpha^* - I_1) = (\mu_1 - \lambda_1) \mathbb{E}[r(\alpha^* - I_1)] + \lambda_1 \mathbb{E}[r(\alpha^* + \text{Exp}(\mu_1 - \lambda_1))]. \quad (30)$$

Proof

$$\begin{aligned} \lim_{\beta \rightarrow 0} \frac{\bar{Cost}(t, \beta)}{\beta} &= \lim_{\beta \rightarrow 0} \frac{1 - \gamma_1}{\beta} \mathbb{E}[r(t + \text{Exp}(\frac{\beta}{1 - \gamma_1}))] \\ &= \mathbb{E}[BP[\lambda_1; S_1]] \cdot \\ &\quad \mathbb{E}[r(t + \text{Exp}(\frac{1}{\mathbb{E}[BP[\lambda_1; S_1]]}))] \\ &\quad (\text{Dominated Convergence Theorem}) \\ &= \frac{1}{\mu_1 - \lambda_1} \mathbb{E}[r(t + \text{Exp}(\mu_1 - \lambda_1))]. \end{aligned} \quad (31)$$

$$\begin{aligned} \lim_{\beta \rightarrow 0} \frac{\Gamma(t, \beta)}{\beta} &= \lim_{\beta \rightarrow 0} \frac{1 - \gamma_2}{\beta} \mathbb{E}[r(t - I_1)] \\ &= \frac{1}{\lambda_1} \mathbb{E}[r(t - I_1)]. \end{aligned} \quad (32)$$

$$\lim_{\beta \rightarrow 0} \frac{\Phi(t, \alpha^*)}{\beta} = \int_{\alpha^*}^t \lim_{\beta \rightarrow 0} e^{-\lambda_1(1 - \gamma_1)(t-s)} \lambda_1 \frac{1}{\beta} \bar{Cost}(s, \beta) ds$$

$$\begin{aligned}
&= \int_{\alpha^*}^t \lim_{\beta \rightarrow 0} \lambda_1 \frac{1}{\beta} \bar{C}ost(s, \beta) ds && \text{because } \gamma_1 \rightarrow 1 \\
&= \frac{\lambda_1}{\mu_1 - \lambda_1} \int_{\alpha^*}^t \mathbb{E}[r(s + \text{Exp}(\mu_1 - \lambda_1))] ds. && (33)
\end{aligned}$$

$$\begin{aligned}
\lim_{\beta \rightarrow 0} \frac{1 - \mathbb{E}[\gamma_1^M]}{\beta} &= \lim_{\beta \rightarrow 0} \frac{1 - e^{-\lambda_1(t - \alpha^*)(1 - \gamma_1)}}{\beta} \\
&= \lambda_1(t - \alpha^*) \lim_{\beta \rightarrow 0} \frac{1 - \gamma_1}{\beta} && \text{L'Hospital's rule} \\
&= \lambda_1(t - \alpha^*) \frac{1}{\mu_1 - \lambda_1}. && (34)
\end{aligned}$$

$$\begin{aligned}
\lim_{\beta \rightarrow 0} \beta U_\beta(\alpha^* - I_1) &= \lim_{\beta \rightarrow 0} \frac{\beta}{1 - \gamma_1 \gamma_2} \left(\frac{1}{\beta} \Gamma(\alpha^*, \beta) + \gamma_2 \frac{1}{\beta} \bar{C}ost(\alpha^*, \beta) \right) \\
&= (\mu_1 - \lambda_1) \lambda_1 \left(\frac{1}{\lambda_1} \mathbb{E}[r(\alpha^* - I_1)] + \frac{1}{\mu_1 - \lambda_1} \mathbb{E}[r(\alpha^* + \text{Exp}(\mu_1 - \lambda_1))] \right) \\
&= (\mu_1 - \lambda_1) \mathbb{E}[r(\alpha^* - I_1)] + \lambda_1 \mathbb{E}[r(\alpha^* + \text{Exp}(\mu_1 - \lambda_1))]. && (35)
\end{aligned}$$

□

4.3.3 Proving claim 1

Finally, we derive the expected terms in Lemma 4.6 by taking the limit $\beta \rightarrow 0$ in the Ω_β terms, which can be characterized using the U_β terms.

Lemma 4.14 *The expected terms in Lemma 4.6 have the expressions given in (36), (37) and (38).*

Proof Define $B_1 := BP[S_1; \lambda; S]$, $B_2 := BP[S_2; \lambda; S]$, $B_W := BP[\widehat{W}(t_1 - \alpha^*); \lambda; S]$.

Since $\Omega_\beta(t, W)$ can be seen as the total cost incurred during the first $BP[W; \lambda; S]$ time, we have that

$$\begin{aligned}
U_\beta(t_1) &= \Omega_\beta(t_1, S_1 + S_2 + \widehat{W}(t_1 - \alpha^*)) + \tilde{B}_1(\beta) \tilde{B}_2(\beta) \tilde{B}_W(\beta) U_\beta(\alpha^* - I_1), \\
U_\beta(t_1) &= \Omega_\beta(t_1, S_1 + \widehat{W}(t_1 - \alpha^*)) + \tilde{B}_1(\beta) \tilde{B}_W(\beta) U_\beta(\alpha^* - I_1), \\
U_\beta(t_1 - I_1) &= \Omega_\beta(t_1 - I_1, S_2 + \widehat{W}(t_1 - \alpha^*)) + \tilde{B}_2(\beta) \tilde{B}_W(\beta) U_\beta(\alpha^* - I_1),
\end{aligned}$$

where $\tilde{B}(\beta) = \mathbb{E}[e^{-\beta B}]$ is the expected discount factor after a busy period B .

Now we take the limit $\beta \rightarrow 0$ and use Lemma 4.11 and Lemma 4.13.

$$\begin{aligned}
& \mathbb{E} \left[\int_0^{\text{BP}[S_1 + S_2 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \middle| A(0) = t_1 - \alpha^* \right] \\
&= \lim_{\beta \rightarrow 0} \Omega_\beta(t_1, S_1 + S_2 + \widehat{W}(t_1 - \alpha^*)) \\
&= \lim_{\beta \rightarrow 0} U_\beta(t_1) - \tilde{B}_1(\beta) \tilde{B}_2(\beta) \tilde{B}_W(\beta) U_\beta(\alpha^* - I_1) \\
&= \lim_{\beta \rightarrow 0} \frac{1}{\beta} \left(\text{Cost}(t_1, \beta) + \gamma_1 \Phi(t_1, \alpha^*) \right) + \left(\mathbb{E}[\gamma_1^{M+1}] - \tilde{B}_1(\beta) \tilde{B}_2(\beta) \tilde{B}_W(\beta) \right) U_\beta(\alpha^* - I_1) \\
&= \frac{1}{\mu_1 - \lambda_1} \mathbb{E}[r(t_1 + \text{Exp}(\mu_1 - \lambda_1))] + \frac{\lambda_1}{\mu_1 - \lambda_1} \int_{\alpha^*}^{t_1} \mathbb{E}[r(s + \text{Exp}(\mu_1 - \lambda_1))] ds \\
&\quad + \lim_{\beta \rightarrow 0} \left(\frac{1 - \tilde{B}_1(\beta) \tilde{B}_2(\beta) \tilde{B}_W(\beta)}{\beta} - \frac{1 - \gamma_1 \mathbb{E}[\gamma_1^M]}{\beta} \right) \beta U_\beta(\alpha^* - I_1) \\
&= \frac{1}{\mu_1 - \lambda_1} \mathbb{E}[r(t_1 + \text{Exp}(\mu_1 - \lambda_1))] + \frac{\lambda_1}{\mu_1 - \lambda_1} \int_{\alpha^*}^{t_1} \mathbb{E}[r(s + \text{Exp}(\mu_1 - \lambda_1))] ds \\
&\quad + \left(\mathbb{E}[B_1 + B_2 + B_W] - \lambda_1(t_1 - \alpha^*) \frac{1}{\mu_1 - \lambda_1} \right. \\
&\quad \left. - \frac{1}{\mu_1 - \lambda_1} \right) \lim_{\beta \rightarrow 0} \beta U_\beta(\alpha^* - I_1). \tag{36}
\end{aligned}$$

Similarly, we have that

$$\begin{aligned}
& \mathbb{E} \left[\int_0^{\text{BP}[S_1 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \middle| A(0) = t_1 - \alpha^* \right] \\
&= \lim_{\beta \rightarrow 0} \Omega_\beta(t_1, S_1 + \widehat{W}(t_1 - \alpha^*)) \\
&= \lim_{\beta \rightarrow 0} U_\beta(t_1) - \tilde{B}_1(\beta) \tilde{B}_W(\beta) U_\beta(\alpha^* - I_1) \\
&= \lim_{\beta \rightarrow 0} \frac{1}{\beta} \left(\text{Cost}(t_1, \beta) + \gamma_1 \Phi(t_1, \alpha^*) \right) + \left(\mathbb{E}[\gamma_1^{M+1}] - \tilde{B}_1(\beta) \tilde{B}_W(\beta) \right) U_\beta(\alpha^* - I_1) \\
&= \frac{1}{\mu_1 - \lambda_1} \mathbb{E}[r(t_1 + \text{Exp}(\mu_1 - \lambda_1))] + \frac{\lambda_1}{\mu_1 - \lambda_1} \int_{\alpha^*}^{t_1} \mathbb{E}[r(s + \text{Exp}(\mu_1 - \lambda_1))] ds \\
&\quad + \lim_{\beta \rightarrow 0} \left(\frac{1 - \tilde{B}_1(\beta) \tilde{B}_W(\beta)}{\beta} - \frac{1 - \gamma_1 \mathbb{E}[\gamma_1^M]}{\beta} \right) \beta U_\beta(\alpha^* - I_1) \\
&= \frac{1}{\mu_1 - \lambda_1} \mathbb{E}[r(t_1 + \text{Exp}(\mu_1 - \lambda_1))] + \frac{\lambda_1}{\mu_1 - \lambda_1} \int_{\alpha^*}^{t_1} \mathbb{E}[r(s + \text{Exp}(\mu_1 - \lambda_1))] ds \\
&\quad + \left(\mathbb{E}[B_1 + B_W] - \lambda_1(t_1 - \alpha^*) \frac{1}{\mu_1 - \lambda_1} - \frac{1}{\mu_1 - \lambda_1} \right) \lim_{\beta \rightarrow 0} \beta U_\beta(\alpha^* - I_1). \tag{37}
\end{aligned}$$

Lemma 4.12

$$\begin{aligned}
& \mathbb{E} \left[\int_0^{\text{BP}[S_2 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* - I_1 \right] \\
&= \lim_{\beta \rightarrow 0} \Omega_\beta(t_1 - I_1, S_2 + \widehat{W}(t_1 - \alpha^*)) \\
&= \lim_{\beta \rightarrow 0} U_\beta(t_1 - I_1) - \tilde{B}_2(\beta) \tilde{B}_W(\beta) U_\beta(\alpha^* - I_1) \\
&= \lim_{\beta \rightarrow 0} \frac{1}{\beta} \Phi(t_1, \alpha^*) + \left(\mathbb{E}[\gamma_1^M] - \tilde{B}_2(\beta) \tilde{B}_W(\beta) \right) U_\beta(\alpha^* - I_1) \\
&= \frac{\lambda_1}{\mu_1 - \lambda_1} \int_{\alpha^*}^{t_1} \mathbb{E}[r(s + \text{Exp}(\mu_1 - \lambda_1))] ds \\
&\quad + \lim_{\beta \rightarrow 0} \left(\frac{1 - \tilde{B}_2(\beta) \tilde{B}_W(\beta)}{\beta} - \frac{1 - \mathbb{E}[\gamma_1^M]}{\beta} \right) \beta U_\beta(\alpha^* - I_1) \\
&= \frac{\lambda_1}{\mu_1 - \lambda_1} \int_{\alpha^*}^{t_1} \mathbb{E}[r(s + \text{Exp}(\mu_1 - \lambda_1))] ds \\
&\quad + \left(\mathbb{E}[B_2 + B_W] - \lambda_1(t_1 - \alpha^*) \frac{1}{\mu_1 - \lambda_1} \right) \lim_{\beta \rightarrow 0} \beta U_\beta(\alpha^* - I_1).
\end{aligned} \tag{38}$$

Lemma 4.12

□

We list here one more lemma in [15] that will be useful.

Lemma 4.15 *(r') The derivative of r is given by*

$$r'(t) = c'(t) + \lambda_1 c(t).$$

Proof See Lemma 4.16 in [15].

□

Now we are ready to prove Claim 1.

Proposition 4.16 *(Claim 1) For all $t_1 \geq \alpha^*, t_2 \geq 0$,*

$$\mu_1(V(t_1, t_2) - V(t_1 - I_1, t_2)) \geq \mu_2(V(t_1, t_2) - V(t_1, t_2 - I_2)).$$

Furthermore, the inequality is an equality when $t_1 = \alpha^$.*

Proof By Lemma 4.6, it suffices to show that

$$\begin{aligned}
& \mu_1 \left(\mathbb{E} \left[\int_0^{\text{BP}[S_1 + S_2 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* \right] \right. \\
& \quad \left. - \mathbb{E} \left[\int_0^{\text{BP}[S_2 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* - I_1 \right] \right) \\
& \geq \mu_2 \left(\mathbb{E} \left[\int_0^{\text{BP}[S_1 + S_2 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* \right] \right)
\end{aligned}$$

$$- \mathbb{E} \left[\int_0^{\text{BP}[S_1 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* \right] \right). \quad (39)$$

By Lemma 4.14, we have that the left hand side is equal to

$$\begin{aligned} & \mu_1 \left(\mathbb{E} \left[\int_0^{\text{BP}[S_1 + S_2 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* \right] \right. \\ & \quad \left. - \mathbb{E} \left[\int_0^{\text{BP}[S_2 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* - I_1 \right] \right) \\ & = \mu_1 \left(\frac{1}{\mu_1 - \lambda_1} \mathbb{E}[r(t_1 + \text{Exp}(\mu_1 - \lambda_1))] + \left(\mathbb{E}[B_1] - \frac{1}{\mu_1 - \lambda_1} \right) \lim_{\beta \rightarrow 0} \beta U_\beta(\alpha^* - I_1) \right). \end{aligned} \quad (40)$$

Also, the right hand side of (39) is equal to:

$$\begin{aligned} & \mu_1 \left(\mathbb{E} \left[\int_0^{\text{BP}[S_1 + S_2 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* \right] \right. \\ & \quad \left. - \mathbb{E} \left[\int_0^{\text{BP}[S_1 + \widehat{W}(t_1 - \alpha^*); \lambda, S]} r(\alpha^* + A(t)) dt \mid A(0) = t_1 - \alpha^* \right] \right) \\ & = \mu_2 \mathbb{E}[B_2] \lim_{\beta \rightarrow 0} \beta U_\beta(\alpha^* - I_1). \end{aligned} \quad (41)$$

Note that $\mathbb{E}[B_1] = \frac{\mathbb{E}[S_1]}{1-\rho} = \frac{1}{\mu_1(1-\rho)}$, $\mathbb{E}[B_2] = \frac{1}{\mu_2(1-\rho)}$. Thus we have that (39) is equivalent to

$$\frac{\mu_1}{\mu_1 - \lambda_1} \mathbb{E}[r(t_1 + \text{Exp}(\mu_1 - \lambda_1))] \geq \frac{1}{\mu_1 - \lambda_1} \lim_{\beta \rightarrow 0} \beta U_\beta(\alpha^* - I_1).$$

Now using Lemma 4.13, the above is equivalent to

$$\mu_1 \mathbb{E}[r(t_1 + \text{Exp}(\mu_1 - \lambda_1))] \geq (\mu_1 - \lambda_1) \mathbb{E}[r(\alpha^* - I_1)] + \lambda_1 \mathbb{E}[r(\alpha^* + \text{Exp}(\mu_1 - \lambda_1))]. \quad (42)$$

Using Lemma 3.6, we have that

$$\begin{aligned} & \mathbb{E}[r(t_1 + \text{Exp}(\mu_1 - \lambda_1))] \\ & = r(\alpha^*) + \frac{1}{\mu_1 - \lambda_1} \mathbb{E}[r'(t_1 + \text{Exp}(\mu_1 - \lambda_1))], \\ & = r(t_1) + \frac{1}{\mu_1 - \lambda_1} (\mathbb{E}[c'(t_1 + \text{Exp}(\mu_1 - \lambda_1))] + \lambda_1 \mathbb{E}[c(t_1 + \text{Exp}(\mu_1 - \lambda_1))]) \end{aligned}$$

Lemma 4.15

$$= r(t_1) + \mathbb{E}[c(t_1 + \text{Exp}(\mu_1 - \lambda_1))] - c(t_1) + \frac{1}{\mu_1 - \lambda_1} \lambda_1 \mathbb{E}[c(t_1 + \text{Exp}(\mu_1 - \lambda_1))]$$

$$= r(t_1) + \frac{\mu_1}{\mu_1 - \lambda_1} \mathbb{E}[c(t_1 + \text{Exp}(\mu_1 - \lambda_1))] - c(t_1) \quad (43)$$

$$\begin{aligned} \mathbb{E}[r(t_1 - I_1)] &= r(t_1) - \frac{1}{\lambda_1} \mathbb{E}[r'(t_1 - I_1)] \\ &= r(t_1) - \frac{1}{\lambda_1} (\mathbb{E}[c'(t_1 - I_1)] + \lambda_1 \mathbb{E}[c(t_1 - I_1)]) \quad \text{Lemma 4.15} \\ &= r(t_1) - (c(t_1) - \mathbb{E}[c(t_1 - I_1)]) - \mathbb{E}[c(t_1 - I_1)] \quad \text{Lemma 3.6} \\ &= r(t_1) - c(t_1). \end{aligned} \quad (44)$$

Note that by (4), we have that $\mathbb{E}[c(\alpha^* + \text{Exp}(\mu_1 - \lambda_1))] = 0$. Thus we have that

$$\mathbb{E}[r(\alpha^* - I_1)] = \mathbb{E}[r(\alpha^* + \text{Exp}(\mu_1 - \lambda_1))].$$

Therefore, (42) is equivalent to

$$\mathbb{E}[r(t_1 + \text{Exp}(\mu_1 - \lambda_1))] \geq \mathbb{E}[r(\alpha^* + \text{Exp}(\mu_1 - \lambda_1))].$$

This inequality holds because for any $s \geq \alpha^*$,

$$\begin{aligned} &\frac{d}{ds} \mathbb{E}[r(s + \text{Exp}(\mu_1 - \lambda_1))] \\ &= \mathbb{E}[r'(s + \text{Exp}(\mu_1 - \lambda_1))] \quad \text{(Dominated Convergence Theorem)} \\ &= \mathbb{E}[c'(s + \text{Exp}(\mu_1 - \lambda_1))] + \lambda_1 \mathbb{E}[c(s + \text{Exp}(\mu_1 - \lambda_1))] \quad \text{(Lemma 4.15)} \\ &\geq 0 + \lambda_1 \mathbb{E}[c(\alpha^* + \text{Exp}(\mu_1 - \lambda_1))] \quad (c \text{ is non-decreasing}) \\ &= 0. \end{aligned}$$

□

4.4 Verifying claim 2

Finally, we use induction to prove Claim 2.

Proposition 4.17 (Claim 2) For all $t_1 \in [0, \alpha^*]$, $t_2 \geq 0$,

$$\mu_1(V(t_1, t_2) - V(t_1 - I_1, t_2)) \leq \mu_2(V(t_1, t_2) - V(t_1, t_2 - I_2)).$$

Proof Define function $g(t_1, t_2) := \mu_1(V(t_1, t_2) - V(t_1 - I_1, t_2)) - \mu_2(V(t_1, t_2) - V(t_1, t_2 - I_2))$. We next show that $g(t_1, t_2) \leq 0$ by induction on t_1 .

We know equality holds at $t_1 = \alpha^*$ by Claim 1 (Proposition 4.16). Now assume that for some $t_1 \leq \alpha^*$, we have that for all $t_2 \geq 0$, $g(t_1, t_2) \leq 0$. We are going to show that there exists a $\Delta > 0$ which does not depend on t_1 , such that for any t_1 and $\delta \in (0, \Delta)$, we have that for all $t_2 \geq 0$,

$$g(t_1 - \delta, t_2) \leq 0.$$

We unfold the states from $(t_1 - \delta, t_2)$ by a δ time-step, and use the inductive hypothesis.

$$\begin{aligned} V(t_1 - \delta, t_2) &= r(t_1 - \delta)\delta + (1 - \mu_2\delta)V(t_1, t_2 + \delta) \\ &\quad + \mu_2\delta V(t_1, t_2 + \delta - I_2) + o(\delta), \end{aligned} \quad (45)$$

$$\begin{aligned} &= r(t_1 - \delta)\delta + V(t_1, t_2 + \delta) - \mu_2\delta(V(t_1, t_2 + \delta) \\ &\quad - V(t_1, t_2 + \delta - I_2)) + o(\delta), \end{aligned} \quad (46)$$

$$\begin{aligned} V(t_1 - \delta - I_1, t_2) &= r(t_1 - \delta - I_1)\delta + (1 - \mu_2\delta)V(t_1 - I_1, t_2 + \delta) \\ &\quad + \mu_2\delta V(t_1 - I_1, t_2 + \delta - I_2) + o(\delta) \end{aligned} \quad (47)$$

$$\begin{aligned} V(t_1 - \delta, t_2 - I_2) &= e^{-\lambda_2 t_2} V(t_1 - \delta, -I_2) + (1 - e^{-\lambda_2 t_2})V(t_1 - \delta, t_2 - \bar{I}_2) + o(\delta), \\ &= r(t_1 - \delta)\delta + e^{-\lambda_2 t_2} ((1 - \mu_1\delta)V(t_1, \delta - I_2) + \mu_1\delta V(t_1 - I_1, \delta - I_2)) \\ &\quad + (1 - e^{-\lambda_2 t_2}) \left((1 - \mu_2\delta)V(t_1, t_2 - \bar{I}_2 + \delta) \right. \\ &\quad \left. + \mu_2\delta V(t_1, t_2 - \bar{I}_2 + \delta - I_2) \right) + o(\delta), \\ &= r(t_1 - \delta)\delta + V(t_1, t_2 + \delta - I_2) \\ &\quad - e^{-\lambda_2 t_2} \cdot \mu_1\delta (V(t_1, \delta - I_2) - V(t_1 - I_1, \delta - I_2)) \\ &\quad - (1 - e^{-\lambda_2 t_2}) \cdot \mu_2\delta (V(t_1, t_2 + \delta - \bar{I}_2) - V(t_1, t_2 + \delta - \bar{I}_2 - I_2)) \\ &\quad + o(\delta), \\ &\leq r(t_1 - \delta)\delta + V(t_1, t_2 + \delta - I_2) \\ &\quad - e^{-\lambda_2 t_2} \cdot \mu_1\delta (V(t_1, \delta - I_2) - V(t_1 - I_1, \delta - I_2)) \\ &\quad - (1 - e^{-\lambda_2 t_2}) \cdot \mu_1\delta (V(t_1, t_2 + \delta - \bar{I}_2) - V(t_1 - I_1, t_2 + \delta - \bar{I}_2)) \\ &\quad + o(\delta), \text{ (by induction hypothesis)} \\ &= r(t_1 - \delta)\delta + V(t_1, t_2 + \delta - I_2) \\ &\quad - \mu_1\delta (V(t_1, t_2 + \delta - I_2) - V(t_1 - I_1, t_2 + \delta - I_2)) + o(\delta). \end{aligned} \quad (48)$$

where $\bar{I}_2 \sim [I_2 | I_2 < t_2]$. Therefore, from Equations (45) and (47), we have that

$$\begin{aligned} \mu_1(V(t_1 - \delta, t_2) - V(t_1 - \delta - I_1, t_2)) &= \mu_1 c(t_1 - \delta)\delta + \mu_1(1 - \mu_2\delta)(V(t_1, t_2 + \delta) \\ &\quad - V(t_1 - I_1, t_2 + \delta)) + \mu_1\mu_2\delta (V(t_1, t_2 + \delta - I_2) - V(t_1 - I_1, t_2 + \delta - I_2)) \\ &\quad + o(\delta), \end{aligned} \quad (49)$$

And from Equations (46) and (48), we have that

$$\begin{aligned} \mu_2(V(t_1 - \delta, t_2) - V(t_1 - \delta, t_2 - I_2)) &\geq \mu_2(V(t_1, t_2 + \delta) - V(t_1, t_2 + \delta - I_2)) - \mu_2^2\delta(V(t_1, t_2 + \delta) - V(t_1, t_2 + \delta - I_2)) \\ &\quad + \mu_1\mu_2\delta (V(t_1, t_2 + \delta - I_2) - V(t_1 - I_1, t_2 + \delta - I_2)) + o(\delta) \\ &= \mu_2(1 - \mu_2\delta)(V(t_1, t_2 + \delta) - V(t_1, t_2 + \delta - I_2)) \end{aligned} \quad (50)$$

$$+ \mu_1 \mu_2 \delta (V(t_1, t_2 + \delta - I_2) - V(t_1 - I_1, t_2 + \delta - I_2))) + o(\delta). \quad (51)$$

Combining Equations (49) and (51), we have that

$$g(t_1 - \delta, t_2) \leq \mu_1 c(t_1 - \delta) \delta + (1 - \mu_2 \delta) g(t_1, t_2) + o(\delta).$$

Since we have that $\mathbb{E}[c(\alpha^* + \text{Exp}(\mu_1 - \lambda_1))] = 0$ and that c is non-decreasing, we have that $c(t_1 - \delta) \leq c(\alpha^*) < 0$.⁴ Thus we have that

$$\begin{aligned} g(t_1 - \delta, t_2) &\leq \mu_1 c(\alpha^*) \delta + (1 - \mu_2 \delta) g(t_1, t_2) + o(\delta) \\ &\leq \mu_1 c(\alpha^*) \delta + o(\delta). \end{aligned} \quad (\text{by induction hypothesis})$$

Observe that the right hand side of the inequality is $o(\delta)$ plus a linear function with respect to δ , where the coefficient of the linear function is a negative constant independent of t_1 . Therefore, we can pick Δ small enough (and independent of t_1) such that for any $\delta \in (0, \Delta)$, the right hand side is smaller than 0. \square

5 Simulations

We now conduct simulations to evaluate the performance of our LookAhead policy from Theorem 1.1. We experiment with different holding cost functions, system loads and arrival rates.

We present our results in the form of 2 experiments. The experiment correspond to a different set of holding cost, arrival and service rate parameters for the two job classes. In each setting, we compare our LookAhead policy with a number of heuristic policies which have been proposed for TVHC problems. We draw this comparison across system loads, while maintaining a fixed ratio of arrival rates from each class.

Via simulation, we demonstrate the following main findings:

- Our policy is not just provably optimal, it also achieves significantly (41-56% in Fig. 8) lower time-average holding cost than other policies.
- Depending on the cost functions, our policy may achieve arbitrarily lower holding cost than other heuristics.

5.1 Policies evaluated

Throughout this section, when we talk about “our policy,” we refer to our LookAhead policy from Theorem 1.1, where the priority of a class i job of age t is given by $V_i(t)$:

$$V_1(t) = \mu_1 \mathbb{E}[c_1(t + X)], \quad \text{where } X \sim \text{Exp}(\mu_1 - \lambda_1), \quad \text{and } V_2(t) = \mu_2 c_2.$$

⁴ Here we can assume $c(\alpha^*)$ is strictly negative. Otherwise we have that for any $t > \alpha^*$, $c(t) = 0$. This falls back to the case when P-Prio(2; 1) is optimal (similar to the arguments at the beginning of Sect. 4).

We compare our policy against the following alternatives:

- **FCFS:** This policy always serves the job that arrived earliest. FCFS is a very simple policy and we compare against it as a baseline.
- **Strict Priority:** This policy assigns a fixed priority to each job class, where jobs from a higher-priority class have preemptive priority over those from a lower-priority class. Jobs within a class are run in FCFS order. In our plots, we present the point-wise better of P-Prio(1; 2) and P-Prio(2; 1), where P-Prio(1; 2) denotes a preemptive priority policy where class 1 has strict priority over class 2.
- **Generalized $c\mu$ Rule [5]:** This policy always serves the job with highest index $c_i(t) \cdot \mu_i$, where t is the age of the class i job. Namely,

$$V_1(t) = \mu_1 c_1(t), \quad \text{and } V_2(t) = \mu_2 c_2.$$

This policy is known to be optimal in the diffusion limit.

- **Aalto's Whittle Index Policy [13]:** In this policy a class i job of age t is given index $V_i(t)$, where

$$V_1(t) = \mu_1 \mathbb{E}[c_1(t + S_1)], \quad \text{where } S_1 \sim \text{Exp}(\mu_1), \quad \text{and } V_2(t) = \mu_2 c_2.$$

This is another Whittle-based heuristic proposed for TVHC in the literature. It also incorporates a LookAhead-like intuition; however, it looks ahead by a shorter time period.

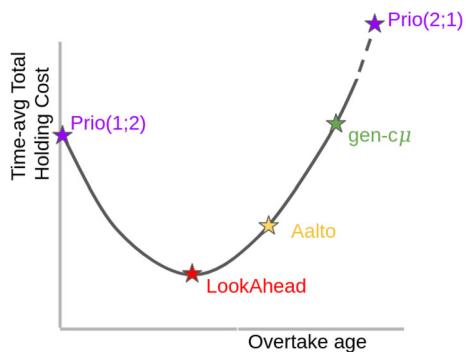
Note that all the above policies, except FCFS belong to the class of Overtake policies. The strict priority policies, P-Prio(1 : 2) and P-Prio(2 : 1) can be represented as Overtake(0) and Overtake(∞), respectively. Generalized $c\mu$, Aalto and our policy each have positive finite overtaking times, with Generalized $c\mu$ having the highest overtaking time and our policy having the lowest. Only our policy's overtaking time varies with load. Our overtaking amount decreases with increasing load, which makes sense since when load increases, we want to overtake sooner so we don't end up with very high holding costs. In the limit of light load, our policy equals Aalto's policy.

Note also that time-average total holding cost is convex in overtaking age, as shown in Fig. 6 and proved in Theorem 3.8. Our policy is at the minimum of this convex function. Aalto, generalized $c\mu$ and P-Prio(2; 1) are to its right; thus, their costs are always ordered from lowest to highest. P-Prio(1; 2), being on the left of the minimum, may perform better or worse than the policies on the right.

5.2 Experimental results

The two experiments are chosen to highlight settings where different policies are superior. In both settings, our policy performs significantly better than the other heuristics. Each experiment shown is represented by (a) a set of holding cost functions (not drawn to scale); (b) the effective overtaking time corresponding to each of the candidate policies (except FCFS which is not an Overtake policy); and (c) the time-average total holding cost obtained by all policies in simulation.

Fig. 6 Expected cost is convex in overtaking age. The minimum is achieved by LookAhead



In Fig. 7, we show an experiment with a quadratic holding cost function for class 1. As shown in Fig. 7a, class 1 holding cost is initially 0 but grows quadratically to surpass class 2's holding cost. When we move to Fig. 7c, we see that FCFS always has a very high holding cost. Strict priority (Prio) is initially the worst policy, but it becomes optimal at high load, when it is P-Prio(1; 2). The generalized $c\mu$ rule and Aalto's policy both start out behaving well at lower loads, but are over 20% worse than our policy at high loads. Finally, in Fig. 7d, we validate the statistical significance of the observed performance differences under high load. To account for the strong correlation in cost variation across policies (due to shared sample paths), we plot the time-average total holding cost ratio of each policy relative to ours with 2σ error bars computed across 10 distinct sample paths. As depicted, our policy obtains consistently lower time-average total holding cost across sample paths.

We can understand the above results by looking at the *overtake age* under each of the above policies, as shown in Fig. 7b. Recall from Fig. 6 that the time-average holding cost is convex in the overtaking age. Recall also that Aalto and the generalized $c\mu$ rule are to the right of our policy, which achieves the minimum cost. Thus it makes sense that Aalto always performs better than the generalized $c\mu$ rule in Figs. 7c and 7d.

Further, for quadratic holding cost functions, we can analytically compute the LookAhead index functions, and thus the optimal overtaking ages as well. The class 1 index function is quadratic in $\mathbb{E}[X] = \frac{1}{\mu_1 - \lambda_1}$. Meanwhile the class 2 index function is constant. Thus the optimal overtaking age, which is the crossing point of the 2 functions, can be derived as the root of a quadratic function. For the particular parameters in Fig. 7a, we have

$$\alpha^*(\rho) = \frac{-1}{1 - 0.9\rho} + \sqrt{90 - \frac{1}{(1 - 0.9\rho)^2}}$$

This explains the rapid decrease in the optimal overtaking age. For $\rho \geq 0.95$, the root of this polynomial is negative, but overtaking age is constrained to be non-negative. Thus optimal overtaking age is zero for $\rho \geq 0.95$.

In Fig. 8, we show an experiment with a deadline-based holding cost function for class 1. As shown in Fig. 8a, a class 1 job at age t incurs instantaneous holding cost c_1

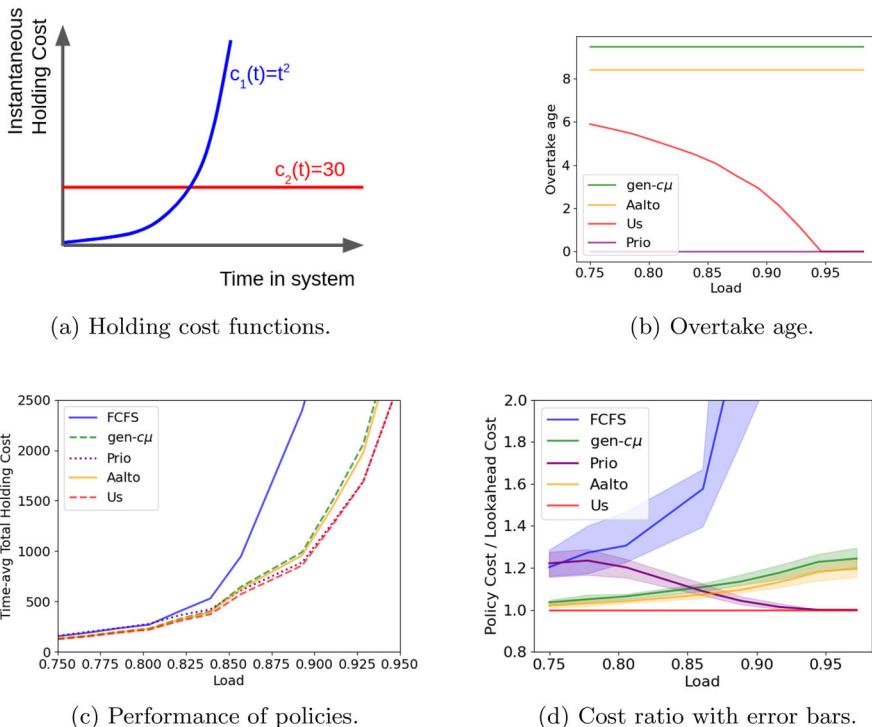


Fig. 7 Comparison of policies on a setting with quadratic holding costs. We fix $\mu_1 = 1$, $\mu_2 = 3$, $\lambda_1 = 0.75\lambda$

past its deadline d_1 . The corresponding time-average total holding costs obtained by the policies is shown in Fig. 8c. Here as well, FCFS always accrues highest holding cost. Strict priority is initially worse than generalized $c\mu$ and Aalto, but is slightly better than Aalto at high load. Our policy does significantly better than all other heuristics. At load 0.9, it is 56% better than Aalto, which is the best of all other heuristics. At load 0.95, it is 41% better than strict priority, which is the best of all other heuristics. We plot performance with error bars in Fig. 8d to test the statistical significance of our policy's performance. Here again, we plot the cost ratio of each policy relative to ours with 2σ error bars computed across 10 distinct sample paths. As depicted, our policy obtains consistently lower time-average total holding cost across sample paths. As depicted, the cost ratio of generalized $c\mu$ and Aalto's heuristic are higher at intermediate loads. They become closer to optimal as we approach heavy traffic in this example.

To understand what's happening, we examine the corresponding overtake ages shown in Fig. 8b. The optimal overtake age decreases with increasing load. However in this setting, it decreases slowly. At load 0, the optimal overtake age is 8.87 (the same as Aalto's overtake age). At load 0.98 (which is the highest load we simulate), the optimal overtake time is 5.71. In fact, in this setting with deadline-based holding costs as well, we can analytically express our LookAhead index functions. This helps

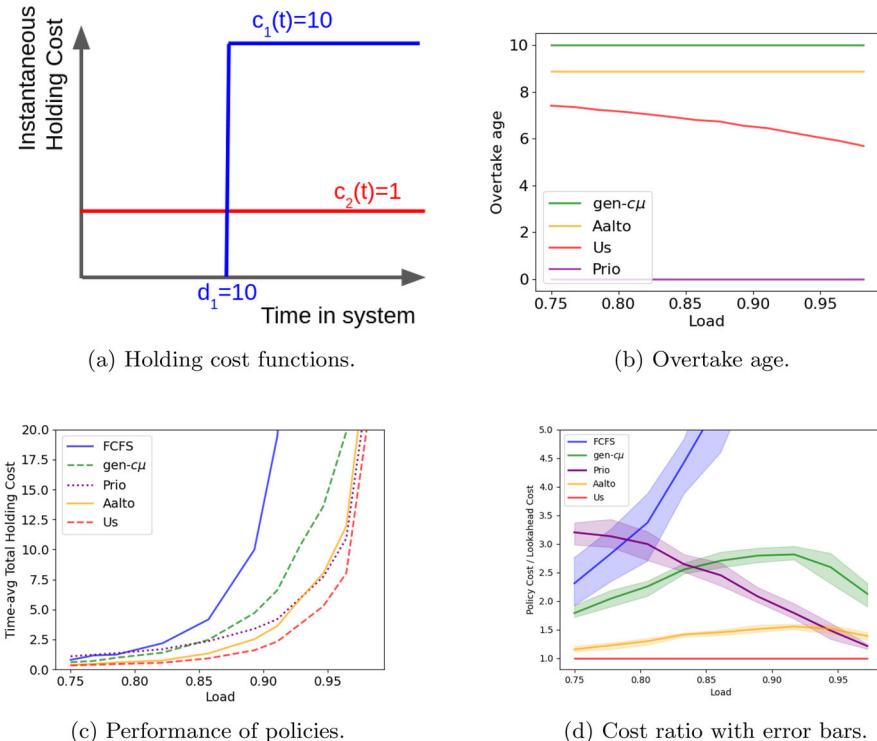


Fig. 8 Comparison of policies on holding cost functions with one deadline. We fix $\mu_1 = 3$, $\mu_2 = 1$, $\lambda_1 = 0.9\lambda$

us derive the optimal overtake age as a function of load to be

$$\alpha^*(\rho) = d_1 - \ln\left(\frac{\mu_1 c_1}{\mu_2 c_2}\right) \cdot \frac{1}{\mu_1 - \lambda_1} = 10 - \frac{\ln(30)}{3 - \frac{9}{4}\rho}.$$

Note that $\lim_{\rho \rightarrow 1} \alpha^*(\rho) = 10 - \frac{4}{3} \ln(30) \approx 5.465$, not zero. Namely, the LookAhead amount and optimal overtake parameter α^* both remain finite in the limit of heavy traffic. Thus none of the other candidate policies (generalized $c\mu$, Aalto's heuristic and strict priority) is optimal in the limit of heavy traffic. This example points out that generalized $c\mu$ is only optimal when we assume *both* the diffusion limit and the limit of heavy traffic. It is not optimal simply under heavy traffic.

6 Conclusion

This paper derives the first optimal scheduling policy for a TVHC problem with two classes of jobs, where one class has a holding cost that increases as jobs age. The policy derived, called LookAhead, has a similar form to the generalized $c\mu$ rule but incorporates the holding cost of a job at a future time X : The index functions

(for class 1 and 2) are, respectively, $V_1(t) = \mu_1 \mathbb{E}[c_1(t + X)]$, and $V_2(t) = \mu_2 c_2$, where $X \sim \text{Exp}(\mu_1 - \lambda_1)$. Our policy is not only optimal but also shows non-trivial improvement over existing policies.

This work opens up promising avenues for future work. First, the TVHC problem is an important but still widely open problem. While our work provides an optimality proof for one simple case of TVHC, we hope that our amortized holding cost intuition from Sect. 3.5 can be useful in more general TVHC instances. Second, there is a large community that works on multi-armed bandit problems. Our work can be viewed as providing an optimality result within the very challenging class of *restless* multi-armed bandit problems. This work can hopefully spur on the discovery of more optimality results within the class of restless multi-armed bandit problems.

Acknowledgements This work was supported by NSF-CIF-2403194, NSF-III-2322973, and NSF-CMMI-2307008.

Funding Open Access funding provided by Carnegie Mellon University. This work is supported by NSF-CIF-2403194, NSF-III-2322973, and NSF-CMMI-2307008.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Cox, D.R.: *Queues*. Chapman and Hall/CRC, Boca Raton (2020)
2. Buyukkoc, C., Varaiya, P., Walrand, J.: The c-mu rule revisited. *Adv. Appl. Probab.* **17**(1), 237–238 (1985). <https://doi.org/10.2307/1427064>
3. Gittins, J.: Dynamic allocation indices for Bayesian bandits. In: *Mathematical Learning Models—Theory and Algorithms: Proceedings of a Conference*, pp. 50–67 (1983)
4. Scully, Z., Harchol-Balter, M.: The Gittins policy in the M/G/1 queue. In: *2021 19th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pp. 1–8 (2021)
5. Van Mieghem, J.A.: Dynamic scheduling with convex delay costs: The generalized c-mu rule. *Annal Appl Probab* 809–833 (1995)
6. Scully, Z., Harchol-Balter, M., Scheller-Wolf, A.: SOAP: one clean analysis of all age-based scheduling policies. *Proc. ACM Measur. Anal. Comput. Syst.* **2**(1), 1–30 (2018)
7. Nuyens, M., Wierman, A.: The foreground-background queue: a survey. *Perform. Eval.* **65**(3–4), 286–307 (2008)
8. Guha, S., Munagala, K., Shi, P.: Approximation algorithms for restless bandit problems. *J. ACM* **58**(1), 1–50 (2010)
9. Nuyens, M., Wierman, A., Zwart, B.: Preventing large sojourn times using smart scheduling. *Oper. Res.* **56**(1), 88–101 (2008)
10. Harchol-Balter, M.: *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge University Press, Cambridge (2013)
11. Rai, I.A., Urvoy-Keller, G., Vernon, M.K., Biersack, E.W.: Performance analysis of LAS-based scheduling disciplines in a packet switched network. *ACM SIGMETRICS Perform. Eval. Rev.* **32**(1), 106–117 (2004)

12. Wierman, A., Harchol-Balter, M., Osogami, T.: Nearly insensitive bounds on SMART scheduling. In: ACM Sigmetrics 2005 Conference on Measurement and Modeling of Computer Systems, pp. 205–215 (2005)
13. Aalto, S.: Whittle index approach to the multi-class queueing systems with convex holding costs and IHR service times. *Math. Methods Oper. Res.* **100**(3), 603–634 (2024)
14. Anand, A., Veciana, G.: A Whittle's index based approach for QoE optimization in wireless networks. *Proc. ACM Measure. Anal. Comput. Syst.* **2**(1), 1–39 (2018)
15. Li, Z., Gurushankar, K., Harchol-Balter, M., Scheller-Wolf, A.: Improving upon the generalized c-mu rule: a Whittle approach. *arXiv preprint arXiv:2504.10622* (2025)
16. Weber, R.R., Weiss, G.: On an index policy for restless bandits. *J. Appl. Probab.* **27**(3), 637–648 (1990)
17. Verloop, I.M.: Asymptotically optimal priority policies for indexable and nonindexable restless bandits. *Ann. Appl. Probab.* **26**(4), 1947–1995 (2016)
18. Gast, N., Gaujal, B., Yan, C.: Exponential asymptotic optimality of whittle index policy. *Queueing Syst.* **104**(1), 107–150 (2023)
19. Stanford, D.A., Taylor, P., Ziedins, I.: Waiting time distributions in the accumulating priority queue. *Queueing Syst.* **77**, 297–330 (2014)
20. Fajardo, V.A., Drekcic, S.: Waiting time distributions in the preemptive accumulating priority queue. *Methodol. Comput. Appl. Probab.* **19**, 255–284 (2017)
21. Mojalal, M., Stanford, D.A., Caron, R.J.: The lower-class waiting time distribution in the delayed accumulating priority queue. *INFOR Inf. Syst. Operational Res.* **58**(1), 60–86 (2020)
22. Cildoz, M., Ibarra, A., Mallor, F.: Accumulating priority queues versus pure priority queues for managing patients in emergency departments. *Operations Res. Health Care* **23**, 100224 (2019)
23. Ansell, P., Glazebrook, K.D., Niño-Mora, J., O'Keeffe, M.: Whittle's index policy for a multi-class queueing system with convex holding costs. *Math. Methods Oper. Res.* **57**, 21–39 (2003)
24. Niño-Mora, J.: Markovian restless bandits and index policies: a review. *Mathematics* **11**(7), 1639 (2023)
25. Kleinrock, L.: A conservation law for a wide class of queueing disciplines. *Naval Res. Logist. Q.* **12**(2), 181–192 (1965)
26. Baird, L.C.: Reinforcement learning in continuous time: advantage updating. In: Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94) 4, pp. 2448–24534 (1994)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.