LEARNING WHILE SCHEDULING JOBS UNDER CONTEXT-DEPENDENT SERVICE RATES: AN ANYTIME RATE-OPTIMAL ALGORITHM

Anonymous authors
Paper under double-blind review

ABSTRACT

We study contextual queueing bandits, where a learner schedules jobs while learning unknown service rates modeled by logistic functions of job-server features. Performance is measured by queue length regret, the expected excess queue length at round t relative to an oracle that knows the service rates. Existing decaying-regret guarantees either have a suboptimal decay rate or require a known fixed horizon. They also assume context-wise slack and a strictly positive minimum eigenvalue of the feature covariance. In this paper, we propose WISE (Widest Interval Selection with Elimination), achieving rate-optimal $\widetilde{\mathcal{O}}(t^{-1/2})$ queue length regret at every sufficiently large time without knowing the horizon. We assume capacity slack, meaning that expected incoming workload under best-server service is below service capacity, and impose no covariance lower bound. Our analysis uses a workload potential measuring the expected service attempts needed by waiting jobs on their best servers. Its drift on nonempty rounds combines a negative term ensured by capacity slack with errors from suboptimal service choices. Then an elliptical potential count bounds how often WISE selects wide confidence intervals, thereby limiting the number of rounds with large service errors. We also sharpen the arrival-rate dependence of an existing lower bound and make its dependence on feature dimension and server count explicit. We prove another lower bound that quantifies the increase in regret as the normalized capacity slack decreases; to our knowledge, this is the first such lower bound for CQB. Simulations show small regret even when context-wise slack fails.

1 INTRODUCTION

Queueing systems play an important role in cloud computing (Vilaplana et al., 2014), call centers (Koole & Mandelbaum, 2002), and services using large language models (LLMs) (Lee et al., 2024a; Mitzenmacher & Shahout, 2025). Scheduling in these systems requires accounting for two characteristics. First, service rates are often difficult to predict a priori and estimated from observed service outcomes. Second, jobs have different contexts, and the service rate of a job depends on both the job itself and the server assigned to it. For example, an LLM's probability of successfully answering a query depends on the query's specifications and the model's capabilities.

Queueing bandits provide a framework for learning unknown service rates while scheduling jobs (Krishnasamy et al., 2021; Choudhury et al., 2021; Stahlbuhk et al., 2021). Scheduling decisions determine both which jobs are served and which service outcomes are observed to estimate these rates. Performance is measured by *queue length regret*, the expected excess queue length relative to an oracle that knows the service rates. A (queue length) regret bound that converges to zero implies that the learner's expected queue length approaches that of the oracle.

Contextual queueing bandits (CQB) extend this framework to incorporate job-specific contexts (Bae et al., 2026a). Earlier work on queueing bandits considers context-independent service rates (Krishnasamy et al., 2021) or a fixed set of queue features (Kim & Oh, 2024). In contrast, CQB allows each job to have its own feature vector. The departure rate is assumed to follow a logistic model of job-server features (Bae et al., 2026a), as in logistic bandits (Filippi et al., 2010; Faury et al., 2020; Jun et al., 2021). This allows scheduling decisions to reflect heterogeneous jobs and servers. Bae

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

et al. (2026a) establish the first decaying queue length regret bound for CQB, of order $\widetilde{\mathcal{O}}(t^{-1/4})$. Bae et al. (2026b) obtain an anytime regret bound of the same order. Bae & Lee (2026a) improve the rate to $\widetilde{\mathcal{O}}(t^{-1/2})$, which is optimal up to logarithmic factors, for a fixed horizon t known in advance.

However, the earlier results on CQB have the following three limitations. First, they either have a suboptimal rate (Bae et al., 2026a;b) or require advance knowledge of the horizon (Bae et al., 2026a; Bae & Lee, 2026a). Cloud and LLM services operate over an indefinite horizon and must maintain stable queues throughout operation, motivating an anytime horizon-free algorithm. Second, all three results require a context-wise slack condition (Bae et al., 2026a;b; Bae & Lee, 2026a): the best service rate exceeds the arrival rate by a uniform positive gap for every context, regardless of how frequently it arrives. A natural relaxation is to require *capacity slack*: the expected incoming workload under best-server service is below the system's service capacity. Third, the results also require that the minimum eigenvalue of the feature covariance is strictly positive. This condition fails when similar jobs have features concentrated in a lower-dimensional subspace.

Motivated by these limitations, we develop an anytime horizon-free algorithm with a rate-optimal regret bound under minimal assumptions. Instead of a context-wise slack condition, we assume capacity slack stated in Assumption 2. Moreover, the result does not require a positive lower bound on the minimum eigenvalue of the feature covariance. Our contributions are summarized as follows.

- We propose WISE (Widest Interval Selection with Elimination), which eliminates job-server pairs that cannot be optimal and selects the widest remaining confidence interval. It achieves anytime horizon-free queue length regret of order $\widetilde{\mathcal{O}}(t^{-1/2})$ without knowing the normalized capacity slack ϵ. The guarantee holds under capacity slack without a covariance lower bound, allowing both difficult job contexts and features confined to a lower-dimensional subspace.
- We refine the $\Omega(\lambda t^{-1/2})$ lower bound of Bae & Lee (2026a) by proving a round of order $\Omega(\min\{\lambda, \min\{d, \log K\}\sqrt{\lambda/t}\})$. This sharpens the dependence on the arrival rate λ and makes the dependence on the feature dimension d and the number of servers K explicit.
- We prove another lower bound of order $\Omega(\min\{\epsilon^{-1}, \min\{d, \log K\}/(\epsilon^2\sqrt{t})\})$. To our knowledge, this is the first lower bound for CQB that captures the increase in regret as the normalized capacity slack ϵ decreases. This quantifies how service errors persisting for $\Theta(\epsilon^{-2})$ rounds increase the difficulty of learning while scheduling.
- We compare WISE with seven baselines in simulations with capacity slack, including settings that violate context-wise slack. Varying slow-job service probabilities and the normalized capacity slack shows how these conditions affect queue length regret. WISE attains small regret even when some contexts violate context-wise slack.

Our analysis for WISE uses two main ideas. First, we replace queue length with a workload potential based on the expected service attempts needed by waiting jobs on their best servers (Gamarnik & Stolyar, 2012). On nonempty rounds, capacity slack gives a negative workload drift term plus an error term from suboptimal service choices, even when queue length itself has positive drift. Second, under valid confidence bounds, WISE's elimination and widest interval selection link large departure rate gaps to wide selected intervals. The elliptical potential count (Abbasi-Yadkori et al., 2011; Kim et al., 2022) limits such selections without a covariance lower bound. Combined with workload drift, this makes long periods without an empty queue unlikely. When a new job arrives after an empty round, WISE serves it using the most uncertain candidate server. We use these observations to bound uncertainty under the arrival distribution, then combine recent empty rounds with confidence interval intersections to control the expected departure rate gap at each sufficiently large time.

2 Problem Setting

2.1 Queueing System and Queue Length Regret

We consider a queueing system with contextual jobs, a single queue and K servers indexed by $[K] = \{1, \dots, K\}$. Let $\mathcal{X}_t$ be the multiset of waiting job contexts in $\mathcal{X}$ at the beginning of round t, with $Q(t) = |\mathcal{X}_t|$ and $\mathcal{X}_1 = \varnothing$. In a nonempty round, the agent selects a job $x_t \in \mathcal{X}_t$ and a server $k_t \in [K]$. The service probability is $p(x, k) = \mu(\langle \boldsymbol{\phi}(x, k), \boldsymbol{\theta}^* \rangle)$, where $\mu(z) = (1 + e^{-z})^{-1}$,

$\boldsymbol{\phi} : \mathcal{X} \times [K] \to \mathbb{R}^d$ is a known measurable feature map, and $\boldsymbol{\theta}^* \in \mathbb{R}^d$ is unknown. Let $\mathcal{F}_t$ contain the history before the service outcome, including the chosen pair and the agent's randomization. The departure indicator satisfies $D(t) \mid \mathcal{F}_t \sim \mathrm{Bern}(p(x_t, k_t))$ when $Q(t) > 0$, and $D(t) = 0$ otherwise.

After service, a job arrives with indicator $A(t) \sim \mathrm{Bern}(\lambda)$, $\lambda \in (0, 1)$, and context $X_t \sim \mathcal{D}$. The arrival pairs are independent across rounds, $A(t)$ is independent of X_t, and $(A(t), X_t)$ is independent of $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(D(t))$. The context is observed when $A(t) = 1$. The queue evolves as

$$\mathcal{X}_{t+1} = \mathcal{X}_t \setminus \{x_t : D(t) = 1\} \cup \{X_t : A(t) = 1\}, \qquad Q(t+1) = Q(t) + A(t) - D(t).$$

Our goal is to minimize queue length regret against an oracle that knows the service probabilities and chooses a job-server pair with maximum service probability $p(x, k)$ in its own queue (Krishnasamy et al., 2021; Bae et al., 2026a). Write $Q^\pi(t) = |\mathcal{X}_t^\pi|$ for the queue length under policy π, and $Q^*(t)$ for the oracle queue length, with arrivals shared across policies. The queue length regret is $R_t^\pi = \mathbb{E}[Q^\pi(t) - Q^*(t)]$. We omit the learner's policy superscript and write $\mathcal{X}_t$, $Q(t)$, and R_t. We seek a single policy whose regret vanishes at every sufficiently large time t, without being given t. For the learner's queue, define $(x_i^*, k_i^*) \in \arg\max_{x \in \mathcal{X}_i, k \in [K]} p(x, k)$ and write $\boldsymbol{\phi}_i = \boldsymbol{\phi}(x_i, k_i)$, $\boldsymbol{\phi}_i^* = \boldsymbol{\phi}(x_i^*, k_i^*)$, setting both features to zero when $Q(i) = 0$. This best pair may differ from the oracle's choice because the two queues may contain different jobs.

2.2 Weaker Assumptions

We weaken the two strong conditions used for queue length analyses in previous CQB works (Bae et al., 2026a;b; Bae & Lee, 2026a); we replace context-wise slack by capacity slack and remove the covariance lower bound. The following assumes standard logistic boundedness and i.i.d. contexts (Jun et al., 2021; Yu & Oh, 2026).

Assumption 1. We have: **1.** For every $(x, k) \in \mathcal{X} \times [K]$, $\|\boldsymbol{\phi}(x, k)\|_2 \le 1$. For some known constant $S > 0$, $\boldsymbol{\theta}^* \in \mathcal{B} := \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_2 \le S\}$.
2. Let $\kappa^{-1} = \dot{\mu}(-S) = \dot{\mu}(S)$. Also, let $\underline{p} = \mu(-S)$ and $\overline{p} = \mu(S)$.
3. The contexts of the arriving jobs are sampled i.i.d. from an unknown distribution $\mathcal{D}$.

Then $\underline{p} \le p(x, k) \le \overline{p}$ and $\dot{\mu}(\langle \boldsymbol{\phi}(x, k), \boldsymbol{\theta} \rangle) \ge 1/\kappa$ for $\boldsymbol{\theta} \in \mathcal{B}$. Next, we impose a capacity slack condition based on the expected incoming workload. Let the best service probability of a context x be $p^*(x) = \max_{k \in [K]} \mu(\langle \boldsymbol{\phi}(x, k), \boldsymbol{\theta}^* \rangle)$.

Assumption 2 (Capacity slack)**.** The (normalized) capacity slack is positive:

$$\epsilon := 1 - \lambda \mathbb{E}_{X \sim \mathcal{D}}[1/p^*(X)] > 0.$$

The quantity $1/p^*(x)$ is the expected number of service attempts needed to complete a job with context x on a best server. Hence, $\rho = \lambda \mathbb{E}_{X \sim \mathcal{D}}[1/p^*(X)]$ is the *traffic intensity* when jobs are served by their best servers. Let $\mu_{\text{eff}} = (\mathbb{E}_{X \sim \mathcal{D}}[1/p^*(X)])^{-1}$ be the corresponding effective service rate. Then $\epsilon = 1 - \rho = (\mu_{\text{eff}} - \lambda)/\mu_{\text{eff}}$. Thus, Assumption 2 is the contextual form of the classical *load condition* $\rho < 1$, equivalently *capacity slack* $\mu_{\text{eff}} - \lambda > 0$. We use *normalized capacity slack* for the parameter ϵ. This condition concerns the system's load under best-server service; stability of the learner's queue depends on its scheduling policy.

Remark 1 (Remark on Assumption 2)**.** The decaying regret bounds in previous works (Bae et al., 2026a;b; Bae & Lee, 2026a) require context-wise slack: there exists $\epsilon_0 > 0$ such that $p^*(x) \ge \lambda + \epsilon_0$ for every $x \in \mathcal{X}$. Here, ϵ_0 is a positive lower bound on the service-rate gap $p^*(x) - \lambda$ across contexts. Our condition is weaker than this since $1 - \lambda \mathbb{E}_{X \sim \mathcal{D}}[1/p^*(X)] \ge 1 - \lambda/(\lambda + \epsilon_0) = \epsilon_0/(\lambda + \epsilon_0) > 0$. Capacity slack allows contexts with $p^*(x) < \lambda$, so even the best available decision can give positive queue length drift. Our first idea measures the expected service attempts required by waiting jobs. For this workload potential, capacity slack gives a negative drift term independent of the selected context. The remaining term is controlled by the departure rate gap. We elaborate this in Section 3.2.

Remark 2 (Remark on Assumption 1-3)**.** The previous works assume that the minimum eigenvalue of the context covariance is positive, i.e., $\lambda_{\min}(\mathbb{E}_{X \sim \mathcal{D}}[\frac{1}{K} \sum_{k \in [K]} \boldsymbol{\phi}(X, k) \boldsymbol{\phi}(X, k)^\top]) > 0$. We relax this condition, thereby allowing more general context distributions, e.g., a distribution supported on a low-dimensional subspace. However, without the condition, random exploration does not give the same uniform control of estimation error. Instead, we use the elliptical potential count, which holds for the observed feature sequence, to control workload and then the departure rate gap at a specified time. We elaborate this in Section 3.3.

3 Challenges And Main Ideas

In this section, we present two main ideas for addressing the challenges in achieving anytime rate-optimal queue length regret under the weaker assumptions: (i) we design new workload potentials that yield negative drift under capacity slack, and (ii) we use the elliptical potential count to control the number of rounds with a large departure rate gap, thereby enabling learning without a covariance lower bound. In Section 3.1, we first explain why previous analyses require stronger assumptions. Then, in Sections 3.2 and 3.3, we show how we address each corresponding difficulty.

3.1 Negative Drift Under Strong Assumptions

A key step in previous analyses of decaying queue length regret is to establish negative drift, $\mathbb{E}[Q(t+1) - Q(t) \mid \mathcal{F}_t] < 0$, thus queue length decreases in expectation on these rounds. Define the departure rate gap $g_t = \mu(\langle \boldsymbol{\phi}_t^*, \boldsymbol{\theta}^* \rangle) - \mu(\langle \boldsymbol{\phi}_t, \boldsymbol{\theta}^* \rangle)$, with $g_t = 0$ when $Q(t) = 0$. Let $D^*(t) \mid \mathcal{F}_t \sim \mathrm{Bern}(\mu(\langle \boldsymbol{\phi}_t^*, \boldsymbol{\theta}^* \rangle))$ be the departure indicator for the best pair in the learner's queue. Then

$$\begin{aligned}\mathbb{E}[Q(t+1) - Q(t) \mid \mathcal{F}_t] &= \mathbb{E}[A(t) - D^*(t) \mid \mathcal{F}_t] + \mathbb{E}[D^*(t) - D(t) \mid \mathcal{F}_t] \\ &= \underbrace{\lambda - \mu(\langle \boldsymbol{\phi}_t^*, \boldsymbol{\theta}^* \rangle)}_{\leq -\epsilon_0,\ \text{context-wise slack}} + \underbrace{\mu(\langle \boldsymbol{\phi}_t^*, \boldsymbol{\theta}^* \rangle) - \mu(\langle \boldsymbol{\phi}_t, \boldsymbol{\theta}^* \rangle)}_{= g_t,\ \text{departure rate gap}}.\end{aligned}$$

Context-wise slack (see Remark 1) makes the first term at most $-\epsilon_0$. A covariance lower bound (see Remark 2) allows random exploration to reduce uncertainty in every feature direction. Therefore, after sufficient exploration, the learner can reduce the departure rate gap g_t below ϵ_0 for all subsequent decisions. Combining the two terms gives negative queue length drift.

However, neither conclusion follows under our assumptions. First, under capacity slack, the first term is not guaranteed to be negative, even for the best decision. Second, without a covariance lower bound, random exploration does not give the same uniform control of uncertainty. In the next subsection, we address the first difficulty by studying the drift of a new potential in place of $Q(t)$.

3.2 Workload Drift Under Capacity Slack

Our first idea is to define a workload potential that measures the *expected number of service attempts* required by the waiting jobs. A job with context x requires $1/p^*(x)$ service attempts on average on a best server. Define the normalized workload and its update by

$$W(t) = \underline{p} \sum_{x \in \mathcal{X}_t} \frac{1}{p^*(x)}, \qquad W(t+1) = W(t) + \frac{\underline{p} A(t)}{p^*(X_t)} - \frac{\underline{p} D(t)}{p^*(x_t)}.$$

The departure term is defined as zero when $Q(t) = 0$. Since $\underline{p} Q(t) \leq W(t)$, workload bounds also control queue length. A service attempt on a best server removes expected workload $p^*(x)\,\underline{p}/p^*(x) = \underline{p}$ for every context x. On a nonempty round,

$$\begin{aligned}\mathbb{E}[W(t+1) - W(t) \mid \mathcal{F}_t] &= \lambda \underline{p}\, \mathbb{E}_{X \sim \mathcal{D}}[1/p^*(X)] - \underline{p}\, p(x_t, k_t)/p^*(x_t) \\ &= \underbrace{-\epsilon \underline{p}}_{\text{Assumption 2: } \epsilon > 0} + \underbrace{\underline{p}\big(p^*(x_t) - p(x_t, k_t)\big)/p^*(x_t)}_{\leq g_t}. \qquad (1)\end{aligned}$$

Capacity slack makes the first term negative for every selected job. The second term is at most g_t because $\underline{p} \leq p^*(x_t) \leq p(x_t^*, k_t^*)$. Consequently, $g_t < \epsilon \underline{p}$ gives negative workload drift. It remains to control g_t without a covariance lower bound.

3.3 Learning With Elliptical Potential Count

Our second idea is to use the *elliptical potential count* to control the departure rate gap $E[g_t]$ at a specified time. The elliptical potential count (Lemma 3) is a consequence of the well-known elliptical potential lemma (Lemma 11 of Abbasi-Yadkori et al., 2011, Lemma 14), which says that we cannot choose too many feature vectors that are far from, or uncertain with respect to, the previously observed feature vectors. We use the count to bound the number of rounds with a large departure rate gap, and show that learning is possible without a covariance lower bound:

Lemma 1. *For some absolute constant $C > 0$, for all sufficiently large t,*

$$\mathbb{E}[g_t] \le C \log^{3/2}(t+1) \min\{\lambda, \beta_t \sqrt{\lambda d/t}\} + C d^2 \beta_t^6 \log^4(t+1)/t.$$

We now explain the main ideas behind the proof of Lemma 1.

Coarse learning over an interval. (1) gives negative drift when $g_i < \epsilon \underline{p}$. But without a covariance lower bound, direct control of g_i for *every round* is hard. Instead, we focus on the number of rounds with a large departure rate gap over an *interval*. Suppose our algorithm gives confidence intervals $[\underline{p}_i(x,k), \overline{p}_i(x,k)]$ that contain the true service probabilities $p(x,k)$ with high probability, thereby gives $g_i \le 2w_i \le \beta_{i-1} \|\boldsymbol{\phi}_i\|_{\mathbf{V}_{i-1}^{-1}}$. We call $w_i(x,k) = \overline{p}_i(x,k) - \underline{p}_i(x,k)$ the width and write $w_i = w_i(x_i, k_i)$ for the selected pair. Here $\beta_i = \mathcal{O}(\sqrt{d \log i})$ is the confidence radius, and $\mathbf{V}_i = \mathbf{I} + \sum_{j=1}^{i} \boldsymbol{\phi}_j \boldsymbol{\phi}_j^\top$. For an interval of length $\ell = t_2 - t_1 + 1$ and $0 < r \le 1$, the elliptical potential count gives

$$\sum_{i=t_1}^{t_2} \mathbf{1}\{w_i > r\} \le 1 + \frac{d\beta_{t_2}^2}{2r^2} \log(1 + \ell/d). \tag{2}$$

Write $\Delta W(i) = W(i+1) - W(i)$. If the interval consists of nonempty rounds, summing (1) gives

$$\begin{aligned} W(t_2+1) - W(t_1) &\le -\epsilon \underline{p}\, \ell + \sum_{i=t_1}^{t_2} g_i + \sum_{i=t_1}^{t_2} \big(\Delta W(i) - \mathbb{E}[\Delta W(i) \mid \mathcal{F}_i]\big) \\ &\le -\epsilon \underline{p}\, \ell/2 + \sum_{i=t_1}^{t_2} \mathbf{1}\{g_i > \epsilon \underline{p}/2\} + \sum_{i=t_1}^{t_2} \big(\Delta W(i) - \mathbb{E}[\Delta W(i) \mid \mathcal{F}_i]\big). \end{aligned} \tag{3}$$

The second inequality uses $g_i \le \epsilon \underline{p}/2 + \mathbf{1}\{g_i > \epsilon \underline{p}/2\}$. Since $g_i \le 2w_i$, (2) with $r = \epsilon \underline{p}/4$ bounds the second term on the right-hand side. The last term is the sum of martingale differences, so the Azuma-Hoeffding inequality (Lemma 16) bounds the last term by $\sqrt{8\ell \log(1/\delta)}$. Therefore, on a sufficiently long nonempty interval, the linear decrease $-\epsilon \underline{p}\, \ell/2$ dominates both terms (Lemma 4). If this decrease exceeds the initial workload $W(t_1)$, this implies that the queue must empty at some round in the interval. This gives an upper bound of order $\widetilde{\mathcal{O}}(\log^2 t)$ on the length of nonempty intervals, and a bound on the workload $W(t)$ at a specified time (Lemma 5).

Fine learning from arrivals after empty rounds. The coarse analysis limits the number of rounds with a large departure rate gap, but does not locate those rounds. To control $\mathbb{E}[g_t]$, we now bound the probability that a newly arriving job has uncertainty. Define a candidate server as one that is not ruled out as best for the job. Define

$$z_i(r) = \mathbb{P}_{X \sim \mathcal{D}} \left(\exists k \in [K] : k \text{ is a candidate for } X \text{ at } i, \quad w_i(X,k) > r\right).$$

$z_i(r)$ measures the probability that a newly arriving job at round i has high uncertainty. Unlike previous works, which control the largest uncertainty over all possible contexts with a covariance lower bound, we only control the uncertainty under the arrival distribution $\mathcal{D}$.

We first connect $z_i(r)$ to the actual number of rounds with selected width greater than r, which is also bounded by the elliptical potential count. Assume the intervals are nested on $[t_1, t_2+1]$, remain unchanged on empty rounds, and we select the widest candidate. If $Q(i) = 0$ and $A(i) = 1$, the new job is alone in round $i+1$. Whenever its context has a candidate wider than r, the selected interval is also wider than r. Independence of the arrival gives $\mathbb{P}(w_{i+1} > r \mid \mathcal{F}_i) \ge \lambda z_i(r) \mathbf{1}\{Q(i) = 0\}$. Taking expectations, summing, and applying (2) yields the upper bound of

$$\lambda \mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r) \mathbf{1}\{Q(i) = 0\}\right] \le \mathbb{E}\left[\sum_{i=t_1+1}^{t_2+1} \mathbf{1}\{w_i > r\}\right] \le 1 + \frac{d\beta_{t_2+1}^2}{2r^2} \log(1 + \ell/d). \tag{4}$$

Here, what we actually need is an upper bound on $\mathbb{E}[z_{t_2}(r)]$. Therefore, we need to account for nonempty rounds when $Q(i) \ne 0$. On an empty round, we have $\mathbb{E}[\Delta W(i) \mid \mathcal{F}_i] \le (1-\epsilon)\underline{p}$. Hence we extend (1) to every empty/nonempty round as $\mathbb{E}[\Delta W(i) \mid \mathcal{F}_i] \le -\epsilon \underline{p} + g_i + \underline{p} \mathbf{1}\{Q(i) = 0\}$. Since $0 \le g_i \le 1$, we have $g_i \le \epsilon \underline{p}/2 + \mathbf{1}\{g_i > \epsilon \underline{p}/2\}$, which gives

$$\epsilon \underline{p}/4 \le \mathbf{1}\{Q(i) = 0\} - \mathbb{E}[\Delta W(i) \mid \mathcal{F}_i] + 2 \cdot \mathbf{1}\{g_i > \epsilon \underline{p}/2\}.$$

We multiply by $z_i(r)$, sum, and take expectations to obtain the left-hand side of (4). Since $z_i(r)$ is known at round i, $\mathbb{E}[z_i(r)\mathbb{E}[\Delta W(i) \mid \mathcal{F}_i]] = \mathbb{E}[z_i(r)\Delta W(i)]$. Therefore, we have

$$\frac{\epsilon \underline{p}}{4}\mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r)\right] \le \mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r)\mathbf{1}\{Q(i)=0\}\right] - \mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r)\Delta W(i)\right] + 2\mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r)\mathbf{1}\{g_i > \epsilon\underline{p}/2\}\right].$$

For the first and third terms on the right-hand side, we apply (4) and the elliptical potential count with $z_i(r) \le 1$, respectively. For the rest of the terms, we assume monotonicity, $z_{i-1}(r) \ge z_i(r)$. We later show that our *interval intersection* procedure in Algorithm 1 can guarantee this property within each epoch. Then, for the second term on the right-hand side,

$$-\sum_{i=t_1}^{t_2} z_i(r)\Delta W(i) = z_{t_1}(r)W(t_1) - z_{t_2}(r)W(t_2+1) - \sum_{i=t_1+1}^{t_2}(z_{i-1}(r)-z_i(r))W(i) \le W(t_1). \tag{5}$$

Also, for the left-hand side, $\ell z_{t_2}(r) \le \sum_{i=t_1}^{t_2} z_i(r)$. Then, we have

$$\frac{\epsilon \underline{p}\,\ell}{4}\mathbb{E}[z_{t_2}(r)] \le \mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r)\mathbf{1}\{Q(i)=0\}\right] + \mathbb{E}[W(t_1)] + 2\mathbb{E}\left[\sum_{i=t_1}^{t_2}\mathbf{1}\{g_i > \epsilon\underline{p}/2\}\right]. \tag{6}$$

Dividing both sides by $\epsilon\underline{p}\,\ell/4$ gives the desired upper bound of $\mathbb{E}[z_{t_2}(r)]$ (Lemma 6).

Next, we need to connect the uncertainty of a new job sampled from $\mathcal{D}$ to that of the currently selected pair. The difficulty is that the job selected at time t is not an independent sample from $\mathcal{D}$, so we cannot write $\Pr(w_t > r) = \mathbb{E}[z_t(r)]$. Instead, we return to the job's arrival time to use independence. First, suppose the queue was empty at some round in $[t-\ell, t-1]$. Then every job remaining at time t must have arrived after that empty round. Therefore, the currently selected job arrived in one of the rounds $t-\ell,\ t-\ell+1,\ \ldots,\ t-1$. The *interval intersection* assumed above ensures that if the current width is greater than r, then that server's width was also greater than r just after arrival and it was a candidate at that time. Thus, the job satisfied the event of $z_i(r)$. Independence of the arrival and its context gives $\Pr(A(i-1)=1,\ X_{i-1}$ has a wide candidate at time $i) = \lambda\,\mathbb{E}[z_i(r)]$. Taking a union bound over the possible arrival times $t-\ell,\ldots,t-1$ gives (see the proof of Lemma 7)

$$\mathbb{P}(w_t > r) \le \lambda \sum_{i=t-\ell+1}^{t} \mathbb{E}[z_i(r)] + \mathbb{P}\big(\text{the job selected at } t \text{ arrived before } t-\ell\big). \tag{7}$$

We bound the first term using (6) and the second term using the fact that busy periods are short with high probability, as shown in the coarse learning analysis. Finally, using the fact that $\mathbb{E}[g_t] = \int_0^1 \Pr(g_t > r)\,dr$ with $g_t \le 2w_t$ we obtain the desired result. The assumptions made in this section are guaranteed by Algorithm 1 and elaborated in Section 4.

4 Algorithm

We propose WISE (Widest Interval Selection with Elimination), shown in Algorithm 1. The algorithm estimates service probabilities, intersects their confidence intervals, and selects the widest interval among pairs that have not been eliminated. We estimate the unknown parameter by regularized logistic maximum likelihood,

$$\widehat{\boldsymbol{\theta}}_t = \operatorname*{arg\,min}_{\boldsymbol{\theta}\in\mathcal{B}}\left\{\sum_{i=1}^{t}\left[-D(i)\log\mu(\langle\boldsymbol{\phi}_i,\boldsymbol{\theta}\rangle) - (1-D(i))\log\big(1-\mu(\langle\boldsymbol{\phi}_i,\boldsymbol{\theta}\rangle)\big)\right] + \|\boldsymbol{\theta}\|_2^2/2\right\}. \tag{8}$$

With $\mathbf{V}_t = \mathbf{I} + \sum_{i=1}^{t}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top$, update β_t by

$$\beta_t = (\kappa/2)\sqrt{d\log(1+t/(d\kappa)) + 4\log(t+1)} + \sqrt{\kappa}S. \tag{9}$$

Algorithm 1 WISE (Widest Interval Selection with Elimination)

input: $d, K, S, \boldsymbol{\phi}$
initialize: $\mathbf{V}_0 \leftarrow \mathbf{I}$, $\widehat{\boldsymbol{\theta}}_0 \leftarrow \mathbf{0}$, $\beta_0 \leftarrow \sqrt{\kappa}S$, $U_0 \sim \mathrm{Unif}(0,1)$, $\mathcal{I}_0(x,k) \leftarrow [\underline{p}, \overline{p}]$
for $t = 1, 2, \ldots$ **do**
  Determine m_t and τ_{m_t} by (10) // current epoch
  if $\mathcal{X}_t \neq \varnothing$ **then**
    for $(x,k) \in \mathcal{X}_t \times [K]$ **do**
      $\mathcal{J} \leftarrow \mathcal{I}_{\tau_{m_t}-1}(x,k)$
      for $i = \tau_{m_t}, \ldots, t-1$ **do**
        if $\mathcal{J} \cap \mathcal{I}_i(x,k) = \varnothing$ **then**
          $\mathcal{J} \leftarrow \{(\min \mathcal{J} + \max \mathcal{J})/2\}$ and **break** // freeze until the next epoch
        end if
        $\mathcal{J} \leftarrow \mathcal{J} \cap \mathcal{I}_i(x,k)$ // intersect confidence intervals
      end for
      $[\underline{p}_t(x,k), \overline{p}_t(x,k)] \leftarrow \mathcal{J}$
    end for
    $\mathcal{Z}_t \leftarrow \{(x,k) \in \mathcal{X}_t \times [K] : \overline{p}_t(x,k) \geq \max_{(x',k') \in \mathcal{X}_t \times [K]} \underline{p}_t(x',k')\}$ // elimination
    $(x_t, k_t) \in \arg\max_{(x,k) \in \mathcal{Z}_t} [\overline{p}_t(x,k) - \underline{p}_t(x,k)]$ // widest interval selection
    Serve (x_t, k_t) and observe $D(t)$
    $\mathbf{V}_t \leftarrow \mathbf{V}_{t-1} + \boldsymbol{\phi}_t \boldsymbol{\phi}_t^\top$, update $\widehat{\boldsymbol{\theta}}_t$ by (8) and β_t by (9)
  else
    $\mathbf{V}_t \leftarrow \mathbf{V}_{t-1}$, $\widehat{\boldsymbol{\theta}}_t \leftarrow \widehat{\boldsymbol{\theta}}_{t-1}$, $\beta_t \leftarrow \beta_{t-1}$ // no updates when queue is empty
  end if
end for

Intersecting confidence intervals. Within each epoch, intersecting confidence intervals makes their widths and $z_i(r)$ nonincreasing. This is the property used in (5) and in comparing a job's current width with its width at arrival. From the estimates, form

$$\mathcal{I}_i(x,k) = [\mu(\langle \boldsymbol{\phi}(x,k), \widehat{\boldsymbol{\theta}}_i \rangle - \beta_i \|\boldsymbol{\phi}(x,k)\|_{\mathbf{V}_i^{-1}}), \mu(\langle \boldsymbol{\phi}(x,k), \widehat{\boldsymbol{\theta}}_i \rangle + \beta_i \|\boldsymbol{\phi}(x,k)\|_{\mathbf{V}_i^{-1}})] \cap [\underline{p}, \overline{p}].$$

We restart the intersection at epoch boundaries to prevent an invalid interval from affecting all later decisions. With $U_0 \sim \mathrm{Unif}(0,1)$ drawn once, set

$$\tau_0 = 1, \qquad \tau_m = \lceil 2^{m-1+U_0} \rceil, \qquad m_t = \max\{m : \tau_m \leq t\}. \tag{10}$$

At an epoch boundary, start from $\mathcal{I}_{t-1}(x,k)$. Within the epoch, intersect this interval with each new confidence interval. If the intersection becomes empty, set both endpoints to the midpoint of the last nonempty intersection and keep them fixed until the next epoch, preserving nesting. Estimation retains all observations across epochs. The random shift bounds the probability of a reset shortly before any specified time (28).

Elimination and widest interval selection. The candidate set $\mathcal{Z}_t$ retains pairs whose upper bound is at least the largest lower bound in the queue. Selecting the widest candidate gives $g_t \leq 2w_t$ (Lemma 2), so the elliptical potential count bounds the number of rounds with a large departure rate gap. For a job arriving alone, a candidate wider than r also guarantees an observation of width greater than r, giving the relation in (4). These are the two properties used in Section 3.3. Appendix A compares the purpose and guarantees of this rule with previous widest interval selection algorithms.

5 Main Results

5.1 Upper Bound

Theorem 1. *Under Assumptions 1 and 2, there exist an absolute constant $C_1 > 0$ and a threshold t_0 satisfying the conditions in (50) such that, for every $t \geq t_0$, the WISE algorithm satisfies*

$$R_t \leq C_1 \log^3(t+1) \min\{\lambda, \beta_t \sqrt{\lambda d / t}\}.$$

Since $\beta_t = \mathcal{O}(\sqrt{d \log t})$, the WISE algorithm attains $\widetilde{\mathcal{O}}(t^{-1/2})$ queue length regret at every sufficiently large time without knowing t.

Proof sketch. Let π_i follow the learner through round i and the oracle afterwards, with queue length $Q^{\pi_i}(j)$ at round j. The coupling of Bae et al. (2026a) (Lemma 9) gives $R_t \le \sum_{i=1}^{t-1} \mathbb{E}[g_i q_{i,t}]$, where g_i is the departure rate gap in Section 3.1 and, for $g_i > 0$,

$$q_{i,t} = \mathbb{P}\big(Q^{\pi_i}(t) - Q^{\pi_{i-1}}(t) = 1 \mid \mathcal{F}_i,\ Q^{\pi_i}(i+1) - Q^{\pi_{i-1}}(i+1) = 1\big).$$

Set $q_{i,t} = 0$ when $g_i = 0$. The conditioning event is a service failure under π_i and a success under π_{i-1} at round i. Thus $q_{i,t}$ measures whether the resulting extra job still affects queue length at t. On this event, $\mathcal{X}_j^{\pi_{i-1}} \subseteq \mathcal{X}_j^{\pi_i}$ for $j \ge i+1$. Therefore,

$$Q^{\pi_i}(j) = 0 \text{ for some } j \in [i+1, t] \quad \Longrightarrow \quad Q^{\pi_i}(t) = Q^{\pi_{i-1}}(t).$$

The workload of Q^{π_i} starts at $\underline{p} \sum_{x \in \mathcal{X}_{i+1}^{\pi_i}} 1/p^*(x) \le W(i) + 1$. Its oracle decisions give workload drift $-\epsilon \underline{p}$ until Q^{π_i} empties. If $W(i)+1$ is small compared with the remaining $t-i$ rounds, Azuma-Hoeffding (Lemma 16) bounds the probability that $Q^{\pi_i}(j) > 0$ throughout $[i+1, t]$. This gives (Lemma 10)

$$\mathbb{E}[q_{i,t}] \le \mathbb{P}\big(W(i) + 1 > \epsilon \underline{p}(t-i)/32\big) + C e^{-(t-i)/C}. \tag{11}$$

Since $0 \le g_i, q_{i,t} \le 1$, we can bound $\mathbb{E}[g_i q_{i,t}]$ by either $\mathbb{E}[q_{i,t}]$ or $\mathbb{E}[g_i]$. We use the former for the first two intervals and the latter for the last. Here C hides dependence on $d, \kappa, S, \epsilon, \underline{p}$, and we choose $t - \tau_{1,t} = \Theta(\log^2 t)$ and $t - \tau_{2,t} = \Theta(\log t)$.

First interval ($1 \le i \le \tau_{1,t}$)**.** The workload bound of order $\log^2(i+1)$ suffices here because we choose $\tau_{1,t}$ so that $\epsilon \underline{p}(t-i)/32 \ge C \log^2(i+1) + 1$. Thus the probability in (11) is at most $\mathbb{P}(W(i) > C \log^2(i+1))$. For $i > \epsilon \underline{p}\, t/96$, Lemma 5, using (3), bounds this by $C(i+1)^{-2}$. For smaller i, $W(i) + 1 \le i + 1$ makes the original probability zero. Summing these probabilities and the exponential terms gives $\sum_{i=1}^{\tau_{1,t}} \mathbb{E}[g_i q_{i,t}] \le C/t$.

Second interval ($\tau_{1,t} < i \le \tau_{2,t}$)**.** Now $\epsilon \underline{p}(t-i)/32$ can be smaller than $C \log^2(i+1) + 1$, so the preceding workload estimate no longer suffices. The learning bounds in (6) and (7) give the tail (Lemma 8), valid for every $z \ge 0$, $\mathbb{P}(W(i) > z) \le C e^{-\epsilon \underline{p} z/2} + C \log^7(i+1)/i$. Applying it with $z = \epsilon \underline{p}(t-i)/32 - 1$ bounds the probability in (11). The exponential terms sum to C/t, while the remainder contributes $C \log^9(t+1)/t$ over $O(\log^2 t)$ rounds. Hence $\sum_{i=\tau_{1,t}+1}^{\tau_{2,t}} \mathbb{E}[g_i q_{i,t}] \le C/t + C \log^9(t+1)/t$.

Third interval ($\tau_{2,t} < i < t$)**.** A service error just before t can still affect $Q(t)$, so we bound the departure rate gap directly using $\mathbb{E}[g_i q_{i,t}] \le \mathbb{E}[g_i]$. Applying Lemma 1 over the remaining $O(\log t)$ rounds multiplies its leading term by $O(\log t)$ and adds at most $C \log^8(t+1)/t$ from its remainder. Adding the three contributions proves the result for sufficiently large t. See Section B for the full proof and cutoff choices. □

5.2 Lower Bounds

We prove lower bounds for two questions. Theorem 2 determines how fast regret can decay with t when the normalized capacity slack is bounded away from zero. Theorem 3 shows how a small normalized capacity slack increases regret.

For Theorem 2, we sharpen the λ dependence in Bae & Lee (2026a) by counting service observations. In our construction, jobs require $O(1)$ expected service attempts, giving $O(\lambda t)$ observations in expectation. Service probabilities separated by order $\min\{1, 1/\sqrt{\lambda t}\}$ can therefore remain hard to distinguish. An independent arrival before the last decision forces service with probability λ, giving a regret lower bound of order $\lambda \min\{1, 1/\sqrt{\lambda t}\} = \min\{\lambda, \sqrt{\lambda/t}\}$. See Section C for the full construction, including d and K.

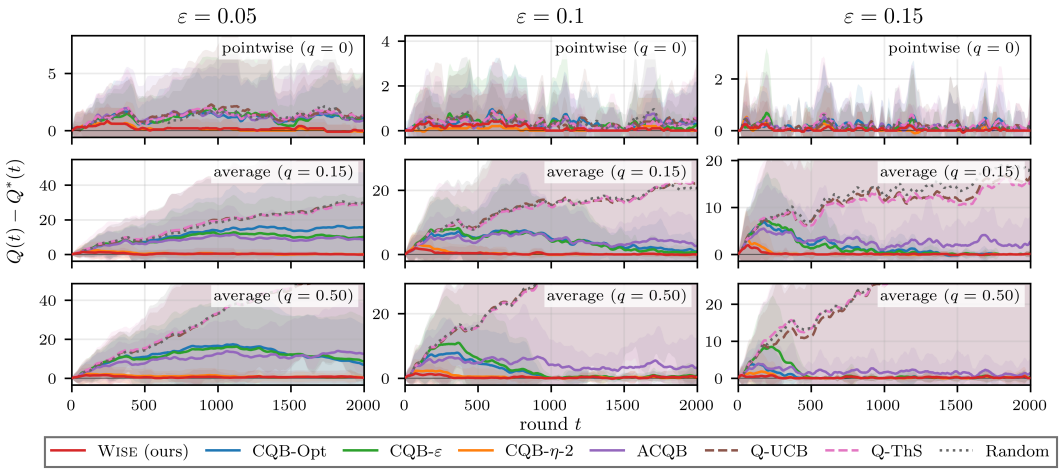


Figure 1: Queue length regret averaged over ten runs, with ± 2 sample standard deviations and 25-round smoothing. Columns use normalized capacity slack $\epsilon = 0.05, 0.1, 0.15$. Rows use pointwise, average ($q = 0.15$), and average ($q = 0.50$) distributions, where $q = \mathbb{P}_{X \sim \mathcal{D}}(p^*(X) < \lambda)$.

Theorem 2. *There is an absolute constant $C_0 \geq 1$ such that for every $d \geq 2$, $K \geq 2$, $\lambda \in (0, 1/4]$, and $t \geq 3$, there is a finite family $\mathcal{V}$ of problem instances with $S = 1/4$ and normalized capacity slack $\epsilon = 1 - 2\lambda \geq 1/2$ for which every algorithm π satisfies*

$$\sup_{\nu \in \mathcal{V}} R_t^{\pi}(\nu) \geq C_0^{-1} \min\{\lambda, \min\{d, \log K\}\sqrt{\lambda/t}\}.$$

Together with Theorem 1, this establishes rate-optimality in t up to logarithmic factors.

For Theorem 3, a small normalized capacity slack lets service errors persist for $\Theta(\epsilon^{-2})$ rounds. In our construction, the best service probability exceeds λ by only $\Theta(\epsilon)$. With constant probability, the queue contains at least a constant times $1/\epsilon$ jobs and stays nonempty for $\Theta(\epsilon^{-2})$ rounds even under the best server (Lemma 11). A service error at i leaves $Q^{\pi_i}(i+1) = Q^{\pi_{i-1}}(i+1) + 1$. This extra job remains while $Q^{\pi_{i-1}}$ is nonempty. We lower-bound the departure rate gap on these states and sum its contributions over the last $\Theta(\epsilon^{-2})$ rounds, giving the ϵ^{-2} factor. See Section D for the proof.

Theorem 3. *There is an absolute constant $C_0 \geq 1$ such that for every $d, K \geq 2$, $0 < \epsilon \leq 1/8$, and $t \geq 64/\epsilon^2$, there is a finite family $\mathcal{V}_\epsilon$ of problem instances with $\lambda = 1/4$, $S = 2$, and normalized capacity slack ϵ for which every algorithm π satisfies*

$$\sup_{\nu \in \mathcal{V}_\epsilon} R_t^{\pi}(\nu) \geq C_0^{-1} \min\{\epsilon^{-1}, \min\{d, \log K\}/(\epsilon^2\sqrt{t})\}.$$

6 Experiments

We evaluate WISE with and without context-wise slack. We compare it with implementations of CQB-Opt and CQB-ϵ (Bae et al., 2026a), CQB-η-2 (Bae & Lee, 2026a), ACQB (Bae et al., 2026b), Q-UCB and Q-ThS (Krishnasamy et al., 2021), and uniformly random selection.

We use $d = 5$, $K = 10$, $\lambda = 0.6$, $\|\phi(x, k)\|_2 = 5$, and one shared parameter with $\|\boldsymbol{\theta}^*\|_2 = S = 10$. All distributions have the same normalized capacity slack within each column. The pointwise distribution satisfies context-wise slack, whereas the two average distributions increase both the frequency and the size of violations. All contextual algorithms use constrained regularized logistic estimates and a fixed confidence or sampling radius $\beta = c = 100$, with input $\kappa = 100$. WISE retains the interval intersections, resets, elimination, and widest interval selection. We run each setting ten times for 2000 rounds. Implementation and distribution details are in Section F.

In all six average-distribution settings, WISE has the smallest mean queue length regret over rounds 1-2000 among the compared algorithms. Its average regret over rounds 1500-2000 is below 0.75 in all nine settings, including those without context-wise slack.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-wise minimax-optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 3691–3699. PMLR, 2021.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of machine learning research*, 3(Nov):397–422, 2002.

Seoungbin Bae and Dabeen Lee. Algorithm for contextual queueing bandits with rate-optimal queue length regret. *arXiv preprint arXiv:2606.09668*, 2026a.

Seoungbin Bae and Dabeen Lee. Logistic bandits with $\widetilde{O}(\sqrt{dT})$ regret without context diversity assumptions. *arXiv preprint arXiv:2604.22161*, 2026b.

Seoungbin Bae, Garyeong Kang, and Dabeen Lee. Queue length regret bounds for contextual queueing bandits. *arXiv preprint arXiv:2601.19300*, 2026a.

Seoungbin Bae, Junyoung Son, and Dabeen Lee. Learning to route and schedule llms from user retrials via contextual queueing bandits. *arXiv preprint arXiv:2602.02061*, 2026b.

Xu Cai, Selwyn Gomes, and Jonathan Scarlett. Lenient regret and good-action identification in gaussian process bandits. In *International Conference on Machine Learning*, pp. 1183–1192. PMLR, 2021.

Tuhinangshu Choudhury, Gauri Joshi, Weina Wang, and Sanjay Shakkottai. Job dispatching policies for queueing systems with unknown service rates. In *Proceedings of the Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, pp. 181–190, 2021.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 208–214. JMLR Workshop and Conference Proceedings, 2011.

Emile Contal, David Buffoni, Alexandre Robicquet, and Nicolas Vayatis. Parallel gaussian process optimization with upper confidence bound and pure exploration. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 225–240. Springer, 2013.

Shi Dong, Tengyu Ma, and Benjamin Van Roy. On the performance of thompson sampling on logistic bandits. In *Conference on Learning Theory*, pp. 1158–1160. PMLR, 2019.

Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning*, pp. 3052–3060. PMLR, 2020.

Louis Faury, Marc Abeille, Kwang-Sung Jun, and Clément Calauzènes. Jointly efficient and optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 546–580. PMLR, 2022.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. *Advances in neural information processing systems*, 23, 2010.

Daniel Freund, Thodoris Lykouris, and Wentao Weng. Quantifying the cost of learning in queueing systems. *Advances in Neural Information Processing Systems*, 36:6532–6544, 2023.

David Gamarnik and Alexander L Stolyar. Multiclass multiserver queueing system in the halfin–whitt heavy traffic regime: Asymptotics of the stationary distribution. *Queueing Systems*, 71(1): 25–51, 2012.

Alkis Gotovos. Active learning for level set estimation. 2013.

Shogo Iwazaki and Shion Takeno. Improved regret analysis in gaussian process bandits: Optimality for noiseless reward, rkhs norm, and non-stationary variance. *arXiv preprint arXiv:2502.06363*, 2025.

Kwang-Sung Jun, Lalit Jain, Blake Mason, and Houssam Nassif. Improved confidence bounds for the linear logistic model and applications to bandits. In *International Conference on Machine Learning*, pp. 5148–5157. PMLR, 2021.

Jung-hun Kim and Min-hwan Oh. Queueing matching bandits with preference feedback. *Advances in Neural Information Processing Systems*, 37:64649–64702, 2024.

Wonyoung Kim, Kyungbok Lee, and Myunghee Cho Paik. Double doubly robust thompson sampling for generalized linear contextual bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 8300–8307, 2023.

Yeoneung Kim, Insoon Yang, and Kwang-Sung Jun. Improved regret analysis for variance-adaptive linear bandits and horizon-free linear mixture mdps. *Advances in Neural Information Processing Systems*, 35:1060–1072, 2022.

Ger Koole and Avishai Mandelbaum. Queueing models of call centers: An introduction. *Annals of Operations Research*, 113(1):41–59, 2002.

Subhashini Krishnasamy, Ari Arapostathis, Ramesh Johari, and Sanjay Shakkottai. On learning the $c\mu$ rule in single and parallel server networks. *arXiv preprint arXiv:1802.06723*, 2018.

Subhashini Krishnasamy, Rajat Sen, Ramesh Johari, and Sanjay Shakkottai. Learning unknown service rates in queues: A multiarmed bandit approach. *Operations research*, 69(1):315–330, 2021.

Jiung Lee, Hongseok Namkoong, and Yibo Zeng. Design and scheduling of an ai-based queueing system. *arXiv preprint arXiv:2406.06855*, 2024a.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. A unified confidence sequence for generalized linear models, with applications to bandits. *arXiv preprint arXiv:2407.13977*, 2024b.

Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*, pp. 2071–2080. PMLR, 2017.

Zihan Li and Jonathan Scarlett. Gaussian process bandit optimization with few batches. In *International Conference on Artificial Intelligence and Statistics*, pp. 92–107. PMLR, 2022.

Thodoris Lykouris and Wentao Weng. Learning to defer in congested systems: the ai-human interplay. *arXiv preprint arXiv:2402.12237*, 2024.

Chenkai Ma, Keqin Chen, and Jonathan Scarlett. Batched kernelized bandits: Refinements and extensions. *arXiv preprint arXiv:2603.12627*, 2026.

Michael Mitzenmacher. Scheduling with predictions and the price of misprediction. *arXiv preprint arXiv:1902.00732*, 2019.

Michael Mitzenmacher and Rana Shahout. Queueing, predictions, and large language models: Challenges and open problems. *Stochastic Systems*, 15(3):195–219, 2025.

Michael J Neely, Scott T Rager, and Thomas F La Porta. Max weight learning algorithms for scheduling in unknown environments. *IEEE Transactions on Automatic Control*, 57(5):1179–1191, 2012.

Ziv Scully, Isaac Grosof, and Michael Mitzenmacher. Uniform bounds for scheduling with job size estimates. *arXiv preprint arXiv:2110.00633*, 2021.

Thomas Stahlbuhk, Brooke Shrader, and Eytan Modiano. Learning algorithms for scheduling in wireless networks with unknown channel statistics. *Ad Hoc Networks*, 85:131–144, 2019. doi: 10.1016/j.adhoc.2018.10.006. URL https://www.mit.edu/~modiano/papers/CV_J_118.pdf.

Thomas Stahlbuhk, Brooke Shrader, and Eytan Modiano. Learning algorithms for minimizing queue length regret. *IEEE Transactions on Information Theory*, 67(3):1759–1781, 2021.

Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In *International conference on machine learning*, pp. 997–1005. PMLR, 2015.

Jordi Vilaplana, Francesc Solsona, Ivan Teixidó, Jordi Mateo, Francesc Abella, and Josep Rius. A queuing theory model for cloud computing. *The Journal of Supercomputing*, 69(1):492–507, 2014.

Mevan Wijewardena, Kamiar Asgari, and Michael J Neely. Bandit-based rate adaptation for a single-server queue. *arXiv preprint arXiv:2512.12016*, 2025.

Zixian Yang, R Srikant, and Lei Ying. Learning while scheduling in multi-server systems with unknown statistics: Maxweight with discounted ucb. In *International Conference on Artificial Intelligence and Statistics*, pp. 4275–4312. PMLR, 2023.

Sanghoon Yu and Min-hwan Oh. Practical and optimal algorithm for linear contextual bandits with rare parameter updates. *arXiv preprint arXiv:2606.00984*, 2026.

Yu-Jie Zhang and Masashi Sugiyama. Online (multinomial) logistic bandit: Improved regret and constant computation cost. *Advances in Neural Information Processing Systems*, 36:29741–29782, 2023.

A Additional Related Work

Queueing systems and learning. Queueing systems model service allocation in call centers and computing platforms (Koole & Mandelbaum, 2002; Vilaplana et al., 2014). Queueing bandits address scheduling when service rates are unknown and must be learned from service outcomes (Krishnasamy et al., 2021; Choudhury et al., 2021; Stahlbuhk et al., 2021). Their guarantees include queue length regret at a given time (Krishnasamy et al., 2021), cumulative queue length or holding-cost regret (Stahlbuhk et al., 2021; Krishnasamy et al., 2018), and the maximum increase in time-averaged queue length due to learning (Freund et al., 2023). The $\widetilde{\mathcal{O}}(t^{-1})$ guarantee of Krishnasamy et al. (2021) depends on a fixed difference between the best and second-best service rates, whereas our lower bounds allow this difference to decrease with t. Contextual extensions assign a fixed feature to each queue (Kim & Oh, 2024) or allow heterogeneous contexts for individual jobs (Bae et al., 2026a;b; Bae & Lee, 2026a). For the latter setting, existing bounds decay as $\widetilde{\mathcal{O}}(t^{-1/4})$ without horizon knowledge (Bae et al., 2026b) or as $\widetilde{\mathcal{O}}(t^{-1/2})$ with a known horizon (Bae & Lee, 2026a). These results require context-wise slack and a covariance lower bound. We replace context-wise slack by capacity slack and remove the covariance lower bound. Other learning approaches for queueing systems study MaxWeight scheduling and stability under unknown service rates (Neely et al., 2012; Stahlbuhk et al., 2019; Yang et al., 2023; Wijewardena et al., 2025), scheduling with predicted service times or job classes (Mitzenmacher, 2019; Scully et al., 2021; Lee et al., 2024a; Mitzenmacher & Shahout, 2025), and learning when to defer decisions to congested human reviewers (Lykouris & Weng, 2024).

Logistic bandits. Our service model connects CQB to logistic bandits, which model binary feedback through a logistic function of action features and an unknown parameter. Following the generalized linear bandit formulation of Filippi et al. (2010), subsequent work improves the dependence on curvature and derives instance-dependent regret bounds (Faury et al., 2020; Abeille et al., 2021; Lee et al., 2024b). Related directions include sharper confidence bounds for finite-action problems (Li et al., 2017; Jun et al., 2021), computationally efficient updates (Faury et al., 2022; Zhang & Sugiyama, 2023), and Thompson sampling (Dong et al., 2019; Kim et al., 2023). Bae & Lee (2026b) remove the assumption of a positive lower bound on the minimum eigenvalue of the context covariance matrix. These works study cumulative reward regret, whereas our analysis requires a bound on the expected departure rate gap at each specified time.

Elimination algorithms. Elimination algorithms use confidence bounds to discard actions that cannot be optimal. In Auer's framework, exploration selects a remaining action whose confidence width exceeds a prescribed threshold (Auer, 2002; Chu et al., 2011; Li et al., 2017; Jun et al., 2021). Related methods for Gaussian process optimization select the action with the largest posterior variance, either in combination with UCB (Contal et al., 2013) or after eliminating suboptimal actions (Cai et al., 2021; Li & Scarlett, 2022; Iwazaki & Takeno, 2025; Ma et al., 2026). Confidence interval intersections also appear in level-set estimation and safe optimization (Gotovos, 2013; Sui et al., 2015). The BLCE algorithm of Yu & Oh (2026) explores by selecting $\boldsymbol{x}_i \in \arg\max_{\boldsymbol{x}\in\mathcal{C}_i} \|\boldsymbol{x}\|_{\mathbf{H}_{i-1}^{-1}}$ among surviving candidates $\mathcal{C}_i$, where $\mathbf{H}_i = \mathbf{I} + \sum_{j=1}^{i} \boldsymbol{x}_j \boldsymbol{x}_j^\top$. For ℓ exploration rounds with $\|\boldsymbol{x}\|_2 \le 1$ and a fixed elimination rule, their analysis gives $\mathbb{E}[\max_{\boldsymbol{x}\in\mathcal{C}_{\ell+1}} \|\boldsymbol{x}\|^2_{\mathbf{H}_\ell^{-1}}] \le \ell^{-1} \sum_{i=1}^{\ell} \mathbb{E}[\|\boldsymbol{x}_i\|^2_{\mathbf{H}_{i-1}^{-1}}] \le (2d/\ell)\log(1+\ell/d)$, where $\mathcal{C}_{\ell+1}$ comes from a fresh i.i.d. context set. The first inequality uses context independence and the selection rule, and the second follows from the elliptical potential lemma. This analysis does not directly apply to CQB because the set of waiting jobs depends on past service decisions and departures, so the first inequality cannot be justified by context independence. We instead combine observations after empty rounds with workload bounds and interval intersections to control $\mathbb{E}[g_t]$ (Section 3.3).

B Proof Of Theorem 1

We first prove the lemmas used to bound $\mathbb{E}[g_i q_{i,t}]$ and then prove Theorem 1 in Section B.6. Throughout this section, C_5 denotes a positive constant depending only on ϵ and $\underline{p}$, and may increase between inequalities. Set $C_2 = 64/(\epsilon\underline{p})^3$, $C_3 = 128/(\epsilon\underline{p})^2$, $C_4 = 32/(\epsilon\underline{p})$, and

$$\ell_{0,t} = \lceil C_2 d \beta_t^2 \log(t+1) \rceil, \qquad \ell_{1,t} = \lceil C_3 \log(t+1) \rceil. \tag{12}$$

These lengths are used only in the proof.

B.1 Proofs Of Lemmas 2 And 3

We use the filtration $\mathcal{F}_t$ in Section 2, including U_0. For the proofs, β_t denotes the deterministic expression in (9), and β_t° denotes the value stored by Algorithm 1. Thus $\beta_t^\circ = \beta_t$ if $Q(t) > 0$, and $\beta_t^\circ = \beta_{t-1}^\circ$ otherwise, with $\beta_0^\circ = \sqrt{\kappa} S$. In particular, $\beta_t^\circ \le \beta_t$. Define

$$\mathcal{C}_0 = \mathcal{B}, \qquad \mathcal{C}_t = \{\boldsymbol{\theta} \in \mathcal{B} : \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t\|_{\mathbf{V}_t} \le \beta_t^\circ\}.$$

Recall that $w_t = \overline{p}_t(x_t, k_t) - \underline{p}_t(x_t, k_t)$, and set $w_t = 0$ when $Q(t) = 0$.

Lemma 2. *For every $t_1 \ge 1$, conditional on U_0,*

$$\mathbb{P}(\exists t \ge t_1 : \boldsymbol{\theta}^* \notin \mathcal{C}_t \mid U_0) \le (\lceil t_1/2 \rceil + 1)^{-2} + e^{-\lambda \lfloor t_1/2 \rfloor}. \tag{13}$$

For sufficiently large t_1, the right-hand side is at most $5(t_1+1)^{-2}$. If $\boldsymbol{\theta}^* \in \mathcal{C}_i$ for every $i = \tau_{m_t} - 1, \dots, t-1$, then

$$\underline{p}_t(x,k) \le p(x,k) \le \overline{p}_t(x,k) \quad \textit{for all } (x,k) \in \mathcal{X} \times [K], \qquad g_t \le 2 w_t.$$

B.1.1 Proof Of Lemma 2

Proof. Fix U_0. Define

$$\eta_i = \big(D(i) - \mu(\langle \boldsymbol{\phi}_i, \boldsymbol{\theta}^* \rangle)\big) \mathbf{1}\{Q(i) > 0\}, \qquad \boldsymbol{G}_t = \sum_{i=1}^t \eta_i \boldsymbol{\phi}_i, \qquad \mathbf{V}_{\kappa,t} = \kappa \mathbf{I} + \sum_{i=1}^t \boldsymbol{\phi}_i \boldsymbol{\phi}_i^\top.$$

By the conditional Bernoulli service law and Hoeffding's lemma (Lemma 15), $\mathbb{E}[\eta_i \mid \mathcal{F}_i] = 0$ and $\mathbb{E}[e^{z\eta_i} \mid \mathcal{F}_i] \le e^{z^2/8}$ for every $z \in \mathbb{R}$. Theorem 1 of Abbasi-Yadkori et al. (2011) (Lemma 12), with initial matrix $\kappa \mathbf{I}$ and noise parameter $1/2$, gives, with probability at least $1-\delta$, simultaneously for all $t \ge 1$,

$$\|\boldsymbol{G}_t\|_{\mathbf{V}_{\kappa,t}^{-1}} \le \frac{1}{2} \sqrt{\log \frac{\det \mathbf{V}_{\kappa,t}}{\det(\kappa \mathbf{I})} + 2\log(1/\delta)} \le \frac{1}{2} \sqrt{d \log(1 + t/(d\kappa)) + 2\log(1/\delta)}. \tag{14}$$

The second inequality follows from the arithmetic-geometric mean inequality and $\sum_{i=1}^t \|\boldsymbol{\phi}_i\|_2^2 \le t$.

Lemma 13 gives

$$\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_{\kappa,t}} \le \kappa \|\boldsymbol{G}_t\|_{\mathbf{V}_{\kappa,t}^{-1}} + \sqrt{\kappa} S.$$

Set $t_2 = \lceil t_1/2 \rceil$ and $\delta = (t_2+1)^{-2}$. Applying (14) gives, simultaneously for every $i \ge t_2$,

$$\begin{aligned} \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}^*\|_{\mathbf{V}_i} &\le \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}^*\|_{\mathbf{V}_{\kappa,i}} \\ &\le \frac{\kappa}{2} \sqrt{d \log(1 + i/(d\kappa)) + 4\log(t_2+1)} + \sqrt{\kappa} S \le \beta_i. \end{aligned} \tag{15}$$

The first inequality follows from $\kappa \ge 1$, and the last follows from $i \ge t_2$ and (9).

Suppose $Q(j) > 0$ for some $j \in [t_2, t_1]$. For any $i \ge t_1$, take $j = \max\{j' \le i : Q(j') > 0\}$. Then $j \ge t_2$, and Algorithm 1 gives

$$\widehat{\boldsymbol{\theta}}_i = \widehat{\boldsymbol{\theta}}_j, \qquad \mathbf{V}_i = \mathbf{V}_j, \qquad \beta_i^\circ = \beta_j.$$

Thus (15) at j implies $\boldsymbol{\theta}^* \in \mathcal{C}_i$. If $Q(j) = 0$ for every $j \in [t_2, t_1]$, then $A(j) = 0$ for every $j \in [t_2, t_1 - 1]$. Hence

$$\begin{aligned} \mathbb{P}(\exists i \ge t_1 : \boldsymbol{\theta}^* \notin \mathcal{C}_i \mid U_0) &\le (t_2+1)^{-2} + \mathbb{P}(A(j) = 0 \text{ for all } j \in [t_2, t_1 - 1] \mid U_0) \\ &= (t_2+1)^{-2} + (1-\lambda)^{t_1 - t_2} \\ &\le (\lceil t_1/2 \rceil + 1)^{-2} + e^{-\lambda \lfloor t_1/2 \rfloor}, \end{aligned}$$

where the equality follows from independence of the arrivals and U_0. This proves (13). If $\lambda \lfloor t_1/2 \rfloor \ge 2\log(t_1+1)$, its right-hand side is at most $4(t_1+1)^{-2} + (t_1+1)^{-2}$.

We next prove the interval inclusion and $g_t \le 2w_t$. If $\boldsymbol{\theta}^* \in \mathcal{C}_i$ and $i \ge 1$, then

$$|\langle \boldsymbol{\phi}(x,k), \widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}^* \rangle| \le \|\boldsymbol{\phi}(x,k)\|_{\mathbf{V}_i^{-1}} \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}^*\|_{\mathbf{V}_i} \le \beta_i^\circ \|\boldsymbol{\phi}(x,k)\|_{\mathbf{V}_i^{-1}}.$$

The first inequality follows from Cauchy-Schwarz, and the second follows from the definition of $\mathcal{C}_i$. Since μ is increasing and $p(x,k) \in [\underline{p}, \overline{p}]$, we obtain $p(x,k) \in \mathcal{I}_i(x,k)$. Also $p(x,k) \in \mathcal{I}_0(x,k) = [\underline{p}, \overline{p}]$. Therefore,

$$p(x,k) \in \bigcap_{i=\tau_{m_t}-1}^{t-1} \mathcal{I}_i(x,k) = [\underline{p}_t(x,k), \overline{p}_t(x,k)].$$

In particular, none of these intersections is empty.

If $Q(t) > 0$, the pair (x_t^*, k_t^*) satisfies

$$\overline{p}_t(x_t^*, k_t^*) \ge p(x_t^*, k_t^*) \ge \max_{(x,k) \in \mathcal{X}_t \times [K]} \underline{p}_t(x,k).$$

Thus $(x_t^*, k_t^*) \in \mathcal{Z}_t$, and

$$\begin{aligned} g_t &\le \overline{p}_t(x_t^*, k_t^*) - \underline{p}_t(x_t, k_t) \\ &= w_t(x_t^*, k_t^*) + \underline{p}_t(x_t^*, k_t^*) - \underline{p}_t(x_t, k_t) \\ &\le w_t(x_t^*, k_t^*) + \overline{p}_t(x_t, k_t) - \underline{p}_t(x_t, k_t) \le 2w_t, \end{aligned}$$

where the first inequality uses the interval inclusion, the second uses $(x_t, k_t) \in \mathcal{Z}_t$, and the third uses widest interval selection in Algorithm 1. For $Q(t) = 0$, $g_t = w_t = 0$. □

Lemma 3. *For $1 \le t_1 \le t_2$, $0 < r \le 1$, and $\ell = t_2 - t_1 + 1$,*

$$\sum_{i=t_1}^{t_2} \mathbf{1}\{w_i > r\} \le 1 + \frac{d\beta_{t_2}^2}{2r^2} \log(1 + \ell/d)$$

on every sample path.

B.1.2 Proof Of Lemma 3

Proof. Within an epoch, Algorithm 1 intersects the current interval with $\mathcal{I}_{i-1}(x,k)$. If the intersection is empty, it keeps the midpoint of the previous interval until the next epoch. Thus the intervals are nested on every sample path. When their endpoints differ, they are contained in $\mathcal{I}_{i-1}(x,k)$. Since $\dot{\mu} \le 1/4$, for $i \ge 2$,

$$\begin{aligned} w_i(x,k) &\le \mu\Big(\langle \boldsymbol{\phi}(x,k), \widehat{\boldsymbol{\theta}}_{i-1} \rangle + \beta_{i-1}^\circ \|\boldsymbol{\phi}(x,k)\|_{\mathbf{V}_{i-1}^{-1}}\Big) \\ &\quad - \mu\Big(\langle \boldsymbol{\phi}(x,k), \widehat{\boldsymbol{\theta}}_{i-1} \rangle - \beta_{i-1}^\circ \|\boldsymbol{\phi}(x,k)\|_{\mathbf{V}_{i-1}^{-1}}\Big) \\ &\le \frac{\beta_{i-1}^\circ}{2} \|\boldsymbol{\phi}(x,k)\|_{\mathbf{V}_{i-1}^{-1}}. \end{aligned} \tag{16}$$

The same bound holds if $w_i(x,k) = 0$.

Set $\widetilde{\mathbf{V}}_{t_1-1} = \mathbf{I}$ and $\widetilde{\mathbf{V}}_i = \mathbf{I} + \sum_{j=t_1}^{i} \boldsymbol{\phi}_j \boldsymbol{\phi}_j^\top$. By Lemma 11 of Abbasi-Yadkori et al. (2011) (Lemma 14),

$$\sum_{i=t_1}^{t_2} \|\boldsymbol{\phi}_i\|_{\mathbf{V}_{i-1}^{-1}}^2 \le \sum_{i=t_1}^{t_2} \min\{1, \|\boldsymbol{\phi}_i\|_{\widetilde{\mathbf{V}}_{i-1}^{-1}}^2\} \le 2 \log\det \widetilde{\mathbf{V}}_{t_2} \le 2d \log(1 + \ell/d). \tag{17}$$

The first inequality follows from $\mathbf{V}_{i-1} \succeq \widetilde{\mathbf{V}}_{i-1} \succeq \mathbf{I}$ and $\|\boldsymbol{\phi}_i\|_2 \le 1$. The second follows from Lemma 14, and the third follows from the arithmetic-geometric mean inequality and $\operatorname{tr}(\widetilde{\mathbf{V}}_{t_2}) \le d + \ell$. Combining (16) and (17) gives

$$\begin{aligned} \sum_{i=t_1}^{t_2} \mathbf{1}\{w_i > r\} &\le 1 + \frac{\beta_{t_2}^2}{4r^2} \sum_{i=\max\{t_1,2\}}^{t_2} \|\boldsymbol{\phi}_i\|_{\mathbf{V}_{i-1}^{-1}}^2 \\ &\le 1 + \frac{d\beta_{t_2}^2}{2r^2} \log(1 + \ell/d), \end{aligned}$$

where the first inequality also uses $\beta_{i-1}^\circ \le \beta_{i-1} \le \beta_{t_2}$. The additive one bounds the indicator at $i = 1$ if that index is present. □

B.2 Proofs Of Lemmas 4 And 5

Lemma 4. *Under any policy,* $|W(i+1) - W(i)| \le 1$, *and*

$$\mathbb{E}[W(i+1) - W(i) \mid \mathcal{F}_i] \le -\epsilon \underline{p} + g_i \qquad \text{if } Q(i) > 0.$$

Under Algorithm 1, fix $1 \le t_1 \le t_2$, $\ell = t_2 - t_1 + 1$, *and* $\delta \in (0,1)$. *With conditional probability at least* $1-\delta$ *given* $\mathcal{F}_{t_1}$, *if* $Q(i) > 0$ *and the waiting intervals contain their true service probabilities for every* $i \in [t_1, t_2]$, *then*

$$W(t_2+1) - W(t_1) \le -\epsilon \underline{p}\, \ell/2 + 1 + \frac{8d\beta_{t_2}^2}{\epsilon^2 \underline{p}^2} \log(1 + \ell/d) + \sqrt{8\ell \log(1/\delta)}.$$

B.2.1 Proof Of Lemma 4

Proof. Write $\Delta W(i) = W(i+1) - W(i)$. By the definition of $W(i)$,

$$\Delta W(i) = \frac{\underline{p} A(i)}{p^*(X_i)} - \frac{\underline{p} D(i)}{p^*(x_i)}, \qquad |\Delta W(i)| \le 1,$$

where the departure term is zero when $Q(i) = 0$. The inequality follows because each fraction belongs to $[0,1]$. When $Q(i) > 0$,

$$\begin{aligned}
\mathbb{E}[\Delta W(i) \mid \mathcal{F}_i] &= \lambda \underline{p}\, \mathbb{E}_{X \sim \mathcal{D}}[1/p^*(X)] - \underline{p}\, p(x_i, k_i)/p^*(x_i) \\
&= -\epsilon \underline{p} + \frac{\underline{p}}{p^*(x_i)} \big(p^*(x_i) - p(x_i, k_i)\big) \\
&\le -\epsilon \underline{p} + g_i \le -\epsilon \underline{p}/2 + \mathbf{1}\{g_i > \epsilon \underline{p}/2\}.
\end{aligned}$$

The first equality follows from the arrival and service laws, and the second follows from Assumption 2. The first inequality uses $\underline{p} \le p^*(x_i) \le p(x_i^*, k_i^*)$, and the second uses $0 \le g_i \le 1$.

If the intervals contain their true service probabilities, then $g_i \le 2w_i$ by Lemma 2. Therefore, on the event in the statement,

$$\begin{aligned}
\sum_{i=t_1}^{t_2} \mathbb{E}[\Delta W(i) \mid \mathcal{F}_i] &\le -\epsilon \underline{p}\, \ell/2 + \sum_{i=t_1}^{t_2} \mathbf{1}\{w_i > \epsilon \underline{p}/4\} \\
&\le -\epsilon \underline{p}\, \ell/2 + 1 + \frac{8d\beta_{t_2}^2}{\epsilon^2 \underline{p}^2} \log(1 + \ell/d), \qquad (18)
\end{aligned}$$

where the second inequality follows from Lemma 3 with $r = \epsilon \underline{p}/4$. Since $|\Delta W(i) - \mathbb{E}[\Delta W(i) \mid \mathcal{F}_i]| \le 2$, the Azuma-Hoeffding inequality (Lemma 16) gives

$$\mathbb{P}\left(\sum_{i=t_1}^{t_2} \big(\Delta W(i) - \mathbb{E}[\Delta W(i) \mid \mathcal{F}_i]\big) > \sqrt{8\ell \log(1/\delta)} \,\middle|\, \mathcal{F}_{t_1} \right) \le \delta.$$

Adding this sum to (18) proves the result. The concentration inequality is conditional only on $\mathcal{F}_{t_1}$, not on the interval-validity or nonempty-queue events. □

Lemma 5. *Let* $\ell_{0,t}$ *be defined in (12). For all sufficiently large* t, *every* $t_2 \in [t/3, t]$ *satisfies*

$$\begin{aligned}
\mathbb{P}(Q(i) > 0 \text{ for all } i \in [t_2 - \ell_{0,t}, t_2] \mid U_0) &\le C_5 (t+1)^{-2}, \\
\mathbb{P}(W(t_2) > \ell_{0,t} \mid U_0) &\le C_5 (t+1)^{-2}, \qquad (19) \\
\mathbb{E}[W(t_2) \mid U_0] &\le \ell_{0,t} + C_5 (t+1)^{-1}.
\end{aligned}$$

The constants and the threshold on t *are uniform in* U_0.

B.2.2 Proof Of Lemma 5

Proof. We apply Lemma 4 first to obtain an empty round before $t_2/2$, and then to bound the probability of $\ell_{0,t}$ consecutive nonempty rounds after an empty round. Fix U_0, and set

$$t_1 = \lfloor \epsilon \underline{p}\, t_2/64 \rfloor, \qquad t_3 = \lfloor t_2/2 \rfloor, \qquad \mathcal{E}_{0,t} = \bigcap_{j=\lfloor \epsilon \underline{p}\, t/1024 \rfloor}^{t} \{\boldsymbol{\theta}^* \in \mathcal{C}_j\}.$$

Take t large enough that the lower indices are positive and the polynomial bound in Lemma 2 applies. Then

$$\mathbb{P}(\mathcal{E}_{0,t}^{\mathrm{c}} \mid U_0) \le C_5 (t+1)^{-2}. \tag{20}$$

This requires $\lambda t \ge C_5 \log(t+1)$, where C_5 depends only on ϵ and $\underline{p}$. By (10), for $i \ge t_1$ and sufficiently large t,

$$\tau_{m_i} - 1 \ge i/2 - 1 \ge \epsilon \underline{p}\, t_2/128 - 3/2 \ge \epsilon \underline{p}\, t/1024.$$

Thus Lemma 2 implies that, on $\mathcal{E}_{0,t}$, all intervals used in rounds $t_1, \dots, t$ contain their true service probabilities.

Increase the threshold on t so that

$$1 + \frac{8 d \beta_t^2}{\epsilon^2 \underline{p}^2} \log(t+1) \le \epsilon \underline{p}\, t_2/64, \qquad t_3 - t_1 \ge 3t_2/8, \qquad \ell_{0,t} \le t/12.$$

Since $W(t_1) \le t_1 \le \epsilon \underline{p}\, t_2/64$, applying Lemma 4 to $[t_1, t_3 - 1]$ with $\delta = \exp(-\epsilon^2 \underline{p}^2 t_2/512)$ gives, on $\mathcal{E}_{0,t} \cap \{Q(i) > 0 \text{ for all } i \in [t_1, t_3]\}$,

$$\begin{aligned} W(t_3) &\le W(t_1) - \frac{\epsilon \underline{p}}{2}(t_3 - t_1) + 1 + \frac{8 d \beta_t^2}{\epsilon^2 \underline{p}^2} \log(t+1) + \frac{\epsilon \underline{p}}{8} \sqrt{(t_3 - t_1) t_2} \\ &\le \frac{\epsilon \underline{p}\, t_2}{64} - \frac{3 \epsilon \underline{p}\, t_2}{16} + \frac{\epsilon \underline{p}\, t_2}{64} + \frac{\epsilon \underline{p}\, t_2}{8} = -\frac{\epsilon \underline{p}\, t_2}{32} < 0, \end{aligned}$$

except on an event of probability at most δ conditional on U_0. The second inequality also uses $t_3 - t_1 \le t_2$. Since $W(t_3) \ge 0$, we conclude that

$$\mathbb{P}(\mathcal{E}_{0,t} \cap \{Q(i) > 0 \text{ for all } i \in [t_1, t_3]\} \mid U_0) \le e^{-\epsilon^2 \underline{p}^2 t_2/512}. \tag{21}$$

For each $j \in [t_1, t_2 - \ell_{0,t}]$, on $\{Q(j) = 0\}$ we have $W(j+1) \le 1$. We apply Lemma 4 to $[j+1, j + \ell_{0,t} - 1]$ with $\delta = (t+1)^{-3}$. First, the definition $C_2 = 64/(\epsilon \underline{p})^3$ in (12) gives

$$\begin{aligned} 1 + \frac{8 d \beta_t^2}{\epsilon^2 \underline{p}^2} \log(1 + \ell_{0,t}/d) &\le 1 + \frac{8 d \beta_t^2}{\epsilon^2 \underline{p}^2} (\log(t+1) - \log 2) \\ &\le \frac{8 d \beta_t^2}{\epsilon^2 \underline{p}^2} \log(t+1) \le \frac{\epsilon \underline{p}\, \ell_{0,t}}{8}, \\ \ell_{0,t} \ge \frac{384 \log(t+1)}{\epsilon^2 \underline{p}^2}, &\qquad \ell_{0,t} \ge \frac{16}{\epsilon \underline{p}}. \end{aligned}$$

The first inequality uses $\ell_{0,t} \le t/12$ and $d \ge 1$. The second uses $8 d \beta_t^2 \log 2 \ge \epsilon^2 \underline{p}^2$. These inequalities and the last two bounds hold for sufficiently large t, since (9) and $\kappa \ge 4$ give $d \beta_t^2 \ge 16 \log(t+1)$. Consequently, Lemma 4 gives, on $\mathcal{E}_{0,t} \cap \{Q(j) = 0,\ Q(i) > 0 \text{ for all } i \in [j+1, j + \ell_{0,t}]\}$,

$$\begin{aligned} W(j + \ell_{0,t}) &\le 1 - \frac{\epsilon \underline{p}}{2}(\ell_{0,t} - 1) + \frac{\epsilon \underline{p}\, \ell_{0,t}}{8} + \sqrt{24 (\ell_{0,t} - 1) \log(t+1)} \\ &\le 1 - \frac{\epsilon \underline{p}}{2}(\ell_{0,t} - 1) + \frac{\epsilon \underline{p}\, \ell_{0,t}}{8} + \frac{\epsilon \underline{p}\, \ell_{0,t}}{4} \\ &= 1 + \epsilon \underline{p}/2 - \epsilon \underline{p}\, \ell_{0,t}/8 < 0, \end{aligned}$$

except on an event of probability at most $(t+1)^{-3}$ conditional on $\mathcal{F}_{j+1}$. Since $W(j+\ell_{0,t}) \geq 0$, taking expectations conditional on U_0 yields

$$\mathbb{P}(\mathcal{E}_{0,t} \cap \{Q(j) = 0,\ Q(i) > 0 \text{ for all } i \in [j+1, j+\ell_{0,t}]\} \mid U_0) \leq (t+1)^{-3}. \tag{22}$$

If the queue is empty in $[t_1, t_3]$ and nonempty throughout $[t_2 - \ell_{0,t}, t_2]$, its last empty round before $t_2 - \ell_{0,t}$ is some $j \in [t_1, t_2 - \ell_{0,t}]$ satisfying $Q(i) > 0$ for all $i \in [j+1, j+\ell_{0,t}]$. Since $t_3 \leq t_2 - \ell_{0,t}$, the union bound and (20)-(22) give

$$\begin{aligned}&\mathbb{P}(Q(i) > 0 \text{ for all } i \in [t_2 - \ell_{0,t}, t_2] \mid U_0) \\ &\quad \leq C_5(t+1)^{-2} + e^{-\epsilon^2 \underline{p}^2 t_2/512} + t(t+1)^{-3} \leq C_5(t+1)^{-2},\end{aligned}$$

where the last inequality holds for sufficiently large t because $t_2 \geq t/3$. The union is over deterministic indices j, so no conditioning on a last empty round is needed.

If $Q(j) = 0$ for some $j \in [t_2 - \ell_{0,t}, t_2]$, then

$$W(t_2) = W(j) + \sum_{i=j}^{t_2-1} \Delta W(i) \leq t_2 - j \leq \ell_{0,t}.$$

This proves (19). Finally, $W(t_2) \leq t_2 \leq t$ gives

$$\mathbb{E}[W(t_2) \mid U_0] \leq \ell_{0,t} + t\mathbb{P}(W(t_2) > \ell_{0,t} \mid U_0) \leq \ell_{0,t} + C_5(t+1)^{-1}.$$

All bounds hold uniformly in the fixed value of U_0. □

B.3 Proofs Of Lemmas 6 To 8

Recall that $\Delta W(i) = W(i+1) - W(i)$ and $\mathcal{G}_i = \mathcal{F}_i \vee \sigma(D(i))$. The arrival $(A(i), X_i)$ is independent of $\mathcal{G}_i$. For $r > 0$, define

$$z_i(r) = \mathbb{P}_{X \sim \mathcal{D}}\left(\exists k \in [K] :\ \overline{p}_i(X,k) \geq \max_{k'} \underline{p}_i(X,k'), \quad w_i(X,k) > r\right). \tag{23}$$

The interval functions are held fixed in this probability.

Lemma 6. *Fix $1 \leq t_1 \leq t_2$, $\ell = t_2 - t_1 + 1$, and $0 < r \leq 1$. Suppose the intervals are nonempty and $\mathcal{F}_i$-measurable at round i, are nested on $[t_1, t_2+1]$, and remain unchanged on empty rounds. Suppose a job arriving after an empty round is served by a candidate server with the widest interval, and the selected widths satisfy Lemma 3. Then*

$$\mathbb{E}[z_{t_2}(r)] \leq \frac{4}{\epsilon \underline{p}\, \ell}\left[\frac{1}{\lambda}\left(1 + \frac{d\beta_{t_2+1}^2}{2r^2}\log(1+\ell/d)\right) + \mathbb{E}[W(t_1)] + 2\mathbb{E}\left[\sum_{i=t_1}^{t_2} \mathbf{1}\{g_i > \epsilon \underline{p}/2\}\right]\right].$$

B.3.1 Proof Of Lemma 6

Proof. We first bound $\mathbb{E}[\sum_{i=t_1}^{t_2} z_i(r)\mathbf{1}\{Q(i) = 0\}]$ by Lemma 3. We then use Lemma 4 to bound $\mathbb{E}[z_{t_2}(r)]$. The variables $z_i(r)$ are $\mathcal{F}_i$-measurable. They are also nonincreasing. Indeed, if x, k satisfy the inequalities in (23) at $i+1$, then

$$\begin{aligned}&\overline{p}_i(x,k) \geq \overline{p}_{i+1}(x,k) \geq \max_{k'} \underline{p}_{i+1}(x,k') \geq \max_{k'} \underline{p}_i(x,k'), \\ &w_i(x,k) \geq w_{i+1}(x,k) > r.\end{aligned}$$

Thus they also satisfy the inequalities at i, and $\ell z_{t_2}(r) \leq \sum_{i=t_1}^{t_2} z_i(r)$.

On $\{Q(i) = 0, A(i) = 1\}$, the job X_i is alone in round $i+1$. Since the intervals are unchanged and the widest candidate is chosen,

$$\mathbb{E}[\mathbf{1}\{w_{i+1} > r\} \mid \mathcal{F}_i] \geq \lambda z_i(r)\mathbf{1}\{Q(i) = 0\}.$$

Taking expectations and summing gives

$$\lambda\mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r)\mathbf{1}\{Q(i) = 0\}\right] \leq \mathbb{E}\left[\sum_{i=t_1+1}^{t_2+1} \mathbf{1}\{w_i > r\}\right] \leq 1 + \frac{d\beta_{t_2+1}^2}{2r^2}\log(1+\ell/d), \tag{24}$$

where the second inequality follows from Lemma 3.

By Lemma 4 when $Q(i) > 0$ and by $\mathbb{E}[\Delta W(i) \mid \mathcal{F}_i] = (1-\epsilon)\underline{p}$ when $Q(i) = 0$,

$$\begin{aligned}\mathbb{E}[\Delta W(i) \mid \mathcal{F}_i] &\le -\epsilon\underline{p} + g_i + \underline{p}\mathbf{1}\{Q(i) = 0\} \\ &\le -\epsilon\underline{p}/4 + \mathbf{1}\{Q(i) = 0\} + 2\mathbf{1}\{g_i > \epsilon\underline{p}/2\}.\end{aligned}$$

The second inequality uses $g_i \le \epsilon\underline{p}/2 + \mathbf{1}\{g_i > \epsilon\underline{p}/2\}$ and $\underline{p} \le 1$. Multiply by $z_i(r)$, sum, and take expectations. Since $z_i(r)$ is $\mathcal{F}_i$-measurable, the tower property gives

$$\begin{aligned}\frac{\epsilon\underline{p}\,\ell}{4}\mathbb{E}[z_{t_2}(r)] &\le \frac{\epsilon\underline{p}}{4}\mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r)\right] \\ &\le \mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r)\mathbf{1}\{Q(i) = 0\}\right] - \mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r)\Delta W(i)\right] \\ &\quad + 2\mathbb{E}\left[\sum_{i=t_1}^{t_2} z_i(r)\mathbf{1}\{g_i > \epsilon\underline{p}/2\}\right]. \end{aligned} \tag{25}$$

For the second term on the right-hand side, summation by parts gives

$$\begin{aligned}-\sum_{i=t_1}^{t_2} z_i(r)\Delta W(i) &= z_{t_1}(r)W(t_1) - z_{t_2}(r)W(t_2+1) - \sum_{i=t_1+1}^{t_2}(z_{i-1}(r) - z_i(r))W(i) \\ &\le W(t_1),\end{aligned}$$

where the inequality follows from $0 \le z_i(r) \le 1$, $z_{i-1}(r) \ge z_i(r)$, and $W(i) \ge 0$. Substituting this inequality and (24) into (25), and using $z_i(r) \le 1$ in its last term, we obtain

$$\frac{\epsilon\underline{p}\,\ell}{4}\mathbb{E}[z_{t_2}(r)] \le \frac{1}{\lambda}\left(1 + \frac{d\beta_{t_2+1}^2}{2r^2}\log(1+\ell/d)\right) + \mathbb{E}[W(t_1)] + 2\mathbb{E}\left[\sum_{i=t_1}^{t_2}\mathbf{1}\{g_i > \epsilon\underline{p}/2\}\right].$$

Dividing by $\epsilon\underline{p}\,\ell/4$ proves the result. □

For use in the next two proofs, define

$$\mathcal{E}_{1,i} = \bigcap_{j=\tau_{m_i}-1}^{i-1}\{\boldsymbol{\theta}^* \in \mathcal{C}_j\}.$$

Lemma 7. *For all sufficiently large t, every $0 < r \le 1$, every $2 \le \ell \le t/8$, and every event $\mathcal{E}$,*

$$\mathbb{P}(w_t > r, \mathcal{E}) \le \frac{C_5\ell\, d\beta_t^2 \log^2(t+1)}{tr^2} + \mathbb{P}\big(\mathcal{E},\ \textit{the job selected at } t \textit{ arrived before } t-\ell\big).$$

The selected-job event is false when $Q(t) = 0$.

B.3.2 Proof Of Lemma 7

Proof. We apply Lemma 6 within the epoch containing t, and then take a union over possible arrival times of the selected job. Fix U_0, set $m = m_t$, and let $\tau_m < i \le t$. The intervals are nested on $[\tau_m, i]$. On an empty round j, Algorithm 1 does not update $\widehat{\boldsymbol{\theta}}_j$, $\mathbf{V}_j$, or β_j°, so $\mathcal{I}_j(x,k) = \mathcal{I}_{j-1}(x,k)$. The intersected intervals therefore remain unchanged.

By Lemma 2,

$$\mathbf{1}\{g_j > \epsilon\underline{p}/2\} \le \mathbf{1}\{w_j > \epsilon\underline{p}/4\} + \mathbf{1}\{\mathcal{E}_{1,j}^{\mathrm{c}}\}.$$

Since all indices $j \in [\tau_m, i-1]$ are in epoch m,

$$\begin{aligned}\mathbb{E}\left[\sum_{j=\tau_m}^{i-1}\mathbf{1}\{g_j > \epsilon\underline{p}/2\}\,\middle|\,U_0\right] &\le 1 + \frac{8d\beta_t^2}{\epsilon^2\underline{p}^2}\log(t+1) + t\mathbb{P}(\exists j \ge \tau_m - 1 : \boldsymbol{\theta}^* \notin \mathcal{C}_j \mid U_0) \\ &\le C_5 d\beta_t^2\log(t+1).\end{aligned} \tag{26}$$

The first inequality follows from Lemma 3. The second follows from Lemma 2 and $\tau_m > t/2$, which give a probability of at most $5\tau_m^{-2}$ for sufficiently large t. Also, Lemma 5 gives

$$\mathbb{E}[W(\tau_m) \mid U_0] \le \ell_{0,t} + C_5(t+1)^{-1} \le C_5 d\beta_t^2 \log(t+1).$$

Apply Lemma 6 on $[\tau_m, i-1]$. Since $z_i(r) \le z_{i-1}(r)$, we have

$$\begin{aligned}\mathbb{E}[z_i(r) \mid U_0] &\le \frac{4}{\epsilon \underline{p}(i-\tau_m)}\left[\frac{1}{\lambda}\left(1 + \frac{d\beta_i^2}{2r^2}\log(1+(i-\tau_m)/d)\right) + \mathbb{E}[W(\tau_m) \mid U_0]\right. \\ &\qquad \left. +2\mathbb{E}\left[\sum_{j=\tau_m}^{i-1} \mathbf{1}\{g_j > \epsilon\underline{p}/2\} \,\middle|\, U_0\right]\right] \\ &\le \frac{C_5 d\beta_t^2 \log(t+1)}{\lambda r^2 (i-\tau_m)}, \end{aligned} \tag{27}$$

where the last inequality follows from (26), the bound on $\mathbb{E}[W(\tau_m) \mid U_0]$, $\beta_i \le \beta_t$, and $r, \lambda \le 1$.

Let U be the fractional part of $\log_2 t - U_0$. Then U is uniform on $[0,1)$, and (10) gives

$$\tau_{m_t} = \lceil t2^{-U} \rceil, \qquad 0 < t - \tau_{m_t} + 1 - t(1-2^{-U}) \le 1.$$

The variable $Y = 1 - 2^{-U}$ has density $1/((1-y)\log 2) \le 2/\log 2$ for $0 \le y \le 1/2$. Consequently,

$$\mathbb{P}_{U_0}(t - \tau_{m_t} + 1 \le 2\ell) \le \mathbb{P}(tY \le 2\ell) \le \frac{4\ell}{t \log 2}, \tag{28}$$

$$\mathbb{E}_{U_0}\left[\frac{\mathbf{1}\{t - \tau_{m_t} + 1 > 2\ell\}}{t - \tau_{m_t} + 1 - \ell}\right] \le \frac{4}{t\log 2}\int_{(2\ell-1)/t}^{1/2} \frac{dy}{y} \le \frac{4\log(t+1)}{t\log 2}. \tag{29}$$

For the first inequality in (29), $t - \tau_{m_t} + 1 > 2\ell$ implies $tY > 2\ell - 1$ and $t - \tau_{m_t} + 1 - \ell \ge tY/2$.

Suppose the selected job arrived in round $i-1$ with $t - \ell + 1 \le i \le t$, and $t - \tau_{m_t} + 1 > 2\ell$. Then $i > \tau_{m_t}$. If $w_t > r$, nesting gives

$$\begin{aligned}&\overline{p}_i(x_t, k_t) \ge \overline{p}_t(x_t, k_t) \ge \max_{k'} \underline{p}_t(x_t, k') \ge \max_{k'} \underline{p}_i(x_t, k'), \\ &w_i(x_t, k_t) \ge w_t > r.\end{aligned}$$

The second inequality in the first line follows from $(x_t, k_t) \in \mathcal{Z}_t$. Thus $X_{i-1} = x_t$ satisfies the inequalities in (23) at i.

The interval functions at round i are $\mathcal{G}_{i-1}$-measurable, and U_0 is included in $\mathcal{G}_{i-1}$. Independence of $(A(i-1), X_{i-1})$ from $\mathcal{G}_{i-1}$ gives

$$\begin{aligned}&\mathbb{P}\big(A(i-1) = 1,\ X_{i-1} \text{ satisfies (23)},\ t - \tau_{m_t} + 1 > 2\ell\big) \\ &\quad = \lambda \mathbb{E}[z_i(r)\mathbf{1}\{t - \tau_{m_t} + 1 > 2\ell\}] \\ &\quad \le \frac{C_5 d\beta_t^2 \log(t+1)}{r^2}\mathbb{E}_{U_0}\left[\frac{\mathbf{1}\{t - \tau_{m_t} + 1 > 2\ell\}}{t - \tau_{m_t} + 1 - \ell}\right], \end{aligned} \tag{30}$$

where the inequality follows from (27) and $i - \tau_{m_t} \ge t - \tau_{m_t} + 1 - \ell$. The union bound over $i = t - \ell + 1, \dots, t$ and (28)-(30) give

$$\begin{aligned}&\mathbb{P}\big(w_t > r,\ \text{the selected job arrived at or after } t - \ell\big) \\ &\quad \le \frac{4\ell}{t\log 2} + \frac{C_5 \ell\, d\beta_t^2 \log^2(t+1)}{tr^2} \le \frac{C_5 \ell\, d\beta_t^2 \log^2(t+1)}{tr^2}.\end{aligned}$$

The last inequality uses $d\beta_t^2 \log(t+1)/r^2 \ge 1$ for sufficiently large t. Intersecting with $\mathcal{E}$ and adding the probability that the selected job arrived before $t - \ell$ proves the result. □

Lemma 8. *For all sufficiently large t and every $z \ge 0$,*

$$\mathbb{P}(W(t) > z) \le C_5 e^{-\epsilon \underline{p} z/2} + \frac{C_5 (d\beta_t^2)^3 \log^4(t+1)}{t}.$$

B.3.3 Proof Of Lemma 8

Proof. We first bound $W(t)$ on an event where $g_i \le \epsilon\underline{p}/2$ after a recent empty round. We also derive two probability bounds used in the proof of Lemma 1.

By Lemmas 2, 5 and 7,

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{1,i}^{\mathrm{c}} \text{ or } w_i > \epsilon\underline{p}/4) &\le C_5(i+1)^{-2} + \frac{C_5\ell_{0,i}d\beta_i^2\log^2(i+1)}{i} \\ &\quad + \mathbb{P}(Q(j) > 0 \text{ for all } j \in [i-\ell_{0,i}, i]) \\ &\le \frac{C_5(d\beta_i^2)^2\log^3(i+1)}{i}. \end{aligned} \tag{31}$$

For the first inequality, Lemma 7 is applied with $r = \epsilon\underline{p}/4$ and $\ell = \ell_{0,i}$. If the selected job arrived before $i - \ell_{0,i}$, then the queue was nonempty throughout $[i-\ell_{0,i}, i]$. The second inequality follows from Lemma 5 and (12).

Set $t_1 = t - \ell_{0,t} - \ell_{1,t}$ and $t_2 = t - \ell_{1,t}$, and define

$$\mathcal{E}_{2,t} = \{Q(j) = 0 \text{ for some } j \in [t_1, t_2]\} \cap \bigcap_{i=t_1}^{t} (\mathcal{E}_{1,i} \cap \{w_i \le \epsilon\underline{p}/4\}).$$

For sufficiently large t, we have $t_1 \ge 3t/4$. By the union bound,

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{2,t}^{\mathrm{c}}) &\le \mathbb{P}(Q(j) > 0 \text{ for all } j \in [t_1, t_2]) + \sum_{i=t_1}^{t} \mathbb{P}(\mathcal{E}_{1,i}^{\mathrm{c}} \text{ or } w_i > \epsilon\underline{p}/4) \\ &\le C_5(t+1)^{-2} + \frac{C_5(\ell_{0,t}+\ell_{1,t}+1)(d\beta_t^2)^2\log^3(t+1)}{t} \\ &\le \frac{C_5(d\beta_t^2)^3\log^4(t+1)}{t}. \end{aligned} \tag{32}$$

The second inequality follows from Lemma 5 at terminal round t_2, $\ell_{0,t_2} \le \ell_{0,t}$, and (31). The third follows from (12). On $\mathcal{E}_{2,t}$, every nonempty round $i \in [t_1, t]$ satisfies $\mathbb{E}[\Delta W(i) \mid \mathcal{F}_i] \le -\epsilon\underline{p}/2$ by Lemmas 2 and 4.

For each fixed j, define the first failure round after j by

$$\tau_{\mathrm{f},j} = \inf\{i \ge j : \mathcal{E}_{1,i}^{\mathrm{c}} \text{ occurs or } w_i > \epsilon\underline{p}/4\}, \qquad Z_i = e^{\epsilon\underline{p}W(i)/2}\mathbf{1}\{\tau_{\mathrm{f},j} > i\} \quad (i \ge j).$$

On $\{W(i) > 0, \tau_{\mathrm{f},j} > i\}$, Hoeffding's lemma (Lemma 15) gives

$$\mathbb{E}[e^{\epsilon\underline{p}\Delta W(i)/2} \mid \mathcal{F}_i] \le \exp\left(\frac{\epsilon\underline{p}}{2}\mathbb{E}[\Delta W(i) \mid \mathcal{F}_i] + \frac{\epsilon^2\underline{p}^2}{8}\right) \le e^{-\epsilon^2\underline{p}^2/8}.$$

The first inequality uses $|\Delta W(i)| \le 1$, and the second follows from Lemmas 2 and 4. On $\{W(i) = 0\}$, we have $W(i+1) \le 1$. Thus

$$\mathbb{E}[Z_{i+1} \mid \mathcal{F}_i] \le e^{-\epsilon^2\underline{p}^2/8}Z_i + e^{\epsilon\underline{p}/2}.$$

On $\{Q(j) = 0\}$, $Z_j \le 1$, so iteration gives, for every $t_3 \ge j$,

$$\begin{aligned}\mathbb{E}[Z_{t_3} \mid \mathcal{F}_j] &\le e^{-\epsilon^2\underline{p}^2(t_3-j)/8} + e^{\epsilon\underline{p}/2}\sum_{i=0}^{t_3-j-1} e^{-\epsilon^2\underline{p}^2 i/8} \\ &\le 1 + \frac{e^{\epsilon\underline{p}/2}}{1 - e^{-\epsilon^2\underline{p}^2/8}} \le \frac{16}{\epsilon^2\underline{p}^2}. \end{aligned}$$

The last inequality follows from $1 - e^{-z} \ge z/(1+z)$, $\epsilon\underline{p} \le 1$, and $e^{1/2} < 5/3$. Markov's inequality gives, on $\{Q(j) = 0\}$,

$$\mathbb{P}(W(t_3) > z, \tau_{\mathrm{f},j} > t_3 \mid \mathcal{F}_j) \le \frac{16}{\epsilon^2\underline{p}^2}e^{-\epsilon\underline{p}z/2}. \tag{33}$$

For $j \in [t_1, t_2]$, let $\mathcal{E}_{3,j} = \{Q(j) = 0,\ Q(i) > 0 \text{ for all } i \in [t_1, j-1]\}$. These events are disjoint and $\mathcal{F}_j$-measurable. For every $t_3 \in [t_2, t]$, $\mathcal{E}_{2,t} \cap \mathcal{E}_{3,j}$ implies $\tau_{\mathrm{f},j} > t_3$. Hence

$$\begin{aligned}\mathbb{P}(W(t_3) > z, \mathcal{E}_{2,t}) &\le \sum_{j=t_1}^{t_2} \mathbb{E}[\mathbf{1}\{\mathcal{E}_{3,j}\}\mathbb{P}(W(t_3) > z, \tau_{\mathrm{f},j} > t_3 \mid \mathcal{F}_j)] \\ &\le \frac{16}{\epsilon^2 \underline{p}^2} e^{-\epsilon \underline{p} z/2} \sum_{j=t_1}^{t_2} \mathbb{P}(\mathcal{E}_{3,j}) \le \frac{16}{\epsilon^2 \underline{p}^2} e^{-\epsilon \underline{p} z/2}, \end{aligned} \tag{34}$$

where the second inequality follows from (33).

We next bound the probability that the selected job arrived before $t - \ell$, for $2 \le \ell \le \ell_{1,t}$. Such a job keeps the queue nonempty throughout $[t-\ell, t]$. On $\mathcal{E}_{2,t} \cap \{W(t-\ell) \le \epsilon \underline{p}\, \ell/4\}$, this implies

$$\begin{aligned}\sum_{i=t-\ell}^{t-1} (\Delta W(i) - \mathbb{E}[\Delta W(i) \mid \mathcal{F}_i]) &= W(t) - W(t-\ell) - \sum_{i=t-\ell}^{t-1} \mathbb{E}[\Delta W(i) \mid \mathcal{F}_i] \\ &\ge -\epsilon \underline{p}\, \ell/4 + \epsilon \underline{p}\, \ell/2 = \epsilon \underline{p}\, \ell/4.\end{aligned}$$

Since $t - \ell \in [t_2, t]$, (34) and the Azuma-Hoeffding inequality (Lemma 16) give

$$\begin{aligned}&\mathbb{P}\big(\mathcal{E}_{2,t},\ \text{the job selected at } t \text{ arrived before } t-\ell\big) \\ &\quad\le \frac{16}{\epsilon^2 \underline{p}^2} e^{-\epsilon^2 \underline{p}^2 \ell/8} + e^{-\epsilon^2 \underline{p}^2 \ell/128} \le C_5 e^{-\ell/C_3},\end{aligned} \tag{35}$$

where $C_3 = 128/(\epsilon \underline{p})^2$. In particular, taking $\ell = \ell_{1,t} = \lceil C_3 \log(t+1) \rceil$ bounds (35) by $C_5 (t+1)^{-1}$.

We also bound $\mathbb{P}(Q(t) > 0, \mathcal{E}_{2,t})$. Fix $j \in [t_1, t-1]$ and condition on $\mathcal{F}_{j+1}$ with $Q(j) = 0$ and $A(j) = 1$. Then $W(j+1) \le 1$. If $Q(i) > 0$ for every $i \in [j+1, t]$ and $\mathcal{E}_{2,t}$ holds, then, for $t - j \ge 1 + 4/(\epsilon \underline{p})$,

$$\begin{aligned}\sum_{i=j+1}^{t-1} (\Delta W(i) - \mathbb{E}[\Delta W(i) \mid \mathcal{F}_i]) &\ge -1 + \epsilon \underline{p}(t-j-1)/2 \\ &\ge \epsilon \underline{p}(t-j-1)/4.\end{aligned}$$

The conditional bound in Lemma 16 gives probability at most $e^{-\epsilon^2 \underline{p}^2 (t-j-1)/128}$. For all $t - j \ge 1$, this is at most $C_5 e^{-(t-j)/C_5}$ after increasing C_5 to cover $t - j < 1 + 4/(\epsilon \underline{p})$. Since $A(j)$ is independent of $\mathcal{G}_j$, we obtain

$$\begin{aligned}&\mathbb{P}\big(Q(j) = 0, A(j) = 1, Q(i) > 0 \text{ for all } i \in [j+1, t], \mathcal{E}_{2,t}\big) \\ &\quad\le C_5 e^{-(t-j)/C_5} \mathbb{P}(Q(j) = 0, A(j) = 1) \le C_5 \lambda e^{-(t-j)/C_5}.\end{aligned}$$

On $\{Q(t) > 0\} \cap \mathcal{E}_{2,t}$, the last empty round before t is some $j \in [t_1, t-1]$ satisfying these conditions. Taking a union over all such indices gives

$$\mathbb{P}(Q(t) > 0, \mathcal{E}_{2,t}) \le C_5 \lambda \sum_{j=t_1}^{t-1} e^{-(t-j)/C_5} \le C_5 \lambda. \tag{36}$$

Finally, taking $t_3 = t$ in (34) and using (32), we obtain

$$\begin{aligned}\mathbb{P}(W(t) > z) &\le \mathbb{P}(W(t) > z, \mathcal{E}_{2,t}) + \mathbb{P}(\mathcal{E}_{2,t}^{\mathrm{c}}) \\ &\le \frac{16}{\epsilon^2 \underline{p}^2} e^{-\epsilon \underline{p} z/2} + \frac{C_5 (d \beta_t^2)^3 \log^4(t+1)}{t}.\end{aligned}$$

Increasing C_5 proves the result. □

B.4 Proof Of Lemma 1

Proof. We bound $\mathbb{P}(g_t > z, \mathcal{E}_{2,t})$ by choosing the interval length ℓ for each z, and then integrate over z. Here $\mathcal{E}_{2,t}$ and $\ell_{1,t}$ are defined in the proof of Lemma 8 and in (12). Throughout this proof, C_5 depends only on ϵ and $\underline{p}$.

Put

$$a_t = \frac{d\beta_t^2 \log^2(t+1)}{t}.$$

Take t sufficiently large that the preceding lemmas apply, $1/t \le a_t \le 1$, and $2 \le \ell_{1,t} \le t/8$. For every $\sqrt{a_t} \le z \le 1$, choose

$$\ell = \max\{2, \lceil C_3 \log(z^2/a_t) \rceil\}.$$

Since $z^2/a_t \le t$ and $\ell_{1,t} = \lceil C_3 \log(t+1) \rceil$, this choice satisfies $2 \le \ell \le \ell_{1,t} \le t/8$. Moreover,

$$\ell \le 2 + C_3 \log(z^2/a_t), \qquad e^{-\ell/C_3} \le a_t/z^2.$$

On $\mathcal{E}_{2,t}$, $g_t \le 2w_t$. Applying Lemma 7 with $r = z/2$ gives

$$\begin{aligned}
\mathbb{P}(g_t > z, \mathcal{E}_{2,t}) &\le \frac{4C_5 \ell\, d\beta_t^2 \log^2(t+1)}{tz^2} + \mathbb{P}\big(\mathcal{E}_{2,t},\ \text{the job selected at } t \text{ arrived before } t-\ell\big) \\
&\le \frac{4C_5 a_t \ell}{z^2} + C_5 e^{-\ell/C_3} \\
&\le \frac{C_5 a_t}{z^2}\big(1 + \log(z^2/a_t)\big). \qquad (37)
\end{aligned}$$

The second inequality follows from (35), which holds for every $2 \le \ell \le \ell_{1,t}$. The last inequality uses the two bounds on ℓ above and increases C_5 to include C_3. Also, since $g_t = 0$ when $Q(t) = 0$,

$$\mathbb{P}(g_t > z, \mathcal{E}_{2,t}) \le \mathbb{P}(Q(t) > 0, \mathcal{E}_{2,t}) \le C_5 \lambda \qquad (z > 0), \qquad (38)$$

where the last inequality follows from (36).

Set $z_0 = \sqrt{a_t \log(e/\lambda)/\lambda}$. If $z_0 \ge 1$, (38) gives

$$\mathbb{E}[g_t \mathbf{1}\{\mathcal{E}_{2,t}\}] = \int_0^1 \mathbb{P}(g_t > z, \mathcal{E}_{2,t})\, dz \le C_5 \lambda \le C_5 \sqrt{\lambda a_t \log(e/\lambda)}.$$

If $z_0 < 1$, then $z_0 \ge \sqrt{a_t}$ because $\log(e/\lambda) \ge 1 \ge \lambda$. Therefore,

$$\begin{aligned}
\mathbb{E}[g_t \mathbf{1}\{\mathcal{E}_{2,t}\}] &= \int_0^{z_0} \mathbb{P}(g_t > z, \mathcal{E}_{2,t})\, dz + \int_{z_0}^1 \mathbb{P}(g_t > z, \mathcal{E}_{2,t})\, dz \\
&\le C_5 \lambda z_0 + C_5 a_t \int_{z_0}^\infty \frac{1 + \log(z^2/a_t)}{z^2}\, dz \\
&= C_5 \lambda z_0 + \frac{C_5 a_t}{z_0}\big(3 + \log(z_0^2/a_t)\big) \\
&\le C_5 \sqrt{\lambda a_t \log(e/\lambda)}.
\end{aligned}$$

The first inequality follows from (37) and (38). The second equality uses $\int (1+\log(z^2/a_t)) z^{-2}\, dz = -(3 + \log(z^2/a_t))/z$. For the last inequality, the definition of z_0 gives

$$3 + \log(z_0^2/a_t) = 2 + \log\log(e/\lambda) + \log(e/\lambda) \le 3\log(e/\lambda).$$

Here we used $\log\log(e/\lambda) \le \log(e/\lambda) - 1$ and $\log(e/\lambda) \ge 1$, and increased C_5.

In both cases, $g_t \le 1$ and (32) imply

$$\begin{aligned}
\mathbb{E}[g_t] &\le \mathbb{E}[g_t \mathbf{1}\{\mathcal{E}_{2,t}\}] + \mathbb{P}(\mathcal{E}_{2,t}^{\mathrm{c}}) \\
&\le C_5 \beta_t \sqrt{\lambda d/t}\, \log(t+1) \sqrt{\log(e/\lambda)} + C_5 (d\beta_t^2)^3 \log^4(t+1)/t. \qquad (39)
\end{aligned}$$

Fix $C_5 \geq 1$ for which (39) holds. Increase the threshold on t so that

$$\frac{d\beta_t^2}{t} \leq \lambda, \qquad \log(t+1) \geq 4C_5^2 \log(e/\lambda),$$
$$t \geq \frac{4C_5^2 (d\beta_t^2)^5 \log^5(t+1)}{\lambda}. \tag{40}$$

These conditions can hold for every sufficiently large t because $\beta_t^2 = \mathcal{O}(\log(t+1))$ with the model parameters fixed. By (40),

$$C_5 \beta_t \sqrt{\lambda d/t}\, \log(t+1) \sqrt{\log(e/\lambda)} \leq \tfrac{1}{2} \log^{3/2}(t+1) \sqrt{\lambda d \beta_t^2 / t},$$
$$\frac{C_5 (d\beta_t^2)^3 \log^4(t+1)/t}{\log^{3/2}(t+1)\sqrt{\lambda d \beta_t^2/t}} = C_5 \sqrt{\frac{(d\beta_t^2)^5 \log^5(t+1)}{\lambda t}} \leq \tfrac{1}{2}.$$

Substituting these inequalities into (39) and using $d\beta_t^2/t \leq \lambda$ yields

$$\mathbb{E}[g_t] \leq \log^{3/2}(t+1) \min\{\lambda, \beta_t \sqrt{\lambda d/t}\}.$$

This proves Lemma 1 with the absolute constant $C = 1$. □

B.5 Proof Of Lemma 10

We use the coupling and regret decomposition of Bae et al. (2026a, Appendices B.1-B.2) (Lemma 9). For $i = 0, \ldots, t-1$, let π_i follow the learner through round i and the oracle afterwards. All queues start empty and share arrivals, contexts, algorithmic randomization, and independent service variables $U_j \sim \mathrm{Unif}(0,1)$. A service attempt with probability p succeeds if $U_j \leq p$. The oracle orders jobs by decreasing $p^*(x)$ and breaks ties by arrival time. Repeated contexts correspond to distinct jobs.

The queues $\mathcal{X}_i^{\pi_i}$ and $\mathcal{X}_i^{\pi_{i-1}}$ agree. Write $p_i = p(x_i, k_i)$ and $p_i^* = p(x_i^*, k_i^*)$, with both zero when $Q(i) = 0$. Then $g_i = p_i^* - p_i$. Define

$$q_{i,t} = \mathbb{P}\big(Q^{\pi_i}(t) - Q^{\pi_{i-1}}(t) = 1 \mid \mathcal{F}_i,\ p_i < U_i \leq p_i^*\big) \quad \text{if } g_i > 0, \qquad q_{i,t} = 0 \quad \text{if } g_i = 0.$$

Lemma 9 (Bae et al. 2026a, Appendices B.1-B.2)**.** *For every $1 \leq i < t$ and $j \geq i+1$, $Q^{\pi_i}(j) - Q^{\pi_{i-1}}(j) \in \{-1, 0, 1\}$. This difference is nonpositive unless $p_i < U_i \leq p_i^*$. On that event,*

$$\mathcal{X}_j^{\pi_{i-1}} \subseteq \mathcal{X}_j^{\pi_i}, \qquad Q^{\pi_i}(j) = 0 \implies \mathcal{X}_{j'}^{\pi_i} = \mathcal{X}_{j'}^{\pi_{i-1}} \quad \textit{for every } j' \geq j.$$

Consequently,

$$R_t = \sum_{i=1}^{t-1} \mathbb{E}[Q^{\pi_i}(t) - Q^{\pi_{i-1}}(t)] \leq \sum_{i=1}^{t-1} \mathbb{E}[g_i q_{i,t}].$$

The cited coupling uses the service probabilities and the common priority order. It therefore applies under Assumptions 1 and 2 without a covariance lower bound or context-wise slack.

Lemma 10. *For every $1 \leq i < t$,*

$$q_{i,t} \leq \mathbf{1}\{W(i) + 1 > \epsilon \underline{p}(t-i)/32\} + e^{-\epsilon^2 \underline{p}^2 (t-i)/128}.$$

Proof. By Lemma 9, it suffices to bound the probability that $Q^{\pi_i}(j) > 0$ for all $j \in [i+1, t]$. If $g_i = 0$, the result follows from $q_{i,t} = 0$. Otherwise, condition on $\mathcal{F}_i$ and $p_i < U_i \leq p_i^*$, and define

$$W^{\pi_i}(j) = \underline{p} \sum_{x \in \mathcal{X}_j^{\pi_i}} 1/p^*(x), \qquad \tau = \inf\{j \geq i+1 : W^{\pi_i}(j) = 0\}.$$

The learner fails service at i and an arrival adds at most one unit of workload. Thus

$$0 < W^{\pi_i}(i+1) \leq W(i) + 1, \qquad \{Q^{\pi_i}(t) - Q^{\pi_{i-1}}(t) = 1\} \subseteq \{\tau > t\}. \tag{41}$$

Let $\mathcal{F}_j^+$ include the coupled history before service at j. For $j \geq i+1$, policy π_i serves each chosen job x on a best server. On $\{j < \tau\}$,

$$\mathbb{E}[W^{\pi_i}(j+1) - W^{\pi_i}(j) \mid \mathcal{F}_j^+] = \lambda \underline{p}\, \mathbb{E}_{X\sim\mathcal{D}}[1/p^*(X)] - p^*(x)\underline{p}/p^*(x) = -\epsilon\underline{p}, \tag{42}$$

where the last equality follows from Assumption 2. Also, $|W^{\pi_i}(j+1) - W^{\pi_i}(j)| \leq 1$.

Set $\ell = t - i$. Suppose $\ell \geq 2$ and $W(i) + 1 \leq \epsilon\underline{p}\ell/32$. Conditional on $\mathcal{F}_{i+1}^+$, define

$$M_j = \sum_{j'=i+1}^{j-1} \mathbf{1}\{j' < \tau\}\big(W^{\pi_i}(j'+1) - W^{\pi_i}(j') + \epsilon\underline{p}\big), \qquad i+1 \leq j \leq t.$$

By (42), M_j is a martingale with increments bounded in absolute value by 2. On $\{\tau > t\}$,

$$\begin{aligned} M_t &= W^{\pi_i}(t) - W^{\pi_i}(i+1) + \epsilon\underline{p}(\ell-1) \\ &> -\epsilon\underline{p}\ell/32 + \epsilon\underline{p}(\ell-1) \geq 7\epsilon\underline{p}(\ell-1)/8, \end{aligned}$$

where the first inequality follows from (41) and $W^{\pi_i}(t) > 0$, and the second uses $\ell \leq 4(\ell-1)$. It follows that

$$\begin{aligned} \mathbb{P}(\tau > t \mid \mathcal{F}_{i+1}^+) &\leq \mathbb{P}(M_t > 7\epsilon\underline{p}(\ell-1)/8 \mid \mathcal{F}_{i+1}^+) \\ &\leq \exp\left(-\frac{(7\epsilon\underline{p}(\ell-1)/8)^2}{8(\ell-1)}\right) = e^{-49\epsilon^2\underline{p}^2(\ell-1)/512} \leq e^{-\epsilon^2\underline{p}^2\ell/128}. \end{aligned}$$

The second inequality follows from Lemma 16, and the last uses $\ell - 1 \geq \ell/2$. Taking conditional expectation given $\mathcal{F}_i$ and $p_i < U_i \leq p_i^*$, and applying (41), gives

$$q_{i,t}\mathbf{1}\{W(i) + 1 \leq \epsilon\underline{p}\ell/32\} \leq e^{-\epsilon^2\underline{p}^2\ell/128}.$$

When $\ell = 1$, the indicator is zero because $W(i) + 1 \geq 1 > \epsilon\underline{p}/32$. The result follows by using $q_{i,t} \leq 1$ on $\{W(i) + 1 > \epsilon\underline{p}\ell/32\}$. □

B.6 Proof Of Theorem 1

Proof. By Lemma 9,

$$\begin{aligned} R_t &\leq \sum_{i=1}^{t-1} \mathbb{E}[g_i q_{i,t}] \\ &= \underbrace{\sum_{i=1}^{\tau_{1,t}} \mathbb{E}[g_i q_{i,t}]}_{I_1} + \underbrace{\sum_{i=\tau_{1,t}+1}^{\tau_{2,t}} \mathbb{E}[g_i q_{i,t}]}_{I_2} + \underbrace{\sum_{i=\tau_{2,t}+1}^{t-1} \mathbb{E}[g_i q_{i,t}]}_{I_3}, \end{aligned} \tag{43}$$

where

$$\tau_{1,t} = t - \lceil C_4(\ell_{0,t} + 1)\rceil, \qquad \tau_{2,t} = t - \ell_{1,t} - 1.$$

For sufficiently large t, $1 \leq \tau_{1,t} < \tau_{2,t} < t$ and $\tau_{1,t} \geq t/2$.

We bound I_1 first. Since $g_i \leq 1$, Lemma 10 gives

$$I_1 \leq \sum_{i=1}^{\tau_{1,t}} \mathbb{P}(W(i) + 1 > \epsilon\underline{p}(t-i)/32) + \sum_{i=1}^{\tau_{1,t}} e^{-\epsilon^2\underline{p}^2(t-i)/128}. \tag{44}$$

For $i \leq \epsilon\underline{p}\, t/96$ and sufficiently large t,

$$W(i) + 1 \leq i + 1 \leq \epsilon\underline{p}\, t/48 \leq \epsilon\underline{p}(t-i)/32.$$

Thus these indices contribute zero to the first sum in (44). For $\epsilon\underline{p}\, t/96 < i \leq \tau_{1,t}$,

$$\ell_{0,i} + 1 \leq \ell_{0,t} + 1 \leq \epsilon\underline{p}(t - \tau_{1,t})/32 \leq \epsilon\underline{p}(t-i)/32.$$

The first inequality follows from the monotonicity of β_i, and the second from the definition of $\tau_{1,t}$. Consequently,

$$\sum_{i=1}^{\tau_{1,t}} \mathbb{P}(W(i)+1 > \epsilon\underline{p}(t-i)/32) \le \sum_{\epsilon\underline{p}\, t/96 < i \le \tau_{1,t}} \mathbb{P}(W(i) > \ell_{0,i}) \le \sum_{\epsilon\underline{p}\, t/96 < i \le \tau_{1,t}} C_5(i+1)^{-2} \le C_5(t+1)^{-1},$$

$$\sum_{i=1}^{\tau_{1,t}} e^{-\epsilon^2\underline{p}^2(t-i)/128} \le \frac{e^{-\epsilon^2\underline{p}^2(t-\tau_{1,t})/128}}{1-e^{-\epsilon^2\underline{p}^2/128}} \le C_5 e^{-\epsilon^2\underline{p}^2\ell_{1,t}/128} \le C_5(t+1)^{-1}.$$

The second inequality in the first chain follows from Lemma 5, and the last uses $(i+1)^{-2} \le C_5(t+1)^{-2}$. The last two inequalities in the second chain use $t - \tau_{1,t} \ge \ell_{1,t}$ and $C_3 = 128/(\epsilon\underline{p})^2$. Substituting into (44) gives

$$I_1 \le C_5(t+1)^{-1}. \tag{45}$$

For I_2, every index satisfies $i \ge t/2$ and $\epsilon\underline{p}(t-i)/32 - 1 \ge 0$ for sufficiently large t. By Lemma 8,

$$\begin{aligned}\mathbb{P}(W(i)+1 > \epsilon\underline{p}(t-i)/32) &\le C_5 e^{-\epsilon\underline{p}(\epsilon\underline{p}(t-i)/32-1)/2} + C_5(d\beta_i^2)^3\log^4(i+1)/i \\ &\le C_5 e^{-\epsilon^2\underline{p}^2(t-i)/128} + C_5(d\beta_t^2)^3\log^4(t+1)/t,\end{aligned}$$

where the second inequality uses $i^{-1} \le 2/t$, $\beta_i \le \beta_t$, and absorbs $e^{\epsilon\underline{p}/2}$ into C_5. Applying Lemma 10, we obtain

$$\begin{aligned}I_2 &\le C_5 \sum_{i=\tau_{1,t}+1}^{\tau_{2,t}} e^{-\epsilon^2\underline{p}^2(t-i)/128} + C_5(\tau_{2,t}-\tau_{1,t})(d\beta_t^2)^3\log^4(t+1)/t \\ &\le C_5 e^{-\epsilon^2\underline{p}^2\ell_{1,t}/128} + C_5(d\beta_t^2)^4\log^5(t+1)/t \\ &\le C_5(t+1)^{-1} + C_5(d\beta_t^2)^4\log^5(t+1)/t.\end{aligned} \tag{46}$$

The second inequality uses $\tau_{2,t} - \tau_{1,t} \le C_5 d\beta_t^2\log(t+1)$ and sums the geometric series. The last follows from (12).

For I_3, there are $\ell_{1,t}$ indices and $q_{i,t} \le 1$. By Lemma 1,

$$\begin{aligned}I_3 &\le \sum_{i=\tau_{2,t}+1}^{t-1} \mathbb{E}[g_i] \\ &\le C_5\ell_{1,t}\log^{3/2}(t+1)\min\{\lambda, \beta_t\sqrt{\lambda d/t}\} + C_5\ell_{1,t}(d\beta_t^2)^3\log^4(t+1)/t \\ &\le C_5\log^{5/2}(t+1)\min\{\lambda, \beta_t\sqrt{\lambda d/t}\} + C_5(d\beta_t^2)^4\log^5(t+1)/t.\end{aligned} \tag{47}$$

The second inequality uses $i \ge t/2$ and $\beta_i \le \beta_t$. The last uses $\ell_{1,t} \le C_5\log(t+1)$ and $d\beta_t^2 \ge 1$. Substituting (45)-(47) into (43) gives

$$R_t \le C_5\log^{5/2}(t+1)\min\{\lambda, \beta_t\sqrt{\lambda d/t}\} + C_5(d\beta_t^2)^4\log^5(t+1)/t. \tag{48}$$

It remains to specify t_0. From (9), for $t \ge 2$,

$$\beta_t^2 \le \left[\frac{\kappa^2}{2}(d+4) + 2\kappa S^2\right]\log(t+1), \tag{49}$$

using $(z_1+z_2)^2 \le 2z_1^2 + 2z_2^2$ and $d\kappa \ge 1$. Fix $C_5 \ge 1$ large enough for (48). Choose t_0 so that (40) holds for every $i \ge t/2$ and the hypotheses of Lemmas 5 and 8 hold for the indices above, $\tau_{1,t} \ge t/2$, $\lambda t \ge C_5\log(t+1)$, and, for all $t \ge t_0$,

$$\log(t+1) \ge 4C_5^2, \qquad d\beta_t^2/t \le \lambda, \qquad t \ge 4C_5^2(d\beta_t^2)^7\log^4(t+1)/\lambda. \tag{50}$$

The second condition makes $\beta_t\sqrt{\lambda d/t}$ the smaller term in the minimum. The first and third conditions imply

$$C_5 \log^{5/2}(t+1)\beta_t\sqrt{\lambda d/t} \le \tfrac{1}{2}\log^3(t+1)\beta_t\sqrt{\lambda d/t},$$
$$\frac{C_5(d\beta_t^2)^4\log^5(t+1)/t}{\log^3(t+1)\beta_t\sqrt{\lambda d/t}} = C_5\sqrt{\frac{(d\beta_t^2)^7\log^4(t+1)}{\lambda t}} \le \tfrac{1}{2}.$$

Applying these inequalities to (48) proves Theorem 1 with the absolute constant $C_1 = 1$. Such a threshold exists by (49). For fixed $\epsilon, \underline{p}$, it is polynomial in $d, \kappa, S, \lambda^{-1}$ up to logarithmic factors. Its dependence on $\epsilon, \underline{p}$ may be exponential because of the first condition in (50). □

C Proof Of Theorem 2

Proof. Fix an algorithm π. Let $d_0 = \min\{d-1, \lfloor \log_2 K \rfloor\}$, $p_0 = \mu(-1/4)$, and

$$\eta = \min\{1/8, \sqrt{p_0}\, d_0/(2\sqrt{\lambda t})\}.$$

There is one context x_0. Index 2^{d_0} servers by $\boldsymbol{u} \in \{-1,+1\}^{d_0}$ and instances by $\boldsymbol{v} \in \{-1,+1\}^{d_0}$. For coordinate vectors $\boldsymbol{e}_0, \dots, \boldsymbol{e}_{d_0} \in \mathbb{R}^d$, define

$$\boldsymbol{\phi}(x_0, k_{\boldsymbol{u}}) = \frac{1}{\sqrt{2}}\left(\boldsymbol{e}_0 + \frac{1}{\sqrt{d_0}}\sum_{a=1}^{d_0} u_a \boldsymbol{e}_a\right), \tag{51}$$

$$\boldsymbol{\theta}^{\boldsymbol{v}} = \sqrt{2}\left(-\eta \boldsymbol{e}_0 + \frac{\eta}{\sqrt{d_0}}\sum_{a=1}^{d_0} v_a \boldsymbol{e}_a\right). \tag{52}$$

If $K > 2^{d_0}$, the remaining servers duplicate these servers and retain their sign vectors. Let d_{H} denote Hamming distance. Then

$$\langle \boldsymbol{\phi}(x_0, k_{\boldsymbol{u}}), \boldsymbol{\theta}^{\boldsymbol{v}} \rangle = -\eta + \frac{\eta}{d_0}\sum_{a=1}^{d_0} u_a v_a = -\frac{2\eta}{d_0} d_{\mathrm{H}}(\boldsymbol{u}, \boldsymbol{v}),$$
$$\|\boldsymbol{\phi}(x_0, k_{\boldsymbol{u}})\|_2^2 = 1, \qquad \|\boldsymbol{\theta}^{\boldsymbol{v}}\|_2^2 = 4\eta^2 \le 1/16, \qquad -1/4 \le \langle \boldsymbol{\phi}(x_0, k_{\boldsymbol{u}}), \boldsymbol{\theta}^{\boldsymbol{v}} \rangle \le 0.$$

Thus the best server has service probability $p^*(x_0) = 1/2$, and every server has service probability at least p_0. The instances satisfy Assumptions 1 and 2 with $S = 1/4$ and normalized capacity slack $\epsilon = 1 - 2\lambda \ge 1/2$. Write $\nu^{\boldsymbol{v}}$ for the instance defined by (52), and $\mathbb{P}^{\boldsymbol{v}}, \mathbb{E}^{\boldsymbol{v}}$ for its probability and expectation.

We first show that $R_t^\pi(\nu^{\boldsymbol{v}}) \ge \mathbb{E}^{\boldsymbol{v}}[g_{t-1}]$. Let π' follow π through round $t-2$ and use the best server in round $t-1$. Couple the queues with common arrivals and independent service uniforms $U_i \sim \mathrm{Unif}(0,1)$. Their updates are $Q(i+1) = Q(i) - \mathbf{1}\{Q(i) > 0,\ U_i \le p_i\} + A(i)$, where p_i is the chosen server's service probability on nonempty rounds and $p_i = 1/2$ otherwise. This update is nondecreasing in $Q(i)$ and nonincreasing in p_i. Hence $Q^{\pi'}(t) \ge Q^*(t)$. Also, $Q^\pi(t-1) = Q^{\pi'}(t-1)$, so

$$\begin{aligned} R_t^\pi(\nu^{\boldsymbol{v}}) &= \mathbb{E}^{\boldsymbol{v}}[Q^\pi(t) - Q^*(t)] \ge \mathbb{E}^{\boldsymbol{v}}[Q^\pi(t) - Q^{\pi'}(t)] \\ &= \mathbb{E}^{\boldsymbol{v}}\left[\mathbf{1}\{Q(t-1) > 0,\ p_{t-1} < U_{t-1} \le 1/2\}\right] = \mathbb{E}^{\boldsymbol{v}}[g_{t-1}]. \end{aligned} \tag{53}$$

It remains to lower-bound $2^{-d_0}\sum_{\boldsymbol{v}} \mathbb{E}^{\boldsymbol{v}}[g_{t-1}]$.

Let $N(t-2) = \sum_{i=1}^{t-2} \mathbf{1}\{Q(i) > 0\}$. By the queue recursion,

$$p_0 \mathbb{E}^{\boldsymbol{v}}[N(t-2)] \le \mathbb{E}^{\boldsymbol{v}}\left[\sum_{i=1}^{t-2} D(i)\right] = \mathbb{E}^{\boldsymbol{v}}\left[\sum_{i=1}^{t-2} A(i)\right] - \mathbb{E}^{\boldsymbol{v}}[Q(t-1)] \le \lambda t, \tag{54}$$

where the first inequality follows from $p_i \ge p_0$ on nonempty rounds, the equality uses $Q(1) = 0$, and the last inequality uses $Q(t-1) \ge 0$.

For $a \in [d_0]$, let $\boldsymbol{v}^{(a)}$ be $\boldsymbol{v}$ with coordinate a flipped. Let $\mathbb{P}^{\boldsymbol{v}}_{t-2}$ be the law of the history through service in round $t-2$, including policy randomization but excluding $A(t-2)$. For $z_1, z_2 \in \mathbb{R}$, we have

$$D_{\mathrm{KL}}\big(\mathrm{Bern}(\mu(z_1))\,\|\,\mathrm{Bern}(\mu(z_2))\big) = \int_{z_1}^{z_2} (\mu(z) - \mu(z_1))\,dz \le (z_1 - z_2)^2/8, \tag{55}$$

where the inequality follows from $0 < \dot{\mu}(z) \le 1/4$. For every server, $|\langle \boldsymbol{\phi}(x_0, k), \boldsymbol{\theta}^{\boldsymbol{v}} - \boldsymbol{\theta}^{\boldsymbol{v}^{(a)}} \rangle| = 2\eta/d_0$. The conditional distributions of arrivals and the action-selection rule are the same in both instances. The chain rule for relative entropy (Lemma 17) gives

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}^{\boldsymbol{v}}_{t-2} \| \mathbb{P}^{\boldsymbol{v}^{(a)}}_{t-2}) &\le \mathbb{E}^{\boldsymbol{v}}\left[\sum_{i=1}^{t-2} \mathbf{1}\{Q(i) > 0\} \frac{1}{8}\left(\frac{2\eta}{d_0}\right)^2\right] \\ &= \frac{\eta^2}{2d_0^2}\mathbb{E}^{\boldsymbol{v}}[N(t-2)] \le \frac{\lambda t \eta^2}{2p_0 d_0^2} \le \frac{1}{8}, \end{aligned}$$

where the first inequality follows from (55), the second follows from (54), and the last follows from $\eta \le \sqrt{p_0} d_0/(2\sqrt{\lambda t})$. Pinsker's inequality (Lemma 18) therefore gives

$$\|\mathbb{P}^{\boldsymbol{v}}_{t-2} - \mathbb{P}^{\boldsymbol{v}^{(a)}}_{t-2}\|_{\mathrm{TV}} \le 1/4.$$

On $A(t-2) = 1$, we have $Q(t-1) > 0$. Let $\boldsymbol{u}_{t-1}$ be the sign vector of the chosen server. The arrival $A(t-2)$ is independent of the preceding history, and the rule for choosing $\boldsymbol{u}_{t-1}$ given that history and $A(t-2) = 1$ is the same in both instances. For $v_a = +1$, Lemma 18 therefore gives

$$\begin{aligned} &\mathbb{P}^{\boldsymbol{v}}(u_{t-1,a} \ne v_a \mid A(t-2) = 1) + \mathbb{P}^{\boldsymbol{v}^{(a)}}(u_{t-1,a} \ne v_a^{(a)} \mid A(t-2) = 1) \\ &\quad = 1 - \mathbb{P}^{\boldsymbol{v}}(u_{t-1,a} = +1 \mid A(t-2) = 1) + \mathbb{P}^{\boldsymbol{v}^{(a)}}(u_{t-1,a} = +1 \mid A(t-2) = 1) \\ &\quad \ge 1 - \|\mathbb{P}^{\boldsymbol{v}}_{t-2} - \mathbb{P}^{\boldsymbol{v}^{(a)}}_{t-2}\|_{\mathrm{TV}} \ge 3/4. \end{aligned}$$

Summing over the 2^{d_0-1} pairs in each coordinate yields

$$\begin{aligned} \frac{1}{2^{d_0}} \sum_{\boldsymbol{v}} \mathbb{E}^{\boldsymbol{v}}[d_{\mathrm{H}}(\boldsymbol{u}_{t-1}, \boldsymbol{v}) \mid A(t-2) = 1] &= \frac{1}{2^{d_0}} \sum_{a=1}^{d_0} \sum_{\boldsymbol{v}} \mathbb{P}^{\boldsymbol{v}}(u_{t-1,a} \ne v_a \mid A(t-2) = 1) \\ &\ge \frac{1}{2^{d_0}} \sum_{a=1}^{d_0} 2^{d_0-1} \cdot \frac{3}{4} = \frac{3d_0}{8}. \end{aligned} \tag{56}$$

Moreover,

$$\frac{1}{2} - \mu\left(-\frac{2\eta}{d_0} d_{\mathrm{H}}(\boldsymbol{u}, \boldsymbol{v})\right) = \int_{-2\eta d_{\mathrm{H}}(\boldsymbol{u},\boldsymbol{v})/d_0}^{0} \dot{\mu}(z)\,dz \ge \frac{2\dot{\mu}(1/4)\eta}{d_0} d_{\mathrm{H}}(\boldsymbol{u}, \boldsymbol{v}),$$

where the inequality follows from $-2\eta d_{\mathrm{H}}(\boldsymbol{u}, \boldsymbol{v})/d_0 \in [-1/4, 0]$ and $\dot{\mu}(z) \ge \dot{\mu}(1/4)$ on this interval. Thus

$$\begin{aligned} \frac{1}{2^{d_0}} \sum_{\boldsymbol{v}} \mathbb{E}^{\boldsymbol{v}}[g_{t-1}] &\ge \frac{\lambda}{2^{d_0}} \sum_{\boldsymbol{v}} \mathbb{E}^{\boldsymbol{v}}[g_{t-1} \mid A(t-2) = 1] \\ &\ge \frac{2\lambda\dot{\mu}(1/4)\eta}{d_0} \cdot \frac{1}{2^{d_0}} \sum_{\boldsymbol{v}} \mathbb{E}^{\boldsymbol{v}}[d_{\mathrm{H}}(\boldsymbol{u}_{t-1}, \boldsymbol{v}) \mid A(t-2) = 1] \ge \frac{3\dot{\mu}(1/4)}{4} \lambda\eta, \end{aligned} \tag{57}$$

where the first inequality uses $g_{t-1} \ge 0$ and $\mathbb{P}(A(t-2) = 1) = \lambda$, the second uses $\dot{\mu}(z) \ge \dot{\mu}(1/4)$ on $[-1/4, 0]$, and the last follows from (56). Substituting (57) into (53), we obtain

$$\sup_{\boldsymbol{v}} R^{\pi}_t(\nu^{\boldsymbol{v}}) \ge \frac{1}{2^{d_0}} \sum_{\boldsymbol{v}} R^{\pi}_t(\nu^{\boldsymbol{v}}) \ge \frac{3\dot{\mu}(1/4)}{4} \min\{\lambda/8, \sqrt{p_0}\, d_0 \sqrt{\lambda/t}/2\}.$$

Since $d - 1 \ge d/2$ and $\lfloor \log_2 K \rfloor \ge \log_2 K/2 \ge \log K/2$, we have $d_0 \ge \min\{d, \log K\}/2$. Choosing a sufficiently large absolute constant C_0 proves the theorem. $\square$

D Proof Of Theorem 3

Lemma 11. *Let $0 < \epsilon \le 1/8$, $\lambda_0 = 1/4$, and $p^* = \lambda_0/(1-\epsilon)$. Consider a queue with Bernoulli arrivals of mean λ_0 and service probability p^* on nonempty rounds. Arrivals and service outcomes are independent, with arrivals after service. Let*

$$q_0 = \lceil 5/(8\epsilon) \rceil + 1, \qquad t_0 = \lceil 60/\epsilon^2 \rceil, \qquad \ell = \lfloor 1/(16\epsilon^2) \rfloor.$$

If $Q(1) = 0$, then $\mathbb{P}(Q(i) \ge q_0) \ge 1/20$ for every $i \ge t_0$. If $Q(1) \ge q_0 - 1$, then $\mathbb{P}(Q(j) > 0$ for every $j \in [1, \ell+1]) \ge 1/2$.

Proof. Let ϖ be the stationary distribution. For $q \ge 1$, the transition probabilities from q to $q-1$ and $q+1$ are $p^*(1-\lambda_0)$ and $(1-p^*)\lambda_0$, respectively. From $q = 0$, the probability of moving to $q = 1$ is λ_0. The balance equations give

$$\varpi(0) = 1 - \lambda_0/p^* = \epsilon,$$

$$\varpi(Q \ge q) = (1-\epsilon)\left(\frac{(1-p^*)\lambda_0}{p^*(1-\lambda_0)}\right)^{q-1} = (1-\epsilon)\left(1 - \frac{\epsilon}{1-\lambda_0}\right)^{q-1}, \qquad q \ge 1.$$

Therefore,

$$\varpi(Q \ge q_0) \ge \frac{7}{8}\exp\left[-\frac{2\epsilon}{1-\lambda_0}\left(\frac{5}{8\epsilon}+1\right)\right] \ge \frac{7}{8}e^{-2} > \frac{1}{10},$$

where the first inequality follows from $\log(1-z) \ge -2z$ for $z \in [0, 1/2]$, $\epsilon/(1-\lambda_0) \le 1/6$, and $q_0 - 1 \le 5/(8\epsilon)+1$. The second inequality follows from $2\epsilon(5/(8\epsilon)+1)/(1-\lambda_0) = 5/3 + 8\epsilon/3 \le 2$.

Let $\tau_0 = \inf\{j \ge 1 : Q(j) = 0\}$, and let $\mathbb{E}_q$ denote expectation when $Q(1) = q$. Before τ_0, $\mathbb{E}[Q(i+1) - Q(i) \mid \mathcal{F}_i] = \lambda_0 - p^* = -p^*\epsilon$. Stopping at $\tau_0 \wedge j$, we obtain

$$0 \le \mathbb{E}_q[Q(\tau_0 \wedge j)] = q - p^*\epsilon\,\mathbb{E}_q[(\tau_0 \wedge j) - 1].$$

By monotone convergence, $\mathbb{E}_q[\tau_0 - 1] \le q/(p^*\epsilon)$. Averaging over q with distribution ϖ gives

$$\mathbb{E}_\varpi[Q(1)] = \sum_{q=1}^{\infty} \varpi(Q \ge q) = \frac{(1-\lambda_0)(1-\epsilon)}{\epsilon},$$

$$\mathbb{E}_\varpi[\tau_0 - 1] \le \frac{\mathbb{E}_\varpi[Q(1)]}{p^*\epsilon} = \frac{3(1-\epsilon)^2}{\epsilon^2}.$$

Couple Q, starting from zero, with Q^ϖ, starting from ϖ, using the same arrivals and service uniforms. Then $Q^\varpi(j) \ge Q(j)$, and the two queues agree from the first round when $Q^\varpi(j) = 0$. Hence, for $i \ge t_0$,

$$\begin{aligned}\mathbb{P}(Q(i) \ge q_0 \mid Q(1) = 0) &\ge \varpi(Q \ge q_0) - \mathbb{P}_\varpi(\tau_0 > i) \\ &\ge \frac{1}{10} - \frac{\mathbb{E}_\varpi[\tau_0 - 1]}{i-1} \ge \frac{1}{10} - \frac{3(1-\epsilon)^2}{60-\epsilon^2} \ge \frac{1}{20},\end{aligned}$$

where the second inequality follows from Markov's inequality, the third follows from $i \ge 60/\epsilon^2$, and the last follows from $60(1-\epsilon)^2 \le 60 - \epsilon^2$.

For the second claim, suppose $Q(1) \ge q_0 - 1$. Define

$$M_j = \sum_{i=1}^{j} \mathbf{1}\{i < \tau_0\}\big(Q(i+1) - Q(i) + p^*\epsilon\big), \qquad M_0 = 0.$$

We have $\mathbb{E}[M_j - M_{j-1} \mid \mathcal{F}_j] = 0$ and $|M_j - M_{j-1}| \le 2$. On $\{\tau_0 \le \ell + 1\}$,

$$M_{\tau_0 - 1} = -Q(1) + p^*\epsilon(\tau_0 - 1) \le -\frac{5}{8\epsilon} + \frac{2}{7}\epsilon\frac{1}{16\epsilon^2} = -\frac{17}{28\epsilon},$$

where the inequality uses $p^* \le 2/7$ and $\tau_0 - 1 \le \ell$. The maximal Azuma-Hoeffding inequality (Lemma 16) gives

$$\mathbb{P}(\tau_0 \le \ell + 1) \le \mathbb{P}\left(\min_{0 \le j \le \ell} M_j \le -\frac{17}{28\epsilon}\right) \le \exp\left[-\frac{(17/(28\epsilon))^2}{8\ell}\right] \le e^{-289/392} < 1/2.$$

Thus $\mathbb{P}(Q(j) > 0$ for every $j \in [1, \ell+1]) \ge 1/2$. □

Proof of Theorem 3. Fix an algorithm π. Let $d_0 = \min\{d-1, \lfloor \log_2 K \rfloor\}$ and use the features in (51), with the same convention for duplicate servers. Set

$$\lambda_0 = 1/4, \qquad p^* = \lambda_0/(1-\epsilon), \qquad z^* = \log(p^*/(1-p^*)), \qquad \eta = \min\{\epsilon, d_0/\sqrt{t}\}/20,$$

and define

$$\boldsymbol{\theta}^{\boldsymbol{v}} = \sqrt{2}\left((z^* - \eta)\boldsymbol{e}_0 + \frac{\eta}{\sqrt{d_0}}\sum_{a=1}^{d_0} v_a \boldsymbol{e}_a\right), \qquad \boldsymbol{v} \in \{-1,+1\}^{d_0}.$$

Then

$$\langle \boldsymbol{\phi}(x_0, k_{\boldsymbol{u}}), \boldsymbol{\theta}^{\boldsymbol{v}} \rangle = z^* - \frac{2\eta}{d_0} d_{\mathrm{H}}(\boldsymbol{u}, \boldsymbol{v}), \qquad \max_k \mu(\langle \boldsymbol{\phi}(x_0, k), \boldsymbol{\theta}^{\boldsymbol{v}} \rangle) = p^*. \tag{58}$$

Since $p^* \in [1/4, 2/7]$, $|z^*| \le \log 3$, and $\eta \le 1/160$,

$$\|\boldsymbol{\theta}^{\boldsymbol{v}}\|_2^2 = 2\big((z^* - \eta)^2 + \eta^2\big) \le 2\big((\log 3 + 1/160)^2 + (1/160)^2\big) < 4.$$

The instances satisfy Assumptions 1 and 2 with $S = 2$ and normalized capacity slack $1 - \lambda_0/p^* = \epsilon$. Write $\nu^{\boldsymbol{v}}$ for each instance and $\mathbb{P}^{\boldsymbol{v}}, \mathbb{E}^{\boldsymbol{v}}$ for its probability and expectation.

We first express regret in terms of $\mathbb{E}^{\boldsymbol{v}}[g_i \mathbf{1}\{Q^\pi(i) \ge q_0\}]$, where q_0, t_0, ℓ are as in Lemma 11. For $i = 0, \dots, t-1$, let π_i follow π through round i and use the best server thereafter. Since $Q^{\pi_0}(t) = Q^*(t)$ and $Q^{\pi_{t-1}}(t) = Q^\pi(t)$,

$$R_t^\pi(\nu^{\boldsymbol{v}}) = \sum_{i=1}^{t-1} \mathbb{E}^{\boldsymbol{v}}[Q^{\pi_i}(t) - Q^{\pi_{i-1}}(t)]. \tag{59}$$

Couple the queues using common arrivals $A(j)$ and service uniforms U_j. In round i, both queues have length $Q^\pi(i)$, but π_i and π_{i-1} use service probabilities p_i and p^*, respectively. Here $p_i = \mu(\langle \boldsymbol{\phi}_i, \boldsymbol{\theta}^{\boldsymbol{v}} \rangle)$ on nonempty rounds, and we set $p_i = p^*$ on empty rounds. Thus $g_i = p^* - p_i$. After round i, both queues use p^*. The monotonicity of the queue update gives $Q^{\pi_i}(j) \ge Q^{\pi_{i-1}}(j)$ for $j \ge i+1$.

Define $\mathcal{E}_i = \{Q^\pi(i) \ge q_0,\ p_i < U_i \le p^*\}$. On $\mathcal{E}_i$,

$$Q^{\pi_{i-1}}(i+1) = Q^\pi(i) - 1 + A(i) \ge q_0 - 1, \qquad Q^{\pi_i}(i+1) = Q^{\pi_{i-1}}(i+1) + 1.$$

If $Q^{\pi_{i-1}}(j) > 0$ for every $j \in [i+1, t]$, then the two queues have the same departures and $Q^{\pi_i}(t) - Q^{\pi_{i-1}}(t) = 1$. For $t - i - 1 \le \ell$, Lemma 11 gives probability at least $1/2$ for this event, conditional on $\mathcal{F}_i \vee \sigma(U_i, A(i))$ and $\mathcal{E}_i$. Consequently,

$$\mathbb{E}^{\boldsymbol{v}}[Q^{\pi_i}(t) - Q^{\pi_{i-1}}(t)] \ge \tfrac{1}{2}\mathbb{P}^{\boldsymbol{v}}(\mathcal{E}_i) = \tfrac{1}{2}\mathbb{E}^{\boldsymbol{v}}[g_i \mathbf{1}\{Q^\pi(i) \ge q_0\}],$$

where the equality follows from $\mathbb{P}(p_i < U_i \le p^* \mid \mathcal{F}_i) = p^* - p_i = g_i$. For $t \ge 64/\epsilon^2$, we have $t - \ell \ge (64 - 1/16)/\epsilon^2 \ge 60/\epsilon^2 + 1 \ge t_0$. Keeping the nonnegative summands with $i \in [t-\ell, t-1]$ in (59) gives

$$R_t^\pi(\nu^{\boldsymbol{v}}) \ge \frac{1}{2}\sum_{i=t-\ell}^{t-1} \mathbb{E}^{\boldsymbol{v}}[g_i \mathbf{1}\{Q^\pi(i) \ge q_0\}]. \tag{60}$$

To bound (60), let $\mathbb{P}_i^{\boldsymbol{v}}$ denote the history law before service at i, including the chosen action. Flipping coordinate a gives $|\langle \boldsymbol{\phi}(x_0, k), \boldsymbol{\theta}^{\boldsymbol{v}} - \boldsymbol{\theta}^{\boldsymbol{v}^{(a)}} \rangle| = 2\eta/d_0$. For $i \le t$, the chain rule for relative entropy (Lemma 17) and (55) give

$$D_{\mathrm{KL}}(\mathbb{P}_i^{\boldsymbol{v}} \| \mathbb{P}_i^{\boldsymbol{v}^{(a)}}) \le \mathbb{E}^{\boldsymbol{v}}\left[\sum_{j=1}^{i-1} \mathbf{1}\{Q(j) > 0\} \frac{1}{8}\left(\frac{2\eta}{d_0}\right)^2\right] \le \frac{t\eta^2}{2d_0^2} \le \frac{1}{800},$$

$$\|\mathbb{P}_i^{\boldsymbol{v}} - \mathbb{P}_i^{\boldsymbol{v}^{(a)}}\|_{\mathrm{TV}} \le \sqrt{\tfrac{1}{2} D_{\mathrm{KL}}(\mathbb{P}_i^{\boldsymbol{v}} \| \mathbb{P}_i^{\boldsymbol{v}^{(a)}})} \le \frac{1}{40},$$

where the last inequality in the first line uses $\eta \le d_0/(20\sqrt{t})$. The second line follows from Pinsker's inequality (Lemma 18).

Since $p_i \le p^*$, common arrivals and uniforms give $Q^\pi(i) \ge Q^*(i)$. By Lemma 11,

$$\mathbb{P}^{\boldsymbol{v}}(Q^\pi(i) \ge q_0) \ge \mathbb{P}^{\boldsymbol{v}}(Q^*(i) \ge q_0) \ge 1/20, \qquad i \ge t_0.$$

Let $\boldsymbol{u}_i$ be the chosen server's sign vector, arbitrary when $Q(i) = 0$. For $i \geq t_0$ and $v_a = +1$,

$$\begin{aligned}
&\mathbb{P}^{\boldsymbol{v}}(Q^\pi(i) \geq q_0,\ u_{i,a} \neq v_a) + \mathbb{P}^{\boldsymbol{v}^{(a)}}(Q^\pi(i) \geq q_0,\ u_{i,a} \neq v_a^{(a)}) \\
&\quad = \mathbb{P}^{\boldsymbol{v}}(Q^\pi(i) \geq q_0) - \mathbb{P}^{\boldsymbol{v}}(Q^\pi(i) \geq q_0,\ u_{i,a} = +1) + \mathbb{P}^{\boldsymbol{v}^{(a)}}(Q^\pi(i) \geq q_0,\ u_{i,a} = +1) \\
&\quad \geq \mathbb{P}^{\boldsymbol{v}}(Q^\pi(i) \geq q_0) - \|\mathbb{P}_i^{\boldsymbol{v}} - \mathbb{P}_i^{\boldsymbol{v}^{(a)}}\|_{\mathrm{TV}} \geq \frac{1}{20} - \frac{1}{40} = \frac{1}{40}.
\end{aligned}$$

The first inequality follows from Lemma 18 applied to $\{Q^\pi(i) \geq q_0,\ u_{i,a} = +1\}$. Summing over 2^{d_0-1} pairs per coordinate gives

$$\begin{aligned}
\frac{1}{2^{d_0}} \sum_{\boldsymbol{v}} \mathbb{E}^{\boldsymbol{v}}[\mathbf{1}\{Q^\pi(i) \geq q_0\} d_{\mathrm{H}}(\boldsymbol{u}_i, \boldsymbol{v})] &= \frac{1}{2^{d_0}} \sum_{a=1}^{d_0} \sum_{\boldsymbol{v}} \mathbb{P}^{\boldsymbol{v}}(Q^\pi(i) \geq q_0,\ u_{i,a} \neq v_a) \\
&\geq \frac{1}{2^{d_0}} \sum_{a=1}^{d_0} \frac{2^{d_0-1}}{40} = \frac{d_0}{80}. \qquad (61)
\end{aligned}$$

On $\{Q^\pi(i) \geq q_0\}$, (58) gives

$$\begin{aligned}
g_i = \mu(z^*) - \mu\left(z^* - \frac{2\eta}{d_0} d_{\mathrm{H}}(\boldsymbol{u}_i, \boldsymbol{v})\right) &= \int_{z^* - 2\eta d_{\mathrm{H}}(\boldsymbol{u}_i, \boldsymbol{v})/d_0}^{z^*} \dot{\mu}(z)\, dz \\
&\geq \frac{2\eta}{d_0(2 + e^2 + e^{-2})} d_{\mathrm{H}}(\boldsymbol{u}_i, \boldsymbol{v}),
\end{aligned}$$

where the inequality follows from $-2 \leq z^* - 2\eta d_{\mathrm{H}}(\boldsymbol{u}_i, \boldsymbol{v})/d_0 \leq z^* \leq 2$ and $\dot{\mu}(z) \geq (2 + e^2 + e^{-2})^{-1}$ on $[-2, 2]$. Taking expectations and applying (61) gives

$$\frac{1}{2^{d_0}} \sum_{\boldsymbol{v}} \mathbb{E}^{\boldsymbol{v}}[g_i \mathbf{1}\{Q^\pi(i) \geq q_0\}] \geq \frac{2\eta}{d_0(2 + e^2 + e^{-2})} \cdot \frac{d_0}{80} = \frac{\eta}{40(2 + e^2 + e^{-2})}. \qquad (62)$$

Substituting (62) into (60) yields

$$\begin{aligned}
\sup_{\boldsymbol{v}} R_t^\pi(\nu^{\boldsymbol{v}}) \geq \frac{1}{2^{d_0}} \sum_{\boldsymbol{v}} R_t^\pi(\nu^{\boldsymbol{v}}) &\geq \frac{\ell\eta}{80(2 + e^2 + e^{-2})} \\
&\geq \frac{1}{51200(2 + e^2 + e^{-2})} \min\{1/\epsilon, d_0/(\epsilon^2\sqrt{t})\},
\end{aligned}$$

where the last inequality uses $\ell \geq 1/(32\epsilon^2)$ and $\eta = \min\{\epsilon, d_0/\sqrt{t}\}/20$. Since $d_0 \geq \min\{d, \log K\}/2$ by Section C, the theorem follows for a sufficiently large absolute constant C_0. □

E Auxiliary Lemmas

Lemma 12 (Theorem 1, Abbasi-Yadkori et al. 2011)**.** *Let $\{\mathcal{F}_i\}_{i \geq 1}$ be a filtration, $\boldsymbol{\phi}_i \in \mathbb{R}^d$ be $\mathcal{F}_i$-measurable, and η_i be $\mathcal{F}_{i+1}$-measurable. Suppose that, for some $\sigma > 0$ and every $z \in \mathbb{R}$, $\mathbb{E}[e^{z\eta_i} \mid \mathcal{F}_i] \leq e^{\sigma^2 z^2/2}$. For a fixed positive definite matrix $\mathbf{V}_0$, define $\mathbf{V}_t = \mathbf{V}_0 + \sum_{i=1}^t \boldsymbol{\phi}_i \boldsymbol{\phi}_i^\top$. For every $\delta \in (0, 1)$, with probability at least $1 - \delta$, simultaneously for all $t \geq 0$,*

$$\|\sum_{i=1}^t \eta_i \boldsymbol{\phi}_i\|_{\mathbf{V}_t^{-1}}^2 \leq \sigma^2 \left(\log \frac{\det \mathbf{V}_t}{\det \mathbf{V}_0} + 2\log(1/\delta)\right).$$

Lemma 13. *Under Assumption 1, let $\widehat{\boldsymbol{\theta}}_t$ be the regularized maximum likelihood estimator in (8), and define*

$$\boldsymbol{G}_t = \sum_{i=1}^t \big(D(i) - \mu(\langle \boldsymbol{\phi}_i, \boldsymbol{\theta}^* \rangle)\big) \boldsymbol{\phi}_i, \qquad \mathbf{V}_{\kappa,t} = \kappa \mathbf{I} + \sum_{i=1}^t \boldsymbol{\phi}_i \boldsymbol{\phi}_i^\top.$$

Then, for every $t \geq 0$, we have

$$\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_{\kappa,t}} \leq \kappa \|\boldsymbol{G}_t\|_{\mathbf{V}_{\kappa,t}^{-1}} + \sqrt{\kappa} S.$$

Proof. Let f_t be the objective in (8), and write $\boldsymbol{\Delta}_t = \widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*$. For every $\boldsymbol{\theta} \in \mathcal{B}$, we have

$$\nabla f_t(\boldsymbol{\theta}) = \boldsymbol{\theta} + \sum_{i=1}^{t} \big(\mu(\langle \boldsymbol{\phi}_i, \boldsymbol{\theta}\rangle) - D(i)\big)\boldsymbol{\phi}_i,$$
$$\nabla^2 f_t(\boldsymbol{\theta}) = \mathbf{I} + \sum_{i=1}^{t} \dot{\mu}(\langle \boldsymbol{\phi}_i, \boldsymbol{\theta}\rangle)\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top \succeq \mathbf{V}_{\kappa,t}/\kappa,$$

where the inequality follows from Assumption 1. Since $\widehat{\boldsymbol{\theta}}_t$ minimizes f_t over $\mathcal{B}$ and $\boldsymbol{\theta}^* \in \mathcal{B}$, $\langle \boldsymbol{\Delta}_t, \nabla f_t(\widehat{\boldsymbol{\theta}}_t)\rangle \le 0$. Therefore,

$$\begin{aligned}
\|\boldsymbol{\Delta}_t\|^2_{\mathbf{V}_{\kappa,t}}/\kappa &\le \int_0^1 \boldsymbol{\Delta}_t^\top \nabla^2 f_t(\boldsymbol{\theta}^* + z\boldsymbol{\Delta}_t)\boldsymbol{\Delta}_t \, dz \\
&= \langle \boldsymbol{\Delta}_t, \nabla f_t(\widehat{\boldsymbol{\theta}}_t) - \nabla f_t(\boldsymbol{\theta}^*)\rangle \\
&\le \langle \boldsymbol{\Delta}_t, \boldsymbol{G}_t - \boldsymbol{\theta}^*\rangle \\
&\le \|\boldsymbol{\Delta}_t\|_{\mathbf{V}_{\kappa,t}}\big(\|\boldsymbol{G}_t\|_{\mathbf{V}^{-1}_{\kappa,t}} + \|\boldsymbol{\theta}^*\|_{\mathbf{V}^{-1}_{\kappa,t}}\big) \\
&\le \|\boldsymbol{\Delta}_t\|_{\mathbf{V}_{\kappa,t}}\big(\|\boldsymbol{G}_t\|_{\mathbf{V}^{-1}_{\kappa,t}} + S/\sqrt{\kappa}\big),
\end{aligned}$$

where the first inequality follows from $\boldsymbol{\theta}^* + z\boldsymbol{\Delta}_t \in \mathcal{B}$ for $z \in [0,1]$ and $\nabla^2 f_t(\boldsymbol{\theta}) \succeq \mathbf{V}_{\kappa,t}/\kappa$ on $\mathcal{B}$, the second inequality follows from $\langle \boldsymbol{\Delta}_t, \nabla f_t(\widehat{\boldsymbol{\theta}}_t)\rangle \le 0$ and $\nabla f_t(\boldsymbol{\theta}^*) = \boldsymbol{\theta}^* - \boldsymbol{G}_t$, the third inequality follows from Cauchy-Schwarz and the triangle inequality, and the last inequality follows from $\mathbf{V}_{\kappa,t} \succeq \kappa\mathbf{I}$ and $\|\boldsymbol{\theta}^*\|_2 \le S$. Dividing by $\|\boldsymbol{\Delta}_t\|_{\mathbf{V}_{\kappa,t}}/\kappa$ gives the result when $\boldsymbol{\Delta}_t \neq 0$. The result is immediate when $\boldsymbol{\Delta}_t = 0$. □

Lemma 14 (Lemma 11, Abbasi-Yadkori et al. 2011)**.** *Let $\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_t \in \mathbb{R}^d$ satisfy $\|\boldsymbol{\phi}_i\|_2 \le 1$, and define $\mathbf{V}_0 = \mathbf{I}$ and $\mathbf{V}_i = \mathbf{I} + \sum_{j=1}^{i} \boldsymbol{\phi}_j\boldsymbol{\phi}_j^\top$. Then*

$$\sum_{i=1}^{t} \min\{1, \|\boldsymbol{\phi}_i\|^2_{\mathbf{V}^{-1}_{i-1}}\} \le 2\log\det \mathbf{V}_t \le 2d\log(1 + t/d).$$

Lemma 15 (Hoeffding's lemma)**.** *Let Y be a random variable with $a \le Y \le b$ almost surely, and let $\mathcal{G}$ be a sigma-field. For every $z \in \mathbb{R}$,*

$$\mathbb{E}\Big[e^{z(Y - \mathbb{E}[Y|\mathcal{G}])} \mid \mathcal{G}\Big] \le e^{z^2(b-a)^2/8}.$$

Lemma 16 (Azuma-Hoeffding inequality)**.** *Let $\{\mathcal{F}_i\}_{i\ge 1}$ be a filtration and η_i be $\mathcal{F}_{i+1}$-measurable with $\mathbb{E}[\eta_i \mid \mathcal{F}_i] = 0$ and $|\eta_i| \le c_i$, where $c_i > 0$ are constants. For $1 \le t_1 \le t_2$ and $z \ge 0$,*

$$\mathbb{P}\left(\max_{t_1 \le j \le t_2} \sum_{i=t_1}^{j} \eta_i \ge z \,\middle|\, \mathcal{F}_{t_1}\right) \le \exp\left(-\frac{z^2}{2\sum_{i=t_1}^{t_2} c_i^2}\right).$$

The same bound holds for $\min_{t_1 \le j \le t_2} \sum_{i=t_1}^{j} \eta_i \le -z$ and for either tail of the terminal sum $\sum_{i=t_1}^{t_2} \eta_i$.

Lemma 17 (Chain rule for relative entropy)**.** *Let P, Q be probability laws of $(Y_1, \dots, Y_t)$ with regular conditional laws $P_i(\cdot \mid Y_1, \dots, Y_{i-1})$ and $Q_i(\cdot \mid Y_1, \dots, Y_{i-1})$, respectively. Then*

$$D_{\mathrm{KL}}(P\|Q) = \sum_{i=1}^{t} \mathbb{E}_P\big[D_{\mathrm{KL}}\big(P_i(\cdot \mid Y_1, \dots, Y_{i-1})\|Q_i(\cdot \mid Y_1, \dots, Y_{i-1})\big)\big],$$

where the $i = 1$ term uses the marginal laws of Y_1.

Lemma 18 (Pinsker's inequality)**.** *For probability measures P, Q on the same measurable space, define $\|P - Q\|_{\mathrm{TV}} = \sup_{\mathcal{E}} |P(\mathcal{E}) - Q(\mathcal{E})|$, where the supremum is over measurable events. For every measurable function f with $0 \le f \le 1$,*

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \le \|P - Q\|_{\mathrm{TV}} \le \sqrt{D_{\mathrm{KL}}(P\|Q)/2}.$$

F EXPERIMENTAL DETAILS

Experimental setup. We use $d = 5$, $K = 10$, $\lambda = 0.6$, $T = 2000$, and one parameter $\boldsymbol{\theta}^*$ with $\|\boldsymbol{\theta}^*\|_2 = 10$, shared by all servers and settings. Each job has K independently sampled features, and jobs are independent. The rows of Figure 1 use $q = \mathbb{P}(p^*(X) < \lambda) = 0, 0.15, 0.50$, and the columns use $\epsilon = 0.05, 0.1, 0.15$.

For each server, we sample a service probability from a low interval with probability $q^{1/K}$ and a high interval otherwise. The low intervals are $[0.57, 0.59]$ for $q = 0.15$ and $[0.54, 0.58]$ for $q = 0.50$. Each high interval has width 0.02 and lies above λ. Within each interval, we sample uniformly from 256 equally spaced midpoints. Thus $\mathbb{P}(p^*(X) < \lambda) = q$. We choose the high-interval center by bisection so that $\lambda\mathbb{E}[1/p^*(X)] = 1 - \epsilon$.

For each sampled probability p, set

$$\boldsymbol{\phi} = \frac{\log(p/(1-p))}{10}\boldsymbol{u} + \sqrt{25 - \frac{\log^2(p/(1-p))}{100}}\boldsymbol{v}, \qquad \boldsymbol{u} = \boldsymbol{\theta}^*/10,$$

where $\boldsymbol{v}$ is a unit vector orthogonal to $\boldsymbol{u}$. Then $\|\boldsymbol{\phi}\|_2 = 5$ and $\mu(\langle\boldsymbol{\phi}, \boldsymbol{\theta}^*\rangle) = p$. We draw $\boldsymbol{v}$ uniformly from 64 fixed directions.

Implementation. All five contextual algorithms use $\beta = 100$ and the constrained logistic MLE in (8) with $S = 10$. For widths and sampling, $\mathbf{V}_t = \mathbf{V}_0 + \sum_{i=1}^{t} \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top$, with $\mathbf{V}_0 = 100\mathbf{I}_5$ for CQB-Opt and CQB-ϵ, and $\mathbf{V}_0 = \mathbf{I}_5$ for WISE, CQB-η-2, and ACQB.

- **WISE** uses Algorithm 1 with probability bounds $[\mu(-50), \mu(50)]$.
- **CQB-Opt** (Bae et al., 2026a) maximizes $\mu(\langle\boldsymbol{\phi}(x,k), \widehat{\boldsymbol{\theta}}_{t-1}\rangle) + \beta\|\boldsymbol{\phi}(x,k)\|_{\mathbf{V}_{t-1}^{-1}}$ over waiting pairs.
- **CQB-ϵ** (Bae et al., 2026a) explores newly arrived jobs using round-robin servers for the first 200 rounds, then a uniformly random server with probability $T^{-1/2}$. Otherwise, it uses the CQB-Opt rule.
- **CQB-η-2** (Bae & Lee, 2026a) explores newly arrived jobs using a uniformly random server for the first 50 rounds and with probability 0.3 for the next 100 rounds. Otherwise, it uses the CQB-Opt rule.
- **ACQB** (Bae et al., 2026b) draws five parameter samples from $\mathcal{N}(\widehat{\boldsymbol{\theta}}_{t-1}, \beta^2\mathbf{V}_{t-1}^{-1})$, shared across all pairs, and maximizes the sampled service probability. After an arrival, with probability $(t+1)^{-1/2}$, it instead serves the new job with a round-robin server.
- **Q-UCB and Q-ThS** (Krishnasamy et al., 2021) serve the earliest arrival. Q-UCB maximizes the empirical success rate plus $\sqrt{\log^2(\max\{t,2\})/(2N_k)}$, where N_k counts attempts on server k and unobserved servers have infinite index. Q-ThS maximizes a draw from each server's Beta posterior with $\mathrm{Beta}(1,1)$ prior. Both choose a uniformly random server with probability $\min\{1, 3K\log^2(\max\{t,2\})/t\}$.
- **Random** chooses a uniformly random waiting job and server.

Results. We use ten runs with initially empty queues. Within each run, all algorithms and the oracle share arrivals, features, and service uniforms. The parameter and feature distributions are fixed across runs. Figure 1 reports the mean of $Q^\pi(t) - Q^*(t)$ and ± 2 sample standard deviations, each smoothed with a 25-round moving average.

USE OF GENERATIVE-AI TOOLS

The authors developed the workload potential and the analysis based on the elliptical potential count. GPT-5.6 Sol assisted the authors in designing Algorithm 1 and establishing the properties required for this analysis. It also assisted with manuscript polishing, reference exploration, and simulation code development. The authors take responsibility for the final arguments and presentation.