Adaptive service rate control of $M/M/1$ queue with breakdowns

YI ZHENG, JUXIHONG JULAITI, AND GUODONG PANG

ABSTRACT. We study service rate control problems for the $M/M/1$ queue with breakdowns in which the breakdown rate is assumed to be a function of the service rate. Assuming that the queue has infinite capacity, we first establish optimality equations for the discounted cost problem and characterize the optimal rate control policies. Then, we study optimality inequities for the average-cost problem and characterize the ergodicity of the queue. By establishing optimality equations, we also study the ergodic control problem when the queue has a finite capacity.

We also consider the adaptive service rate control problem for the model with a finite capacity. We assume that the relationship between the server breakdown rate and the service rate is linear with unknown parameters. Studying this problem is useful because the relationship between the server breakdown rate and the service rate is costly to observe in practice. We propose an adaptive (self-tuning) control algorithm, and prove that the regret vanishes under the algorithm and the proposed policies are asymptotically optimal. In addition, numerical studies are conducted to validate the algorithm.

1. Introduction

In this paper, we study service rate control problems for a single-server queue with Poisson arrivals, exponential service times, and server breakdowns. The server availability is modeled by a random process with ‘up’ and ‘down’ states. The system functions normally in the ‘up’ state, while the server stop serving customers in the ‘down’ state. The sojourn time that the server stays in the ‘down’ state follows an exponential distribution with a constant parameter. The time for the server from ‘up’ to ‘down’ obeys an exponential distribution with a parameter depending on the service rate. The control problems for the system with infinite or finite capacity have been addressed in this paper. The controller is allowed to choose a service rate at each state during the ‘up’ times.

The objective is to minimize the total cost, which has an effort cost that increases with the service rate, a holding/delay cost associated with the system congestion, a maintenance cost that occurs during the down state and a rejection cost for the system with finite capacity.

It is well known that any effort to reduce unplanned downtime can create considerable savings in industries; see, for example, [1]. Many queueing models that have been developed and analyzed include interruptions, such as the models in [2–7]. In many applications, the breakdown rate of a system may be closely related to its service rate. To the best of our knowledge, the queueing model with a service breakdown rate depending on the service rate has not been developed previously.

The service rate control problems for a single-server queue with infinite capacity have been extensively addressed. The closely related papers are [8–13]. In [10], the authors study service rate control problem of the long-run average cost for a single-server queue. The service rate control problem of a single-server queue with a Markov-modulated Poisson arrival process has been studied in [12]. Badian-Pessot et al. [13] address the control problem under the average cost criterion for $M/M/1$ queue, where the controller is allowed to remove the server and adjust the service rate when the server is turned on.

The Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, College of Engineering, Pennsylvania State University, University Park, PA 16802

E-mail addresses: yxz282, jpj5196, gup3@psu.edu.

Key words and phrases. single-server queue, server breakdowns, service rate control, adaptive (self-tuning) control, Markov decision process, discounted and long-run average (ergodic) cost criteria.
We first consider the service rate control problem under the discounted cost and long-run average cost criteria for a queue with infinite capacity. The controlled process is identified as a Markov decision process. For the discounted cost problem, we show the optimality equations for the optimal value function and the existence of optimal controls in Theorem 2.1. Since the breakdown rate for the server depends on the service rate, the characterization equation (2.4) for optimal controls contains an additional term corresponding to the breakdown rate, and the dynamic equations for the optimal values have two different forms for ‘up’ and ‘down’ states, respectively. Furthermore, under the assumptions of the cost functions and assuming the convexity and monotonicity of the breakdown rate function, we provide a representation of the optimal service rate controls in Theorem 2.2. Because of the dependent structure between the service rate and the breakdown rate, the property that the optimal service rate is monotone with respect to the number of jobs in the system (see [9, 10]) does not hold in general for the queue in this paper. For a system with a high congestion level and assuming that the breakdown rate is nondecreasing in the service rate, the controller may run the system at a low service level since a breakdown may cause more waiting jobs and delay costs (and lost jobs in the case of finite capacity). In this study, assuming that the relationship between the breakdown and service rates is linear, we provide a sufficient condition that the optimal service rate is nondecreasing in the number of jobs in the system. For the service rate control problem under the long-run average cost, we apply the vanishing discount approach in the spirit of [14]. The results for the long-run average cost problem are given in Theorem 2.3. Its proof relies on the stability condition of the joint Markov process consisting of the state and server availability processes. A necessary and sufficient condition related to the effective service rate for the stability of the joint Markov process is given in Proposition 2.4. The transition rate matrix results for the quasi-birth-and-death process play an important role in the proof of the stability. Similarly, by applying the vanishing discounted method, we provide a sufficient condition for the optimal service rate in Propositions 2.3 for the long-run average cost problem.

In [10], the optimal value under the long-run average cost criterion for $M/M/1$ queue with infinite capacity is approximated by solving a sequence of problems with the truncated holding cost function. They show that the optimal policies of the approximating problems converge monotonically to the optimal policy of the original problem. Their proof crucially relies on the monotonicity property of the optimal controls. For $M/M/1$ queue with breakdown, because the monotonicity of the optimal controls may not hold, it remains open to study the approximation of the original service rate control problem with truncated holding costs. Another natural approach for the approximation of the service rate control problem with infinite capacity is to use a sequence of problems with the truncated state space; see [15]. Adusumilli and Hasenbein [15] construct a sequence of approximation policies and show that the limiting policy exists and is optimal for the control problem with infinite capacity. Their proof for the convergence of approximating problems also relies on the monotonicity of the optimal controls in the number of jobs in the system.

For a single server queue with finite capacity, the service rate control problem under the ergodic cost has been studied in [15, 16]. In this paper, we study the service rate control problem under the ergodic cost with breakdowns and finite capacity, where the relation between the breakdown rate and the service rate is assumed to be a general continuous function and the action space is assumed to be a compact set. The results of the characterization for optimal controls are stated in Theorem 3.1. We consider a joint Markov process associated with the state process and the background process related to ‘up’ and ‘down’ states. Because the joint Markov process has finite states, we prove Theorem 3.1 by using the property of the stationary distribution of the process. This approach was previously used in [15, 16]. However, in this paper, the construction of the verification equations, which depend on the background process, is different from those in [15, 16]. Then, because we assume a compact action space, the uniqueness of the solution to the system of equations is presented in Lemma 3.2. Its proof, which uses contraction, is much similar in spirit to the method in [8]. The result in this lemma is used in the study of the adaptive control problem.
In practice, the relationship between the service rate and the breakdown rate of the system may not be known and may change over time. In an online learning framework, the aforementioned optimal control problem with finite capacity and an unknown relationship becomes an adaptive control problem. The adaptive control problems of Markov chains have been studied in [17–19]. They consider finite-state controlled Markov chains whose transition probabilities depend on an unknown parameter. It is shown in [19] that the long-run average costs under the adaptive controls converge to the optimal cost of the true model almost surely. For the long-run average (ergodic) cost criterion, the adaptive control problem of diffusion with an unknown parameter in the drift has been addressed in [20,21], in which the almost sure convergence for optimality is studied under stability conditions and using maximum likelihood estimation. We refer readers to [22,23] for the study of adaptive control for some stochastic systems. To the best of our knowledge, the adaptive control problem for general continuous-time Markov processes has not been studied in the literature.

In this paper, we consider the adaptive control problem for $M/M/1$ queue with breakdowns and finite capacity, where the relationship between the service rate and the breakdown rate depends on unknown parameters taking values in a compact set. This implies that the transition rate matrix of the Markov process is specified up to the unknown parameters. For the sake of simplicity, we assume that the breakdown rate is a linear function of the service rate with two unknown parameters. We apply a self-tuning approach to the adaptive control problem. The self-tuning scheme is introduced in [24]. Mandl [24] provides several models for controlled Markov processes with unknown parameters. The self-tuning approach is identified as a procedure that the controlled policy is continuously modified based on the estimation of unknown parameters to approach the optimal policy for the problem with true parameters. In this paper, we estimate the unknown parameters in the relation between the breakdown and service rates based on the historical data at each jump time of the Markov process, and then the control implemented is characterized by an optimality equation with the current parameter estimate. The optimality equation used is the same as the equation used in the aforementioned optimal control problem when the true parameters are known. Since the linear relation between the service and breakdown rates leads to a nonlinear relation between the mean service times and mean ‘up’ times of the system, the quasi-maximum likelihood estimates are used for the estimation of unknown parameters. This method has been studied in [25,26] and references therein. Asymptotic optimality under work-conserving Markov rate control policies, which results in uniform ergodicity for the joint Markov process of the state and background processes, is established in Theorem 4.1.

We also evaluate the performance of the method by conducting numerical analysis. For the system dynamics, we consider three scenarios, in which the proportions of up times and the rejection probabilities in the long run are different. We plot the functions for optimal service rates of congestion levels in these scenarios. Cost parameters are chosen to cover different cases in practice. We observe that optimal service rates may not be monotone in the number of jobs in the system. This is different from the monotonicity of optimal service rates for an $M/M/1$ queue in [10,15]. For the adaptive service rate control problem, we conduct simulation experiments and show that the proposed policies converge to the optimal service rate control policy.

It is worth mentioning that online learning problems are also related to adaptive control problems, since estimation methods for unknown parameters are used in both types of problems. When the unknown parameters are fitted statistically, the regret of the algorithm is commonly used to measure its performance [27]. In this paper, we show that the regret of the proposed algorithm vanishes and the proposed policies are asymptotically optimal. Blackwell’s Approachability Theorem [28] provides conditions to analyze the regret of online learning algorithms [29,30]. Online problems for demand models have been addressed extensively; see, for example, [31–34]. In [33], the authors consider the unknown demand model, where the demand for products is assumed to be linear with unknown parameters. They provide modified greedy iterated least-squares policies to achieve asymptotic optimality. However, studies on online problems for queueing systems
are scarce. Recently, Chen et al. [35] study the dynamic pricing and capacity sizing problem in GI/GI/1 queue. They develop an online stochastic gradient descent method and show the regret bound for the convergence. Applications of reinforcement learning to queueing systems are not new; see, for example, [36–38]. In a recent work, Liu et al. [37] propose a model-based learning algorithm to a server allocation and routing problem in a queueing system. They show that the proposed algorithm can obtain the optimal policy using a Lyapunov analysis. However, the convergence of reinforcement learning is not well understood [27] and there are limited studies on service rate control with an unreliable server. In this paper, we consider adaptive service rate control problems for M/M/1 queue with breakdowns, where the optimal controls do not have closed-form representations and are characterized by a system of Bellman equations.

1.1. Organization of the paper. The paper is organized as follows. Section 2.1 contains a detailed description of the M/M/1 queueing model with service breakdowns. After defining control policies considered in this study, we state the assumptions and describe the system dynamics. In Section 2.2, we establish optimality equations for the discounted cost problem. The characteristics of the value function and cost functions are then presented. The optimality equations of the ergodic problems are stated in Section 2.3, followed by the properties of the optimal controls. In Section 3, we define an ergodic control problem when the queue has a finite capacity and show the optimality of the controls. In Section 4, we present the results of the adaptive control problem for a system with finite capacity when the relationship between the server breakdown rate and the service rate is a linear function and the parameters for the function are unknown. Finally, numerical examples of the adaptive control problem are presented in Section 5.

2. M/M/1 QUEUE WITH INFINITE CAPACITY AND BREAKDOWNS

2.1. The model description. We consider an M/M/1 queue with adjustable service rate and service breakdowns. Let \{A(t)\}_{t \geq 0} denote the Poisson arrival process with arrival rate \( \lambda > 0 \). The queueing system is in an up-down environment. In the ‘up’ state, the system functions normally and the controller chooses a dynamic service rate \( \mu \) from the compact set \( U := [0, \bar{\mu}] \) with \( 0 < \bar{\mu} < \infty \). In the ‘down’ state, the server stops, while jobs keep joining the queue and any job in service will wait for the system to resume.

Let \{X(t)\}_{t \geq 0} and \{K(t)\}_{t \geq 0} denote the number of jobs in the system (including those in queue and in service, either interrupted or not) and the server availability process, respectively. At time \( t \geq 0 \), \( K(t) = 1 \) if the system is in the “up” state, and \( K(t) = 0 \) otherwise. We say that a rate control policy \( U \) is admissible if it is non-anticipative, takes values in \( U \), and satisfies that \( U(t) = 0 \) if \( K(t) = 0 \) for \( t \geq 0 \). The set of admissible rate control policies is denoted by \( \Omega \). Let \( \mathbb{Z}_+ \) denote the set of nonnegative integers. An admissible rate control policy \( U \) is called stationary Markov if

\[
U(t) = \nu(X(t), K(t)) \quad t \geq 0,
\]

for some \( \nu: \mathbb{Z}_+ \times \{0, 1\} \mapsto U \) satisfying \( \nu(0, 1) \equiv 0 \) and \( \nu(x, 0) \equiv 0 \) for \( x \in \mathbb{Z}_+ \). The set of stationary Markov control policies is denoted by \( \Omega_{\text{sm}} \).

Let \( \mathbb{R}_+ \) denote the set of nonnegative real numbers. We assume that given \( U \in \Omega \), \{K(t)\}_{t \geq 0} is a continuous-time Markov process with state space \( \{0, 1\} \), and its transition rate matrix is given by

\[
\begin{bmatrix}
-\beta_d & \beta_d \\
\beta_u(U(t)) & -\beta_u(U(t))
\end{bmatrix}
\]

for \( t \geq 0 \), where \( \beta_u: U \mapsto \mathbb{R}_+ \) is a measurable function, and \( \beta_d \) is a positive constant. \( \beta_u \) and \( \beta_d \) represent the breakdown (from “up” to “down”) and repair (from “down” to “up”) rates, respectively. This implies that the repairment of the server is started immediately when the server stops. The repair rate is usually assumed to be a constant, see, for example, Section 2 of [39], where they assume a constant mean repair time. In practice, the functioning time of a server, for
example, a machine, depends on its effort. Thus, the breakdown rate is assumed to be a function of service rate. In addition, we assume that $\beta_u$ is strictly positive and continuously differentiable. Given $\nu \in \mathfrak{U}_{\text{sm}}$ and $X(0)$, the state process $\{X(t)\}_{t \geq 0}$ evolves as the following

$$X(t) = X(0) + A(t) - S \left( \int_0^t \nu(X(s), K(s))(X(s) \wedge 1) \, ds \right) \quad \forall t \geq 0,$$  

(2.1)

where $\{S(t)\}_{t \geq 0}$ is a unit rate Poisson process, independent of the arrival process $\{A(t)\}_{t \geq 0}$. Note that provided $\nu \in \mathfrak{U}_{\text{sm}}$, $\{(X(t), K(t))\}_{t \geq 0}$ is a well-defined Markov process.

The costs of our optimization problems consist of the effort, holding/delay and repair costs. The effort cost function is denoted by $R$, that is, the cost rate is $R(\mu)$ per unit time when a service rate $\mu \in \mathbb{U}$ is selected. We assume that $R$ is strictly increasing, continuously differentiable, and such that $R(0) = 0$. The holding cost function is defined by $H(x)$ for $x \in \mathbb{Z}_+$, which is assumed to be convex and nondecreasing. We also assume that during the down times, the system incurs a cost at a positive constant rate $C_m$. The total cost function is defined by

$$f(x, k, \mu) := R(\mu) + H(x) + C_m(1 - k)$$

(2.2)

for $(x, k, \mu) \in \mathbb{Z}_+ \times \{0, 1\} \times \mathbb{U}$. In the next two sections, we consider the discounted and long-run average cost minimization problems.

### 2.2. The discounted cost problem

In this subsection, we present the results for the service rate control problem under the discounted cost criterion. We study the optimality equations and show the properties of the optimal controls.

For $U \in \mathfrak{U}$, the $\alpha$-discounted cost criterion is given by

$$J_{\alpha}^{U}(x, k) := \mathbb{E}_{x, k}^{U} \left[ \int_{0}^{\infty} e^{-\alpha s} f(X(s), K(s), U(s)) \, ds \right] \quad \forall \alpha > 0.$$  

(2.3)

The optimal $\alpha$-discounted value function is denoted by

$$V_{\alpha}(x, k) := \min_{U \in \mathfrak{U}} J_{\alpha}^{U}(x, k) \quad \forall \alpha > 0.$$  

(2.4)

We say that a control $U^* \in \mathfrak{U}$ is optimal if $J_{\alpha}^{U^*} = V_{\alpha}$.

In the next theorem, we show that a stationary optimal control for the $\alpha$-discounted problem exists, and that $V_{\alpha}$ is the solution of optimality equations. We first define

$$\phi(w, y) := \max_{\mu \in \mathbb{U}} \{\mu - \beta_u(\mu)y - R(\mu)\} \quad \forall (w, y) \in \mathbb{R} \times \mathbb{R}.$$  

(2.5)

**Theorem 2.1.** There exists an optimal control policy $\nu_{\alpha}^* \in \mathfrak{U}_{\text{sm}}$ for the $\alpha$-discounted problem (2.3). The function $V_{\alpha}$ satisfies the following discounted cost optimality equations

$$V_{\alpha}(0, 1) = \frac{1}{\alpha + M} \left( H(0) + \lambda V_{\alpha}(1, 1) + \beta_u(0)V_{\alpha}(0, 0) + (M - \lambda - \beta_u(0))V_{\alpha}(0, 1) \right),$$

$$V_{\alpha}(x, 1) = \frac{1}{\alpha + M} \left( H(x) - \phi(W_{\alpha}(x, 1), Y_{\alpha}(x)) + \lambda V_{\alpha}(x + 1, 1) + (M - \lambda)V_{\alpha}(x, 1) \right)$$

(2.6)

for $x \in \mathbb{N}$, and

$$V_{\alpha}(x, 0) = \frac{1}{\alpha + M} \left( H(x) + C_m + \lambda V_{\alpha}(x + 1, 0) + \beta_d V_{\alpha}(x, 1) + (M - \lambda - \beta_d)V_{\alpha}(x, 0) \right)$$

for $x \in \mathbb{Z}_+$, where $\mathbb{N}$ denotes the set of natural numbers, $M := \bar{\mu} + \lambda + \beta_d + \beta_u(\bar{\mu})$.

$$W_{\alpha}(x, 1) := V_{\alpha}(x, 1) - V_{\alpha}(x - 1, 1) \quad \text{and} \quad Y_{\alpha}(x) := V_{\alpha}(x, 0) - V_{\alpha}(x, 1).$$  

(2.7)
Proof. We use the uniformization technique to prove this theorem (see, for example, [9, 12]). Let $V_{n,\alpha}(x, k)$ be the optimal $\alpha$-discounted expected value obtained during the last $n$ transitions starting from the state $(x, k)$. We assume $V_{0,\alpha}(x, 0) = V_{0,\alpha}(x, 1) = 0$ for all $x \in \mathbb{Z}_+$. 

The recursive formula for $V_{n,\alpha}$ is given by

\begin{align}
V_{n+1,\alpha}(0, 1) &= \frac{1}{\alpha + M} \left( H(x) + \lambda V_{n,\alpha}(1, 1) + \beta_u(0)V_{n,\alpha}(0, 0) \right. \\
&\quad + \left. (M - \lambda - \beta_d(0))V_{n,\alpha}(0, 1) \right), \tag{2.7}
\end{align}

\begin{align}
V_{n+1,\alpha}(x, 1) &= \frac{1}{\alpha + M} \min_{\mu \in \mathcal{U}} \{ \mathcal{R}(u) + H(x) + \mu V_{n,\alpha}(x-1, 1) + \lambda V_{n,\alpha}(x+1, 1) \}
&\quad + \beta_u(\mu)V_{n,\alpha}(x, 0) + (M - \lambda - \beta_d(\mu))V_{n,\alpha}(x, 1) \tag{2.8}
\end{align}
for $x \in \mathbb{N}$, and

\begin{align*}
V_{n+1,\alpha}(x, 0) &= \frac{1}{\alpha + M} \left( H(x) + C_m + \lambda V_{n,\alpha}(x+1, 0) + \beta_d V_{n,\alpha}(x, 1) + (M - \lambda - \beta_d) V_{n,\alpha}(x, 0) \right) \tag{2.9}
\end{align*}
for $x \in \mathbb{Z}_+$. Let

$W_{n,\alpha}(\cdot, k) := V_{n,\alpha}(\cdot, k) - V_{n,\alpha}(\cdot, 0)$ and $Y_{n,\alpha}(\cdot) := V_{n,\alpha}(\cdot, 0) - V_{n,\alpha}(\cdot, 1)$ \tag{2.10}

for $k \in \{0, 1\}$. Then, (2.9) takes the form

\begin{align}
V_{n+1,\alpha}(x, 1) &= \left( H(x) - \phi(W_{n,\alpha}(x, 1), Y_{n,\alpha}(x)) + \lambda V_{n,\alpha}(x+1, 1) + (M - \lambda) V_{n,\alpha}(x, 1) \right). \tag{2.11}
\end{align}

Applying [40, Proposition 3.1 (iii)], taking $n \to \infty$ in (2.7), (2.9), and (2.11) yields the equations in the statement. By part (ii) of Proposition 3.1 of [40], there exists an optimal control $\nu_{\alpha}^n \in \mathcal{U}_{\text{sm}}$. This completes the proof. \hfill \Box

The next two propositions are used to derive the characterization of the optimal controls. In the following proposition, we show that the $\alpha$-discounted value functions are nondecreasing in the number of jobs in the system.

**Proposition 2.1.** For $k \in \{0, 1\}$, the value function $V_{\alpha}(\cdot, k)$ is a nondecreasing function.

Proof. We prove this result by induction. Recall $V_{n}(x, 1)$ and $V_{n}(x, 0)$ given in (2.8) and (2.9), respectively. This result trivially holds when $n = 0$. Suppose that for $k \in \{0, 1\}$, $V_{n,\alpha}(\cdot, k)$ is a nondecreasing function. We let $\nu_{\alpha}^{n+1}$ be an optimal stationary Markov control of $(n+1)$-stage problem in (2.8). For the notational convenience, we denote

\begin{align}
\mu_x := \nu_{\alpha}^{n+1}(x, 1). \tag{2.12}
\end{align}

Without loss of generality, we assume $\mu_0 = 0$. It follows by the inductive hypotheses and equations (2.7)–(2.9) that for $x \in \mathbb{N}$,

\begin{align}
V_{n+1,\alpha}(x-1, 1) &\leq \frac{1}{\alpha + M} \left( H(x-1) + \mathcal{R}(\mu_x) + \mu_x V_{n,\alpha}(x-2, 1) + \lambda V_{n,\alpha}(x, 1) \right. \\
&\quad + \left. \beta_u(\mu_x)V_{n,\alpha}(x-1, 0) + (M - \lambda - \mu_x - \beta_d(\mu_x))V_{n,\alpha}(x-1, 1) \right) \tag{2.13}
\end{align}

and

\begin{align}
V_{n+1,\alpha}(x-1, 0) &\leq \frac{1}{\alpha + M} \left( C_m + H(x-1) + \lambda V_{n,\alpha}(x, 0) \right. \\
&\quad + \left. \beta_d V_{n,\alpha}(x, 1) + (M - \lambda - \beta_d) V_{n,\alpha}(x, 0) \right) \tag{2.14}
&\quad \leq V_{n+1,\alpha}(x, 0).
\end{align}

The result follows by taking $n \to \infty$. \hfill \Box
In the next proposition, we show that if the repair cost is higher than the effort cost, then the \( \alpha \)-discounted cost starting from ‘down’ state is higher than that starting from ‘up’ state.

**Proposition 2.2.** Assume \( C_m \geq \mathcal{R}(\hat{\mu}) \). Then, \( V_\alpha(x,0) \geq V_\alpha(x,1) \), for \( x \in \mathbb{Z}_+ \).

**Proof.** We use induction. It is evident that the result holds for \( V_{0,\alpha} \). Suppose \( V_{n,\alpha}(\cdot,0) \geq V_{n,\alpha}(\cdot,1) \). By taking the difference of (2.7) and (2.9), and using the inductive hypothesis, the result trivially holds when \( x = 0 \) for \((n+1)\)-stage problem. Recall \( \mu_x \) in (2.12), and \( W_{n,\alpha} \) and \( Y_{n,\alpha} \) in (2.10). By using (2.8) and (2.9), we obtain

\[
(\alpha + M)Y_{n+1,\alpha}(x + 1) = C_m - \mathcal{R}(\mu_{x+1}) + \mu_{x+1}W_{n,\alpha}(x + 1,1) + \lambda Y_{n,\alpha}(x + 2) + (M - \lambda - \beta_u(\mu_{x+1}) - \beta_d)Y_{n,\alpha}(x + 1)
\]

\[
\geq 0,
\]

where the inequality follows by (2.13), the inductive hypothesis, and \( C_m \geq \mathcal{R}(\hat{\mu}) \). By taking \( n \to \infty \), the result follows. \( \square \)

**Assumption 2.1.** In addition to the assumptions stated in Section 2.1, we assume that the function \( \beta_u \) is nondecreasing and convex, \( \mathcal{R} \) is strictly convex, and \( C_m \geq \mathcal{R}(\hat{\mu}) \).

The assumption of \( \beta_u \) in Assumption 2.1 implies that the system is more likely to breakdown when the service rate is at high level.

**Theorem 2.2.** Grant Assumption 2.1. Then, the results in Theorem 2.1 hold, and there exists an optimal control \( \nu^*_\alpha \in U_{\text{sm}} \) for the \( \alpha \)-discounted problem such that

\[
\nu^*_\alpha(x,1) = \psi(W_\alpha(x,1),Y_\alpha(x)),
\]

where the function \( \psi \) is the maximizer of \( \phi \) in (2.4) and satisfies

\[
\psi(w,y) = \begin{cases} 0, & \text{for } w \leq \mathcal{R}'(0) + \beta_u'(0)y, \\ (y\beta_u' + \mathcal{R})^{-1}(w) & \text{for } \mathcal{R}'(0) + \beta_u'(0)y < w \leq \mathcal{R}'(\hat{\mu}) + \beta_u'(\hat{\mu})y, \\ \hat{\mu} & \text{for } w > \mathcal{R}'(\hat{\mu}) + \beta_u'(\hat{\mu})y. \end{cases}
\]

**Proof.** Since \( \mathcal{R} \) is a strictly convex function, then by the assumption that \( \beta_u \) is convex, the set of maximizers of \( \phi \) becomes a singleton if \( y \geq 0 \). But it follows by Proposition 2.2 that \( Y_{\alpha}(x) \geq 0 \) for any \( x \in \mathbb{Z}_+ \). Using Assumption 2.1, \((y\beta_u' + \mathcal{R})^{-1}\) is continuous and strictly increasing. Thus, (2.15) holds, and the rest of the proof is the same as that of Theorem 2.1. \( \square \)

**Remark 2.1.** The following special case is frequently used in the rest of paper. If

\[
\beta_u(\mu) = \kappa_1 + \kappa_2\mu
\]

for some positive constants \( \kappa_1 \) and \( \kappa_2 \), then

\[
\psi(w,y) = \begin{cases} 0, & \text{for } w - \kappa_2y \leq \mathcal{R}'(0), \\ (\mathcal{R}')^{-1}(w - \kappa_2y) & \text{for } \mathcal{R}'(0) < w - \kappa_2y \leq \mathcal{R}'(\hat{\mu}), \\ \hat{\mu} & \text{for } w - \kappa_2y > \mathcal{R}'(\hat{\mu}). \end{cases}
\]

From (2.15), we observe that the optimal control policy may not be monotone in the queue length. In the following assumption, we provide a sufficient condition under which a monotonically optimal control exists.

**Assumption 2.2.** We assume that (2.17) holds, \( \mathcal{R} \) is strictly convex, \( C_m \geq \mathcal{R}(\hat{\mu}) \) and

\[
\beta_d \geq (\kappa_2 + 1)\hat{\mu}.
\]
The inequality (2.19) implies that the repair rate is higher than the sum of the maximum service rate and the increment of the breakdown rate. In the following proposition, we provide the structural properties of the optimal policies. We show that if the repair cost is not less than the highest cost of effort and the repair rate satisfies (2.19), then there exists an optimal control that is nondecreasing in queue length.

**Proposition 2.3.** Grant Assumption 2.2. Then, the following hold.

(i) For \( k \in \{0, 1\} \), \( V_\alpha(\cdot, k) \) is a convex function.

(ii) \( Y_\alpha \) is a nondecreasing function.

(iii) There exists an optimal Markov control policy \( \nu_\alpha^* \in \mathcal{U}_\text{adm} \) such that \( \nu_\alpha^*(\cdot, 1) \) is a nondecreasing function.

**Proof.** Recall \( V_{n,\alpha}(\cdot, k) \), \( k \in \{0, 1\} \), in (2.8) and (2.9), respectively, and \( \{W_{n,\alpha}(\cdot, k): k \in \{0, 1\}\} \) and \( Y_{n,\alpha} \) in (2.10). Note that if \( W_{n,\alpha}(x + 1, k) - W_{n,\alpha}(x, k) \geq 0 \) for \( x \in \mathbb{Z}_+ \), then \( V_{n,\alpha}(x, k) \) is a convex function. We define \( Z_{n,\alpha}(x) := W_{n,\alpha}(x, 1) - \kappa_2 Y_{n,\alpha}(x) \) for \( x \in \mathbb{N} \). We choose

\[
\mu^n_x := \psi(W_{n,\alpha}(x, 1), Y_{n,\alpha}(x)).
\]

Then, \( \mu^n = \{\mu^n_x: x \in \mathbb{N}\} \) is optimal for \((n + 1)\)-stage problem in (2.8). Applying (2.18), and the assumption that \( R \) is strictly increasing, it is straightforward to check that \( \mu^n_x \) is nondecreasing in \( x \) if \( Z_{n}\) is nondecreasing.

We prove (i)–(iii) by induction. For \( n = 0 \), it is evident that \( W_{0,\alpha}(\cdot + 1, k) - W_{0,\alpha}(\cdot, k) \equiv 0 \) for \( k \in \{0, 1\} \), and \( Y_{0,\alpha} \) and \( Z_{0,\alpha} \) are nondecreasing functions. For a fixed \( n \in \mathbb{N} \), suppose that \( \{W_{n,\alpha}(\cdot + 1, k) - W_{n,\alpha}(\cdot, k): k \in \{0, 1\}\} \) are nonnegative functions, and \( Y_{n,\alpha} \) and \( Z_{n,\alpha} \) are nondecreasing functions. We first consider the case when the initial state of the system is ‘down’. By (2.9), we have

\[
(\alpha + M)W_{n+1,\alpha}(x + 1, 0) = H(x + 1) - H(x) + \lambda W_{n,\alpha}(x + 2, 0) + \beta_d W_{n,\alpha}(x + 1, 1) + (M - \lambda - \beta_d)W_{n,\alpha}(x + 1, 0),
\]

and thus

\[
(\alpha + M)(W_{n+1,\alpha}(x + 1, 0) - W_{n+1,\alpha}(x, 0)) = H(x + 1) - 2H(x) + H(x - 1) + \lambda W_{n,\alpha}(x + 2, 0) - W_{n,\alpha}(x + 1, 0) + \beta_d (W_{n,\alpha}(x + 1, 1) - W_{n,\alpha}(x, 1)) + (M - \lambda - \beta_d)(W_{n,\alpha}(x + 1, 0) - W_{n,\alpha}(x, 0)) \geq 0,
\]

where the second inequality follows by inductive hypotheses and convexity of \( H \). Since \( Z_{n,\alpha} \) is a nondecreasing function by inductive hypotheses, \( \mu^n_x \) is also nondecreasing in \( x \) by (2.18). Then, for the system starting from the ‘up’ state, it follows by (2.8) that

\[
(\alpha + M)W_{n+1,\alpha}(x + 1, 1) \geq H(x + 1) - H(x) - \mu^n_{x+1}(W_{n,\alpha}(x + 1, 1) - W_{n,\alpha}(x, 1)) + \lambda W_{n,\alpha}(x + 2, 1) + \beta_u(\mu^n_{x+1})W_{n,\alpha}(x + 1, 0) + (M - \lambda - \beta_u(\mu^n_{x+1}))W_{n,\alpha}(x + 1, 1),
\]

and

\[
(\alpha + M)W_{n+1,\alpha}(x, 1) \leq H(x) - H(x - 1) - \mu^n_{x-1}(W_{n,\alpha}(x, 1) - W_{n,\alpha}(x - 1, 1)) + \lambda W_{n,\alpha}(x + 1, 1) + \beta_u(\mu^n_{x-1})W_{n,\alpha}(x, 0) + (M - \lambda - \beta_u(\mu^n_{x-1}))W_{n,\alpha}(x, 1).
\]

By inductive hypotheses, we have \( Z_{n,\alpha}(x + 1) \geq Z_{n,\alpha}(x) \geq Z_{n,\alpha}(x - 1) \), an then it follows by (2.18) that

\[
\mu^n_{x+1} \geq \mu^n_x \geq \mu^n_{x-1}.
\]
By (2.22) and (2.23), we obtain
\[
(\alpha + M)(W_{n+1,\alpha}(x+1,1) - W_{n+1,\alpha}(x,1)) \\
\geq H(x+1) - 2H(x) + H(x-1) + \lambda(W_{n,\alpha}(x+2,1) - W_{n,\alpha}(x+1,1)) \\
+ (M - \lambda - \mu_{x+1}^n - \beta_u(\mu_{x-1}^n))(W_{n,\alpha}(x+1,1) - W_{n,\alpha}(x,1)) \\
+ \mu_{x-1}^n(W_{n,\alpha}(x,1) - W_{n,\alpha}(x-1,1)) + \beta_u(\mu_{x-1}^n)(W_{n,\alpha}(x+1,0) - W_{n,\alpha}(x,0)) \\
+ (\beta_u(\mu_{x+1}^n) - \beta_u(\mu_{x-1}^n))(W_{n,\alpha}(x+1,0) - W_{n,\alpha}(x+1,1)) \\
\geq 0,
\]
where the second inequality follows by inductive hypotheses and (2.24). Note that
\[
W_{n,\alpha}(x+1,0) - W_{n,\alpha}(x+1,1) = Y_{n,\alpha}(x+1) - Y_{n,\alpha}(x).
\]
Applying (2.14), we have
\[
(\alpha + M)(Y_{n+1,\alpha}(x+1) - Y_{n+1,\alpha}(x)) \\
\geq \mu_{x}^n(W_{n,\alpha}(x+1,1) - W_{n,\alpha}(x,1)) + \lambda(Y_{n,\alpha}(x+2) - Y_{n,\alpha}(x+1)) \\
+ (M - \lambda - \beta_u(\mu_{x}^n) - \beta_d)(Y_{n,\alpha}(x+1) - Y_{n,\alpha}(x)) \\
\geq 0,
\]
where the second inequality follows by inductive hypotheses. By (2.14) and (2.22), we obtain
\[
(\alpha + M)Z_{n+1,\alpha}(x+1) \\
\geq H(x+1) - H(x) + \kappa_2(\mathcal{R}(\mu_{x+1}^n) - C_m) - \kappa_2\mu_{x+1}^nW_{n,\alpha}(x+1,1) \\
+ \beta_u(\mu_{x+1}^n)W_{n,\alpha}(x+1,0) + \lambda Z_{n,\alpha}(x+2) + \kappa_2\beta_dY_{n,\alpha}(x+1) \\
- \mu_{x+1}^n(W_{n,\alpha}(x+1,1) - W_{n,\alpha}(x,1)) \\
+ (M - \lambda - \beta_u(\mu_{x+1}^n))(W_{n,\alpha}(x+1,1) - \kappa_2Y_{n,\alpha}(x+1)).
\]
Similarly, applying (2.14) and (2.23), we have
\[
(\alpha + M)Z_{n+1,\alpha}(x) \\
\leq H(x) - H(x-1) + \kappa_2(\mathcal{R}(\mu_{x}^n) - C_m) - \kappa_2\mu_{x+1}^nW_{n,\alpha}(x,1) + \beta_u(\mu_{x+1}^n)W_{n,\alpha}(x,0) \\
+ \lambda Z_{n,\alpha}(x+1) + \kappa_2\beta_dY_{n,\alpha}(x) - \mu_{x-1}^n(W_{n,\alpha}(x,1) - W_{n,\alpha}(x-1,1)) \\
+ (M - \lambda - \beta_u(\mu_{x-1}^n))Y_{n,\alpha}(x) - \kappa_2(M - \lambda - \beta_u(\mu_{x+1}^n))Y_{n,\alpha}(x).
\]
Recall the uniformization constant $M$ defined in (2.5). By taking the difference of the last two inequalities above, and applying inductive hypotheses and convexity of $H$, it follows that
\[
(\alpha + M)(Z_{n+1,\alpha}(x+1) - Z_{n+1,\alpha}(x)) \\
\geq \kappa_2(\mathcal{R}(\mu_{x+1}^n) - C_m) - \kappa_2(\mathcal{R}(\mu_{x}^n) - C_m) \\
+ \kappa_2(\beta_d - \kappa_2\mu_{x}^n - \mu_{x+1}^n)(Y_{n,\alpha}(x+1) - Y_{n,\alpha}(x)) \\
+ \kappa_2(\beta_u(\mu_{x+1}^n) - \beta_u(\mu_{x-1}^n))Y_{n,\alpha}(x+1) - \kappa_2(\beta_u(\mu_{x}^n) - \beta_u(\mu_{x-1}^n))Y_{n,\alpha}(x) \\
- (\kappa_2\mu_{x+1}^n - \kappa_2\mu_{x}^n)W_{n,\alpha}(x+1,1) \\
+ (M - \lambda - \beta_u(\mu_{x-1}^n) - \kappa_2\mu_{x}^n - \mu_{x+1}^n)(Z_{n,\alpha}(x+1) - Z_{n,\alpha}(x)) \\
\geq \kappa_2(\mathcal{R}(\mu_{x+1}^n) - \mathcal{R}(\mu_{x}^n) - \mathcal{R}(\mu_{x}^n) - \mathcal{R}(\mu_{x+1}^n) - \mathcal{R}(\mu_{x}^n) - \mathcal{R}(\mu_{x+1}^n))Z_{n,\alpha}(x+1) \\
+ (M - \lambda - \beta_u(\mu_{x-1}^n) - \kappa_2\mu_{x}^n - \mu_{x+1}^n)(Z_{n,\alpha}(x+1) - Z_{n,\alpha}(x)) \\
\geq \kappa_2\bar{\mu}(Z_{n,\alpha}(x+1) - Z_{n,\alpha}(x)) - \kappa_2(\mu_{x+1}^n - \mu_{x}^n)(Z_{n,\alpha}(x+1) - \mathcal{R}(\mu_{x}^n)),
\]
where we use the inductive hypotheses, (2.17), (2.19), and (2.24) in the second and third inequalities. It is evident that by (2.28), \( Z_{n+1,\alpha}(x+1) \geq Z_{n+1,\alpha}(x) \) when \( \mu_{x+1}^{n} = \mu_{x}^{n} \). Note that \( \mu_{x+1}^{n} - \mu_{x}^{n} \leq \bar{\mu} \).

If \( \mu_{x}^{n} = 0 \), then \( Z_{n,\alpha}(x) \leq R'(\mu_{x}^{n}) \) from (2.18). Similarly, \( Z_{n,\alpha}(x) \leq R'(\mu_{x}^{n}) \) when \( 0 < \mu_{x}^{n} < \bar{\mu} \). Thus, it follows by (2.28) that \( Z_{n+1,\alpha}(x+1) \geq Z_{n+1,\alpha}(x) \).

Therefore, by (2.21), (2.25), (2.27), and (2.28), and taking \( n \to \infty \), we have shown (i)-(iii). This completes the proof.

\[ \square \]

2.3. **The long-run average cost problem.** In this subsection, we first establish a necessary and sufficient condition for the stability of the joint Markov process. Under the stability condition, we show the existence and characterization of optimal controls for the long-run average expected cost problem by utilizing the vanishing discounted method.

Recall \( f \) in (2.2). Under \( U \in \mathcal{U} \), the expected long-run average cost criterion is given by

\[
\phi^{U}(x,k) := \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}_{x,k} \left[ \int_{0}^{T} f(X(s),K(s),U(s)) \, ds \right].
\]

The optimal expected long-run average cost is defined by

\[
\phi^{*}(x,k) := \inf_{U \in \mathcal{U}} \phi^{U}(x,k).
\]  

(2.29)

We say that a policy \( \nu \) is optimal for long-run average cost if \( \phi^{*}(x,k) = \phi^{\nu}(x,k) \). In Proposition 2.4, we show that \( \phi^{*} \) is finite and independent of \( (x,k) \) under the following assumption on the holding cost function. The finiteness of \( \phi^{*} \) implies that the average cost problem is well-posed. Similar assumptions on the holding cost function were also used in [10,12].

**Assumption 2.3.** The holding cost function \( H \) is subgeometric, that is,

\[
\sum_{n=0}^{\infty} H(n) \gamma^{n} < \infty, \quad \forall \gamma \in [0,1).
\]  

(2.30)

In the next proposition, we provide a sufficient and necessary condition for the long-run average cost to be finite. We say that a policy \( \nu \in \mathcal{U}_{\text{sm}} \) is stable if \( (X,K) \) is positive recurrent under \( \nu \).

**Proposition 2.4.** There exists a stable policy \( \nu \in \mathcal{U}_{\text{sm}} \) if and only if

\[
\min_{\mu \in \mathcal{U}} \left\{ \frac{\lambda(\beta_{u}(\mu) + \beta_{d})}{\beta_{d}\mu} \right\} < 1.
\]  

(2.31)

Moreover, under Assumption 2.3, there exists a stable policy \( \nu \in \mathcal{U}_{\text{sm}} \) such that \( \phi^{\nu} \) is finite and independent of \( (x,k) \).

**Proof.** Under \( \nu \in \mathcal{U}_{\text{sm}} \), the infinitesimal generator of \( (X,K) \) is given by

\[
Q := \begin{bmatrix}
Q_{0} & A \\
B_{1} & Q_{1} & A \\
B_{2} & Q_{2} & A \\
B_{3} & Q_{3} & A \\
\vdots & \ddots & \ddots \\
\end{bmatrix},
\]

where

\[
A := \begin{bmatrix}
\lambda & 0 \\
0 & \lambda \\
\end{bmatrix}, \quad Q_{0} := \begin{bmatrix}
-(\beta_{d} + \lambda) & \beta_{d} \\
\beta_{u}(0) & -(\beta_{u}(0) + \lambda) \\
\end{bmatrix},
\]

\[
B_{x} := \begin{bmatrix}
0 & 0 & \nu(x,1) \\
0 & \nu(x,1) \\
\end{bmatrix}, \quad Q_{x} := \begin{bmatrix}
-(\beta_{d} + \lambda) & \beta_{d} \\
\beta_{u}(\nu(x,1)) & -(\beta_{u}(\nu(x,1)) + \nu(x,1) + \lambda) \\
\end{bmatrix},
\]

with \( x \in \mathbb{N} \). We first prove the sufficiency. It is evident that under \( \nu \in \mathcal{U}_{\text{sm}} \), \( (X,K) \) is irreducible when \( \inf_{x \in \mathbb{N}} \nu(x,1) > 0 \). Let \( \mu_{o} \) be a minimizer of (2.31), and \( \nu(x,1) \equiv \mu_{o} \) for \( x \in \mathbb{N} \). Then,
the embedded Markov chain of the joint process \((X, K)\) becomes a discrete time level-independent quasi-birth-and-death process, see [41] for the detailed definition. We define
\[
Q := A + B_1 + Q_1.
\]
For \(x \in \mathbb{R}^d\), \(x^T\) denotes the transpose of \(x\). It is straightforward to verify that
\[
\eta := \left(\frac{\beta_u(\mu_o)}{\beta_u(\mu_o) + \beta_d}, \frac{\beta_d}{\beta_u(\mu_o) + \beta_d}\right)^T
\]
solves the system of equations \(\eta^T Q = 0\) and \(e^T \eta = 1\). Then, it follows by [41, Theorem 3.2.1] that under \(\nu \in \mathcal{U}_{\text{adm}}\), \((X, K)\) is recurrent if
\[
\eta^T A e < \eta^T B_1 e. \tag{2.32}
\]
It is evident that (2.32) is equivalent to \(\lambda(\beta_u(\mu_o) + \beta_d) < \beta_u(\mu_o) \mu_o\). Since the spectral radius of \(Q\) is bounded, the stationary distribution of \((X, K)\) exists and thus \(\nu \in \mathcal{U}_{\text{adm}}\).

For the necessity, it follows by Theorem 3.1.1 of [42] that \((X, K)\) is positive recurrent under \(\mu_o\) only if \(\eta^T A e < \eta^T B_1 e\).

Similarly as above, we choose \(\nu(x, 1) \equiv \mu_o\). We use \(R_\nu\) to denote the rate matrix that satisfies
\[
(\pi_\nu(n - 1, 0), \pi_\nu(n - 1, 1)) R_\nu = (\pi_\nu(n, 0), \pi_\nu(n, 1)) \quad \text{for } n \geq 1,
\]
where \(\pi_\nu\) denotes the stationary distribution of \((X, K)\) governed by \(\nu\). Applying Theorem 3.1.1 of [42] again, \(R_\nu\) is the minimal nonnegative solution of the quadratic matrix equation
\[
(R_\nu)^2 B_x + R_\nu Q_x + A = 0, \tag{2.33}
\]
and the spectral radius of \(R_\nu\) is less than 1. We refer the reader to Lemma 3.1 for an explicit representation of \(R_\nu\). Thus, the spectral radius of \((R_\nu)^m\) converges to 0 at a geometric decay as \(m \to \infty\). Therefore, by Assumption 2.3, we have that \(\varrho^*\) is finite. This completes the proof. \(\square\)

**Remark 2.2.** We rewrite (2.31) as
\[
\lambda < \max_{\mu \in \mathcal{U}} \left\{ \frac{\mu \beta_d}{\beta_u(\mu) + \beta_d} \right\}. \tag{2.34}
\]
Note that \(\frac{\beta_d}{\beta_u(\mu) + \beta_d}\) can be viewed as the fraction of time that the server functions normally, and thus, the expression in the maximum on the right-hand side of (2.34) represents the effective service rate. Therefore, (2.34) means that there exists a service rate such that the arrival rate is less than the effective service rate, that is, the stability condition for \((X, K)\).

In the next theorem, we present the existence of the solution to the average cost optimality inequalities (ACOI) and the optimal policy for the long-run average expected cost. To prove this theorem, we apply the vanishing discounted method.

**Theorem 2.3.** Suppose that \(C_m \geq \mathcal{R}(\bar{\mu})\), and Assumption 2.3 and (2.31) hold. Then, the following items hold:

(i) As \(\alpha \searrow 0\), \(V_a(\cdot, k) - V_a(0, 1)\) converges, along a subsequence, to a function \(V(\cdot, k)\), for \(k \in \{0, 1\}\), and \(\varrho^* = \lim_{\alpha \to 0} \alpha V_a(x, k)\) for every \((x, k) \in \mathbb{Z}_+ \times \{0, 1\}\). Moreover, \(\{V(\cdot, k) : k \in \{0, 1\}\}\) and \(\varrho^*\) satisfy the ACOI:

\[
\begin{align*}
\begin{cases}
V(x, 1) &\geq \frac{1}{M} (H(x) + \lambda V(x + 1, 1) - \phi(W(x, 1), Y(x)) - \varrho_s + (M - \lambda) V(x, 1)), \\
V(x, 0) &\geq \frac{1}{M} (H(x) + \lambda V(x + 1, 0) + C_m - \beta_d Y(x) - \varrho_s + (M - \lambda) V(x, 0)),
\end{cases}
\end{align*}
\]

for \(x \in \mathbb{N}\), and

\[
0 \geq \frac{1}{M} (H(0) + \mathcal{R}(0) + \lambda V(1, 1) + \beta_u(0) Y(0) - \varrho_s), \tag{2.36}
\]

where \(W(x, 1) := V(x + 1, 1) - V(x, 1)\) and \(Y(x) := V(x, 0) - V(x, 1)\) for \(x \in \mathbb{Z}_+\).
(ii) There exists a long-run average cost optimal control \( \nu^* \in \mathcal{U}_{ssm} \), which is a limit of a sequence of optimal controls for the discounted cost problem.

Proof. To prove this theorem, we verify Assumptions 1-8 in [14]. Since the uniformization rate \( M \) is positive and finite, then Assumptions 1 and 2 of [14] are satisfied. To verify Assumptions 3 and 4 of [14], it suffices to show that Lemma 2.1 (i) of [14] holds. But it follows directly by Proposition 2.4, which implies that there exists a stable policy such that the long-run average cost is finite and \((X, K)\) is an ergodic Markov process. Let \( V_\alpha(0, 1) \) be the distinguishing point. Applying Propositions 2.1 and 2.2, we obtain

\[
V_\alpha(x, k) - V_\alpha(0, 1) \geq 0 \quad \forall (x, k) \in \mathbb{Z}_+ \times \{0, 1\}.
\]

Thus, Assumption 5 of [14] is satisfied. Let

\[
C_{x,0}(\alpha, 0) := \frac{C_m + H(x)}{\alpha + M} \quad \text{and} \quad C_{x,1}(\alpha, \mu) := \frac{R(\mu) + H(x)}{\alpha + M}.
\]

It is evident that for each \((x, k) \in \mathbb{Z}_+ \times \{0, 1\}, C_{x,0} \) and \( C_{x,1} \) are continuous functions on \([0, \infty) \times U\). Hence, Assumption 6 of [14] is verified. To verify Assumption 7 of [14], we define

\[
L_{(x,0),(x',k')}(\alpha, \mu) = \begin{cases} \frac{\lambda}{\alpha + M} & \text{if } x' = x + 1 \text{ and } k' = 0, \\
\frac{\beta_x(0)}{\alpha + M} & \text{if } x' = x \text{ and } k' = 1, \\
0 & \text{otherwise}, \end{cases}
\]

for \( x \in \mathbb{Z}_+ \),

\[
L_{(0,1),(x',k')}(\alpha, \mu) = \begin{cases} \frac{\lambda}{\alpha + M} & \text{if } x' = 1 \text{ and } k' = 1, \\
\frac{\beta_x(0)}{\alpha + M} & \text{if } x' = 0 \text{ and } k' = 0, \\
0 & \text{otherwise}, \end{cases}
\]

and

\[
L_{(x,1),(x',k')}(\alpha, \mu) = \begin{cases} \frac{\lambda}{\alpha + M} & \text{if } x' = x + 1 \text{ and } k' = 1, \\
\frac{\mu}{\alpha + M} & \text{if } x' = x - 1 \text{ and } k' = 1, \\
\frac{\beta_x(0)}{\alpha + M} & \text{if } x' = x \text{ and } k' = 0, \\
0 & \text{otherwise}. \end{cases}
\]

for \( x \in \mathbb{N} \). Then, it is evident that for all \((x, k) \in \mathbb{Z}_+ \times U\), \( L_{(x, k),(x', k')}(\alpha, \mu) \) is a continuous function on \([0, \infty) \times U\), and Assumption 7 of [14] is satisfied. Since the expected sojourn times for \((X, K)\) under any \( \nu \) are equal to \( 1/M \), then Assumption 8 of [14] holds. Thus, by Theorem 12 of [14], we have shown (i) and (ii) \( \square \)

Corollary 2.1. Suppose that the hypotheses in Theorem 2.3 and Assumption 2.2 hold. There exists an optimal policy \( \nu^* \in \mathcal{U}_{ssm} \) such that \( \nu^*(\cdot, 1) \) is nondecreasing.

Proof. By Theorem 2.3, we choose \( \{\nu_m : m \in \mathbb{N}\} \) with \( \alpha_m \searrow 0 \) as a sequence of optimal policies in Proposition 2.3 (iii) for the discounted cost problem such that \( \nu_m \) converges to an average cost optimal policy \( \nu^* \in \mathcal{U}_{ssm} \). Since \( \nu_m(\cdot, 1) \) is nondecreasing, the limit \( \nu^*(\cdot, 1) \) is a nondecreasing function. \( \square \)

Remark 2.3. In practice, to obtain the optimal policies, people usually study the approximation of the original problem, since (2.35) is almost impossible to solve directly for the infinite state space. In the literature, there are two approaches for the approximation. One approach is to use a sequence of control problems with truncated holding cost functions to generate the approximations. In [10], an asymptotic method that uses truncated a holding cost function to compute the optimal policy is developed for the average-cost problem of the \( M/M/1 \) queue. They show that the optimal policies of the approximating problems converge monotonically to the optimal policy of the original problem, and the optimal controls are nondecreasing functions in the congestion level of the system.
The other approach is to use a sequence of optimal objective values for the finite state space to approximate the original optimal value. This approach has been studied in [15,16] for the \( M/M/1 \) queue with rejection cost. For notation convenience, we let \( \nu > 0 \) denote a fixed penalty to reject a customer. For notation convenience, we let \( \nu_x \equiv \nu(x,1) \) for \( x \in \mathbb{Z}_+ \). The ergodic cost is given by
\[
\varrho' := \sum_{x=0}^{N} \left( \pi_x (x,1) (H(x) + \mathcal{R}(\nu_x)) + \pi_x(x,0) (H(x) + C_m) \right) + \lambda p (\pi_x(N,1) + \pi_x(N,0)), \tag{3.1}
\]
and the optimal ergodic cost is defined by
\[
\varrho_* := \inf_{\nu \in \mathcal{M}_{ssm}} \nu' . \tag{3.2}
\]
We replace the assumptions on the action space, the cost functions and the breakdown rate function \( \beta_u \) in the previous section with the following relaxed assumptions.

**Assumption 3.1.** We assume that

(i) The action space \( \mathbb{U} \) is a compact subset of \([0, \infty)\) satisfying \( 0 \in \mathbb{U} \).

(ii) The holding cost function \( H(x) \) is nondecreasing in \( x \).

(iii) The effort cost function \( \mathcal{R} \) is nondecreasing and continuous satisfying \( \mathcal{R}(0) = 0 \).

(iv) The function \( \beta_u \) is strictly positive and continuous.

Recall \( \bar{\mu} := \max\{ \mu \in \mathbb{U} \} \) and \( M := \bar{\mu} + \lambda + \beta_u(\bar{\mu}) + \beta_d \). By applying the uniformization technique, the optimality equations for the ergodic control problem are given by
\[
V(0,1) = \frac{1}{M} \left( \mathcal{R}(0) + H(0) + \lambda V(1,1) + \beta_u(0) V(0,0) - \varrho + (M - \lambda - \beta_u(0)) V(0,1) \right), \tag{3.3}
\]
and
\[
V(x,1) = \frac{1}{M} \min_{\mu \in \mathbb{U}} \left\{ \mathcal{R}(\mu) + H(x) + \lambda V(x+1,1) + \mu V(x-1,1) \right. \\
+ \beta_u(\mu) V(x,0) - \varrho + (M - \lambda - \mu - \beta_u(\mu)) V(x,1) \} \tag{3.4}
\]
for \( 1 \leq x \leq N - 1 \), and
\[
V(x,0) = \frac{1}{M} \left( H(x) + C_m + \lambda V(x+1,0) + \beta_d V(x,1) - \varrho + (M - \lambda - \beta_d) V(x,0) \right) \tag{3.5}
\]
for \( 0 \leq x \leq N - 1 \), and
\[
V(N,1) = \frac{1}{M} \min_{\mu \in \mathbb{U}} \left\{ \mathcal{R}(\mu) + H(N) + \lambda p + \mu V(N-1,1) \right. \\
+ \beta_u(\mu) V(N,0) - \varrho + (M - \mu - \beta_u(\mu)) V(N,1) \} , \tag{3.6}
\]
and
\[
V(N,0) = \frac{1}{M} \left( H(N) + C_m + \lambda p + \beta_d V(N,1) - \varrho + (M - \beta_d) V(N,0) \right) . \tag{3.7}
\]
To simplify the notation, we define the relative cost differences $W_1 := V(0,1)$, $W_x := V(x,1) − V(x−1,1)$ for $1 ≤ x ≤ N$, and $Y_x := V(x,0) − V(x,1)$ for $0 ≤ x ≤ N$. By using the relative cost differences, the optimality equations take the form

\begin{align}
\begin{cases}
\lambda W_1 = -\beta_u(0)Y_0 - H(0) + \varrho, \\
\lambda(W_1 + Y_1) = (\lambda + \beta_d)Y_0 - C_m - H(0) + \varrho, \\
\lambda(W_{x+1} + Y_{x+1}) = (\lambda + \beta_d)Y_x - C_m - H(x) + \varrho,
\end{cases}
\end{align}

(3.8)

for $1 ≤ x ≤ N − 1$, and

\begin{align}
\begin{cases}
\lambda p = \phi(W_N,Y_N) - H(N) + \varrho, \\
\lambda p = \beta_dY_N - C_m - H(N) + \varrho,
\end{cases}
\end{align}

(3.9)

for $x = N$, where $\phi$ is defined in (2.4). Note that the system of equations (3.3)–(3.7) is equivalent to that of equations (3.8)–(3.10).

Recall that $\{\pi_\nu(x,k) : 0 ≤ x ≤ N, k \in \{0,1\}\}$ denotes the stationary distribution of the irreducible joint Markov process $(X,K)$ under a policy $\nu \in \mathcal{U}_{sm}$. In the following lemma, we provide the recursive formula for the stationary distribution of $(X,K)$ when the system is not full.

**Lemma 3.1.** Let $\nu \in \mathcal{U}_{sm}$ be such that the stationary distribution of $(X,K)$ exists under $\nu$. Then, for $1 ≤ x ≤ N − 1$,

\begin{align}
(\pi_\nu(x,1),\pi_\nu(x,0)) R_\nu(x) &= (\pi_\nu(x,0),\pi_\nu(x,1)),
\end{align}

(3.11)

where

\begin{align}
R_\nu(x) = \left( \frac{\lambda(\beta_u(\nu_x) + \nu_x)}{\lambda(\lambda + \beta_d)} \frac{\lambda}{\lambda(\lambda + \beta_d)} \right),
\end{align}

and $(R_\nu(x))^{-1} = \left( \frac{\lambda + \beta_d}{\lambda + \beta_d} \frac{\beta_u(\nu_x)}{\nu_x} \right)$.

**Proof.** It is evident that the balance equations for $(X,K)$ take the form

\begin{align}
\begin{cases}
(\lambda + \beta_u(0))\pi_\nu(0,1) = \beta_d\pi_\nu(0,0) + \nu_1\pi(1,1), \\
(\lambda + \beta_d)\pi_\nu(0,0) = \beta_u(0)\pi_\nu(0,1),
\end{cases}
\end{align}

(3.12)

and

\begin{align}
\begin{cases}
(\lambda + \beta_u(\nu_x) + \nu_x)\pi_\nu(x,1) = \beta_d\pi_\nu(x,0) + \nu_{x+1}\pi_\nu(x + 1,1) + \lambda\pi_\nu(x − 1,1), \\
(\lambda + \beta_d)\pi_\nu(x,0) = \beta_u(\nu_x)\pi_\nu(x,1) + \lambda\pi_\nu(x − 1,0),
\end{cases}
\end{align}

(3.13)

for $1 ≤ x ≤ N − 1$. We sum the equations in (3.12) and get

\begin{align}
\lambda(\pi_\nu(0,0) + \pi_\nu(0,1)) = \nu_1\pi(1,1).
\end{align}

(3.14)

By applying (3.14) and adding the equations in (3.13), we have that for $1 ≤ x ≤ N − 1$,

\begin{align}
\lambda(\pi_\nu(x,1) + \pi_\nu(x,0)) = \nu_{x+1}\pi_\nu(x + 1,1).
\end{align}

(3.15)

Thus, by (3.13) and (3.15), we obtain

\begin{align}
\begin{cases}
\lambda\pi_\nu(x − 1,0) = (\lambda + \beta_d)\pi_\nu(x,0) - \beta_u(\nu_x)\pi_\nu(x,1), \\
\lambda\pi_\nu(x − 1,1) = -(\lambda + \beta_d)\pi_\nu(x,0) + (\beta_u(\nu_x) + \nu_x)\pi_\nu(x,1),
\end{cases}
\end{align}

for $1 ≤ x ≤ N − 1$. Thus, equation (3.11) holds, and this completes the proof. □

In the following theorem, we characterize the optimal controls. Its proof relies on the results of Lemma 3.1. Recall that the function $\psi$ represents the minimizer of (2.4).

**Theorem 3.1.** Let $\varrho < \infty$ and $(W_1,W_2,...,W_N,Y_0,Y_1,...,Y_N)$ be a solution to (3.8)–(3.10). Under Assumption 3.1, if $Y_x ≥ 0$ for $0 ≤ x ≤ N$, $W_x ≥ 0$ and $\nu_\varrho^* := \psi(W_x,Y_x) > 0$ for $1 ≤ x ≤ N$, then $\{\nu_\varrho^* : 1 ≤ x ≤ N\}$ is optimal and $\varrho_\varrho^* = \varrho = \varrho_\star$. 
Proof. Let $\nu$ be an optimal rate control policy of (3.2). Multiplying both sides of the first equation in (3.9) by $\pi\nu(x,1)$ and both sides of the second equation in (3.9) by $\pi\nu(x,0)$, we obtain

$$
\begin{cases}
(H(x) + CR(\nu_x) - \varrho)\pi\nu(x,1) \geq W_x\nu_x\pi\nu(x,1) - Y_x\beta_u(\nu_x)\pi\nu(x,1) - \lambda W_x\pi\nu(x,1), \\
(H(x) + C_m - \varrho)\pi\nu(x,0) = (\lambda + \beta_d)Y_x\pi\nu(x,0) - \lambda(W_{x+1} + Y_{x+1})\pi\nu(x,0),
\end{cases}
$$

(3.16)

for $1 \leq x \leq N - 1$. It follows by (3.11) and (3.15) that

$$
W_x\nu_x\pi\nu(x,1) + Y_x((\lambda + \beta_d)\pi\nu(x,0) - \beta_u(\nu_x)\pi\nu(x,1)) = W_x\lambda(\pi\nu(x-1,1) + \pi\nu(x-1,0)) + Y_x\lambda\pi\nu(x-1,0).
$$

(3.17)

Then, summing the equations in (3.16) and applying (3.17), we obtain that for $1 \leq x \leq N - 1$,

$$
H(x)(\pi\nu(x,1) + \pi\nu(x,0)) + R(\nu_x)\pi\nu(x,1) + C_m\pi\nu(x,0) - \varrho(\pi\nu(x,1) + \pi\nu(x,0))
\geq W_x\lambda(\pi\nu(x-1,1) + \pi\nu(x-1,0)) + Y_x\lambda\pi\nu(x-1,0)
$$

(3.18)

$$
- (W_{x+1}\lambda(\pi\nu(x,1) + \pi\nu(x,0)) + Y_{x+1}\lambda\pi\nu(x,0)).
$$

It follows by the balance equation

$$
\begin{cases}
(\lambda + \beta_u(\nu_N) + \nu_N)\pi\nu(N,1) = \beta_d\pi\nu(N,0) + \lambda\pi\nu(N-1,1), \\
(\lambda + \beta_d)\pi\nu(N,0) = \beta_u(\nu_N)\pi\nu(N,1) + \lambda\pi\nu(N-1,0),
\end{cases}
$$

(3.19)

that

$$
\beta_d\pi\nu(N,0) - \beta_u(\nu_N)\pi\nu(N,1) = \lambda\pi\nu(N-1,0).
$$

(3.19)

In analogy to (3.18), applying (3.10), (3.15), and (3.19), we obtain

$$
(\lambda p + H(N) - \varrho)(\pi\nu(N,1) + \pi\nu(N,0)) + R(\nu_N)\pi\nu(N,1) + C_m\pi\nu(N,0)
\geq W_N\lambda(\pi\nu(N-1,1) + \pi\nu(N-1,0)) + Y_N\lambda\pi\nu(N-1,0).
$$

(3.20)

Adding all the equations in (3.18) for $1 \leq x \leq N - 1$ together with (3.20), and applying (3.1), we have

$$
\varrho_* - \varrho - (H(0) - \varrho)(\pi\nu(0,0) + \pi\nu(0,1)) - C_m\pi\nu(0,0)
\geq W_1\lambda(\pi\nu(0,1) + \pi\nu(0,0)) + Y_1\lambda\pi\nu(0,0).
$$

(3.21)

It follows by (3.12) that

$$
(\lambda + \beta_d)\pi\nu(0,0) - \beta_u(0)\pi\nu(0,1) = 0.
$$

(3.22)

Then, by (3.8) and (3.22), we have

$$
\lambda W_1\pi\nu(0,1) + \lambda(W_1 + Y_1)\pi\nu(0,0) = -C_m\pi\nu(0,0) - (H(0) - \varrho)(\pi\nu(0,0) + \pi\nu(0,1)).
$$

(3.23)

Thus, applying (3.21) and (3.23), we obtain $\varrho_* \geq \varrho$. It implies $\varrho_* = \varrho$, and then the proof is completed.

By Lemma 3.1, it follows that $\pi\nu(x,k)$ is continuous with respect to $\nu$ for each $1 \leq x \leq N$ and $k \in \{0,1\}$. It is evident that $\pi\nu$, for each $\nu \in \mathcal{U}_{\text{sm}}$, takes values in a compact set. Then the minimum of (3.2) exists.

In the next lemma, we show the uniqueness of the value functions. The result of the lemma is also used in the study of the adaptive control problem in the next section.

**Lemma 3.2.** If a solution $\{V(x,k)\}_{(x,k) \in S}$ to (3.3)–(3.7) exists, then the solution is unique.

**Proof.** We prove this lemma by contraction. It follows by Theorem 3.1 that $\varrho = \varrho_*$ for any solutions to (3.3)–(3.7). Let $S$ denote the state space of $(X,K)$. We define the operator $\mathcal{T}_V : S \mapsto \mathbb{R}$ by

$$
\mathcal{T}_V(0,1) := \frac{1}{M}(R(0) + H(0) + \lambda V(1,1) + \beta_u(0)V(0,0) - \varrho_* + (M - \lambda - \beta_u(0))V(0,1)),
$$

(3.24)
\[
\mathcal{T}_V(x, 1) := \frac{1}{M} \left( H(x) - \varrho_* - \phi(V(x, 1) - V(x - 1, 1), V(x, 0) - V(x, 1)) + \lambda V(x + 1, 1) + (M - \lambda) V(x, 1) \right)
\]

for \(1 \leq x \leq N - 1,\)
\[
\mathcal{T}_V(N, 1) := \frac{1}{M} \left( H(N) - \varrho_* - \phi(V(N, 1) - V(N - 1, 1), V(N, 0) - V(N, 1)) + \lambda p \right),
\]
\[
\mathcal{T}_V(x, 0) := \frac{1}{M} \left( H(x) + C_m + \lambda V(x + 1, 0) + \beta_d V(x, 1) - \varrho_* + (M - \lambda - \beta_d) V(x, 0) \right)
\]

for \(0 \leq x \leq N - 1,\) and
\[
\mathcal{T}_V(N, 0) := \frac{1}{M} \left( H(N) + C_m + \lambda p + \beta_d V(N, 1) - \varrho_* + (M - \beta_d) V(N, 0) \right).
\]

Let \(V_1\) and \(V_2\) be any functions satisfying (3.3)–(3.7). It suffices to show that there exists some positive constant \(C < 1\) such that
\[
\max_{(x, k) \in S} |\mathcal{T}_{V_1}(x, k) - \mathcal{T}_{V_2}(x, k)| \leq C \max_{(x, k) \in S} |V_1(x, k) - V_2(x, k)|.
\]  

We show that (3.25) satisfies (3.29). Note that for any functions \(g\) and \(f,\)
\[
|\max_x g(x) - \max_x f(x)| \leq \max_x |g(x) - f(x)|.
\]

Thus, for any \(0 \leq x \leq N - 1,\)
\[
|\mathcal{T}_{V_1}(x, 1) - \mathcal{T}_{V_2}(x, 1)|
\leq \frac{1}{M} \left( \bar{\mu} \left( |V_1(x, 1) - V_2(x, 1)| + |V_1(x - 1, 1) - V_2(x - 1, 1)| \right)
\right.
\]
\[
+ \beta(\bar{\mu}) \left( |V_1(x, 1) - V_2(x, 1)| + |V_1(x, 0) - V_2(x, 0)| \right) + \lambda |V_1(x + 1, 1) - V_2(x + 1, 1)|
\]
\[
\left. + (M - \lambda) |V_1(x + 1, 1) - V_2(x, 1)| \right)
\]
\[
\leq C \max_{(x, k) \in S} |V_1(x, k) - V_2(x, k)|,
\]

where \(C\) is some constant such that
\[
\max \left\{ \frac{\bar{\mu}}{M}, \beta(\bar{\mu}) \frac{\lambda}{M}, \frac{M - \lambda}{M} \right\} \leq C < 1.
\]

By repeating the procedure described above and applying (3.24) and (3.26)–(3.28), we obtain (3.28). It follows that \(\mathcal{T}_V\) is a contraction, and it has a unique fixed point. This completes the proof.

Remark 3.1. For the service rate control problem of \(M/M/1\) queue, Adusumilli et al. [15] consider a sequence of approximating problems with rejection and truncated state space to the ergodic control problem with infinite capacity. They show that under a sequence of terminating optimal solutions such that the optimal values are decreasing, the limiting policy exists and is optimal for the ergodic control problem with infinite capacity. Their proof for the convergence of the approximating problems crucially relies on the property that the optimal service rates are nondecreasing in the number of jobs in the system. However, this monotone property does not hold in general for the queueing model discussed in this paper, since the server breakdown rate depends on the service rate. In Section 5.1, we provide some numerical examples for cases in which the optimal service rates are not monotone in the number of jobs in the system. For the service rate control problem of \(M/M/1\) queue with breakdowns and finite capacity, it remains open to show that there exists a sequence of optimal policies under which the optimal values of the problems with a truncated number of states converges to the optimal value of the ergodic control problem with infinite capacity.
Remark 3.2. For the service rate control problem of M/M/1 queue with breakdowns and finite capacity under the discounted cost criterion, one may apply the same approach, which uses the uniformization technique, as in Section 2.2 to obtain the existence and characterization of optimal controls. For the simplicity, in this section, we focus on the service rate control problem under the ergodic cost criterion. We apply the approach by involving the stationary distribution of the joint Markov process \((X, K)\). This approach is different from the vanishing discounted method used for the problem with infinite capacity as described in Section 2.3. Since we consider a queue with finite capacity, that is, the state process has finite states, the stationary distribution for the joint process \((X, K)\) can be expressed as a finite-dimensional vector. Therefore, it is natural to consider the approach with a stationary distribution as shown in the proof of Theorem 3.1.

4. The adaptive control problem

In practice, the function \(\beta_u\) in (3.3)–(3.7) may be unobservable from data. At each state \(x_i, i = 1, 2, \cdots\), of the state process \(X\), the controller must choose the service rate \(\nu_i(x_i)\) for the server. The function \(\beta_u\) can be inferred from the history of \(\{\nu_i(x_i)\}_{i \in \mathbb{N}}\) and the sojourn times when the system is in the up state.

To simplify the notation, throughout this section, we assume that (2.17) holds with \(\kappa_1\) and \(\kappa_2\) taking values in compact sets \(K_1\) and \(K_2\), respectively. We define \(\kappa := (\kappa_1, \kappa_2)^T\), and assume that \(\kappa\) is initially unknown and \(K_1, K_2 \subset \mathbb{R}_+\). We consider a queueing system with finite capacity and the objective function as in Section 3. Note that in Remark 4.1, we provide some approaches to relax the assumption in (2.17).

We let the sequence \(\{\nu_i(x_i)\}_{i \in \mathbb{N}}\) with \(\nu_i \in \mathcal{U}_{ssm}\) denote the design variables corresponding to the service rates during the "up" times, that is, when the process \(K = 1\). We use \(T(\nu_i(x_i))\) to denote the sojourn time of the joint Markov process governed by \(\nu_i \in \mathcal{U}_{ssm}\) when the processes \(X = x_i\) and \(K = 1\). Given \(\{\nu_i(x_i)\}_{i \in \mathbb{N}}\), it is evident that each \(T(\nu_i(x_i))\) is exponentially distributed with the parameter \(\lambda + \kappa_1 + (\kappa_2 + 1)\nu_i(x_i)\). Let \(\{t_i\}_{i \in \mathbb{N}}\) be a sequence of realizations of random variables \(\{T(\nu_i(x_i))\}_{i \in \mathbb{N}}\). Then, given a sample of design variables and responses with size \(n\), we let \(\hat{\kappa}^n := (\hat{\kappa}_1^n, \hat{\kappa}_2^n)^T\) denote a solution of the quasi-likelihood equations taking the form

\[
\begin{align*}
\sum_{i=1}^n \nu_i(x_i) \left( t_i - \left( \hat{\kappa}_1^n + \hat{\kappa}_2^n \nu_i(x_i) \right)^{-1} \right) &= 0, \\
\sum_{i=1}^n \left( t_i - \left( \hat{\kappa}_1^n + \hat{\kappa}_2^n \nu_i(x_i) \right)^{-1} \right) &= 0.
\end{align*}
\]

(4.1)

The solution of (4.1) may not be unique. However, we may choose the "correct" root with the lowest mean-least square error, see, for example, [25, Chapter 13.3]. Such a root of (4.1) is a quasi maximum likelihood estimate of \(\hat{\kappa} := (\kappa_1 + \lambda, \kappa_2 + 1)^T\). Given the true parameter \(\hat{\kappa}\), we define the error terms \(\{\varepsilon_i\}_{i \in \mathbb{N}}\) by

\[\varepsilon_i := t_i - \left( \lambda + \kappa_1 + (\kappa_2 + 1)\nu_i(x_i) \right)^{-1}.\]

Note that there is no decision to make during the down times, and if decisions are made based on current parameter estimations during the up times, then \(\{\varepsilon_i\}_{i \in \mathbb{N}}\) forms a martingale with respect to its natural filtration. Therefore, we consider the adaptive design case in this section.

It is well known that the estimations from (4.1) may be inconsistent. Therefore, we provide a family of rate control policies under which the estimations from (4.1) are consistent. First, we present a sufficient condition for the strong consistency of \(\hat{\kappa}^n\). Let \(\hat{\xi}^n\) and \(\hat{\zeta}^n\) be the smallest and the largest eigenvalues of the design matrix

\[
\sum_{i=1}^n (1, \nu_i(x_i))^T (1, \nu_i(x_i)) = \left[ \sum_{i=1}^n \nu_i(x_i) \sum_{i=1}^n (\nu_i(x_i))^2 \right],
\]

respectively. The following lemma directly follows by Theorem 2.1 of [26].
Lemma 4.1. Assume that $\zeta^n \to \infty$ as $n \to \infty$ a.s., and
\[
\liminf_{n \to \infty} \frac{\zeta^n}{(\zeta^n \log \zeta^n)^{1/2}(\log \log \zeta^n)^{1/2+\delta}} > 0 \quad \text{a.s.}
\]
for some $\delta > 0$. Then, (4.1) has a solution $\hat{\kappa}^n$ such that
\[
|\hat{\kappa}^n - \kappa| = o\left(\frac{(\zeta^n \log \zeta^n)^{1/2}(\log \log \zeta^n)^{1/2+\delta}}{\zeta^n}\right) \quad \text{a.s.} \quad (4.2)
\]

In the following lemma, we verify the conditions in Lemma 4.1 for a class of rate control policies, and then show that there exists a sequence of estimations from (4.1) is strongly consistent.

Lemma 4.2. Under any sequence of work-conserving Markov rate control policies, the conditions in Lemma 4.1 are satisfied. Moreover, (4.1) has a solution $\hat{\kappa}^n$ such that
\[
|\hat{\kappa}^n - \kappa| = o\left(\frac{(\log n)^{1/2}(\log \log n)^{1/2+\delta}}{\sqrt{n}}\right) \quad \text{a.s.} \quad (4.3)
\]
for any $\delta > 0$.

Proof. Since $\{\nu_i(x_i)\}_{i \in \mathbb{N}}$ is uniformly bounded, it is evident that $\bar{\zeta}^n = O(n)$. We show the lower bound of $\zeta^n$ for all large $n$. Recall that for any constant rate policy $\nu \equiv \mu \in \mathbb{U}$, the Markov process $(X, K)$ is ergodic. Since the service rate for work-conserving Markov policy is bounded away from 0 by assumption, it follows that the mean return time of the embedded Markov chain of $(X, K)$ to the state $(0,1)$ is uniformly bounded under any sequence of work-conserving Markov policies. Recall that $\{\nu_i(x_i)\}_{i \in \mathbb{N}}$ denote the design variables during the up times. Let $\bar{\nu}^n := n^{-1} \sum_{i=1}^n \nu_i(x_i)$ and $\nu$ denote the policy satisfying $\nu(0,1) = 0$ and
\[
\nu(x,1) \equiv \nu = \arg\min_{\mu \in \mathbb{U}} \left\{ \mu^2 \left(1 - \pi_\mu(0,1)\right)^2 \pi_\mu(0,1) \right\} \quad \forall x \in \mathbb{N}. \quad (4.4)
\]
The minimum of (4.4) can be attained because $\{\pi_\mu(x,k) : (x,k) \in \mathbb{Z}_+ \times \{0,1\}\}$ are continuous functions of $\mu$ on $\mathbb{U}$ by Lemma 3.1, (3.12) and (3.19). Let $\mathbb{I}$ denote the indicator function. Thus, based on the ergodic theory,
\[
\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \left(\nu_i(x_i) - \bar{\nu}^n\right)^2 \geq \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \left(\nu_i(x_i) - \bar{\nu}^n\right)^2 \mathbb{I}(x_i = 0)
\]
\[
= \liminf_{n \to \infty} (\bar{\nu}^n)^2 \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i = 0) \geq \frac{\nu^2 (1 - \pi_\nu(0,1))^2 \pi_\nu(0,1)}{(\sum_{j=0}^N \pi_\nu(j,1))^3} \quad (4.5)
\]
a.s., where the last inequality follows by (4.4) and the fact $\nu_i(0) = 0$ for any $i$. Applying Lemma 2 of [33], we have
\[
\zeta^n \geq C_1 \sum_{i=1}^n \left(\nu_i(x_i) - \bar{\nu}^n\right)^2 \quad (4.6)
\]
where $C_1 = \frac{3}{(1+2\bar{\mu}-\bar{\mu})^2}$. Then, by (4.5) and (4.6), it follows that
\[
\zeta^n \geq C_2(1 + n)
\]
for some positive constant $C_2$. Therefore, both $\zeta^n$ and $\bar{\zeta}^n$ are at the order of $n$, and we have verified the conditions in Lemma 4.1. It follows by (4.2) that (4.3) holds, and this completes the proof. \qed

Let $\tilde{\nu}^n$ be the optimal value of (3.2) with the parameters $\kappa_1$ and $\kappa_2$ replaced by $\hat{\kappa}_1^n - \lambda$ and $\hat{\kappa}_2^n - 1$, respectively. Recall that $\nu^*$ denotes the true optimal ergodic cost in (3.2). In the following lemma, we show the convergence of optimal values.
Lemma 4.3. If \( \hat{\kappa}_n \to \kappa \) as \( n \to \infty \) a.s., then \( \hat{\varphi}_n \to \varphi^* \) as \( n \to \infty \) a.s.

The result of Lemma 4.3 directly follows by the expression in (3.1) together with the Lemma 3.1, and applying Theorem 2.3 of [43], and we omit its proof.

Lemma 4.4. Assume that (2.17) holds and \( \mathcal{R} \) is strongly convex and continuously differentiable. Let \( \nu^i \) be the optimal control obtained by solving (3.8)–(3.10) under the estimator \( \hat{\kappa}^i \). Then, it follows that
\[
\nu^i(x,1) \to \nu^*(x,1) \quad \text{a.s.} \quad (4.7)
\]
as \( i \to \infty \), for \( 1 \leq x \leq N \).

Proof. It is evident that \( \phi(w,y) \) is a continuous function of \( w \) and \( y \). For any \( (w,y) \), \( \phi(w,y) \) is also continuous with respect to the parameter \( \kappa \). Let \( \{W^i_x,Y^i_x; 1 \leq i \leq N\} \) be the solution of equations (3.8)–(3.10) under the estimate \( \hat{\kappa}^i \). Let \( \hat{\varphi}_i \) denote the optimal value in (3.2) under the estimate \( \hat{\kappa}^i \).

Note that \( \{W_x,Y_x; 0 \leq x \leq N\} \) are continuous functions of \( \kappa \) and \( \varphi \). Since \( \hat{\kappa}^i \to \kappa \) and \( \hat{\varphi}_i \to \varphi^* \) a.s. as \( i \to \infty \) by Lemma 4.3, then it follows by the continuous mapping theorem that
\[
W^i_x \to W_x^*, \quad \text{and} \quad Y^i_x \to Y_x^*, \quad \text{a.s.}
\]
as \( i \to \infty \) for \( 1 \leq x \leq N \). By the strong convexity and continuity of \( \phi \), the convergence of optimal values \( \phi(W^i_x,Y^i_x) \), \( i \in \mathbb{N} \), implies the convergence of maximizers of \( \{\phi(W^i_x,Y^i_x); i \in \mathbb{N}\} \). We refer the readers to (2.18) for the representation of maximizers. We have shown (4.7). \( \square \)

For the queue with a finite capacity, the dynamics in (2.1) becomes
\[
X(t) = \left( X(0) + A(t) - S \left( \int_0^t \nu(X(s),K(s))(X(s) \wedge 1) \, ds \right) \right) \wedge N \quad \forall t \geq 0.
\]

Recall the cost function \( f(x,k,\mu) \) in (2.2). To simplify the notation, we let \( f^i(x,k) = f(x,k,\nu^i(x,k)) \) for \( \nu^i \in \mathbb{U}_{\text{fin}} \). Let \( m_i \) denote the number of jumps for the process \( (X,K) \) before time \( t \), \( \tau_0 = 0 \), and \( \tau_i, 1 \leq i \leq m_t \), denote the \( i \)-th jump time of \( (X,K) \). We define the cumulative cost function \( F \) by
\[
F(t) := \sum_{i=1}^{m_t} \int_{\tau_{i-1}}^{\tau_i} f^i_{\nu^i-1}(X(s),K(s)) \, ds + \int_{\tau_{m_t}}^{t} f^i_{\nu^i}(X(s),K(s)) \, ds
\]
\[
+ p \sum_{i=1}^{m_t} \left( A(\tau_i) - A(\tau_{i-1}) \right) \int_{\tau_{i-1}}^{\tau_i} \mathbb{1}(X(s) = N) \, ds
\]
\[
+ p \left( A(t) - A(\tau_{m_t}) \right) \int_{\tau_{m_t}}^{t} \mathbb{1}(X(s) = N) \, ds \quad (4.8)
\]
for \( t \geq 0 \), where \( \nu^i \) denotes the policy updated by solving (3.8)–(3.10) under the estimator \( \hat{\kappa}^i \), and third and fourth terms on the left hand side (LHS) correspond to the penalty of rejections.

In the next theorem, we present the main result of this section. The theorem implies that if we estimate the unknown parameters \( \kappa_1 \) and \( \kappa_2 \) under work-conserving rate controls at each state and update the rate controls by solving (3.8)–(3.10) and under estimated parameters, then the long-run average cost converges to the optimal cost. Because the transition rate matrix is updated over time due to the change of parameters, the joint Markov process is time-varying, and the proof of the theorem relies on Kruglov strong law of large numbers; see Theorem 2 in [44].

Theorem 4.1. Assume that (4.7) holds. Then,
\[
\lim_{t \to \infty} \frac{1}{t} \mathbb{E}[F(t)] = \varphi^* . \quad (4.9)
\]
Proof. It is evident that
\[
\frac{1}{t} \int_0^t f_{\mu(s)}(X(s), K(s)) \, ds = \sum_{x=0}^N \sum_{k=0}^1 \frac{1}{t} \int_0^t \mathbb{1}(X(s) = x, K(s) = k) f_{\mu(s)}(x, k) \, ds ,
\] (4.10)
where \( \mu(s) = \nu^i(X(s), K(s)) \) for \( \tau_i \leq s < \tau_{i+1} \). For \( (x, k) \in \mathcal{S} \), let \( \tau_n(x, k) \) denote the \( n \)-th time at which the Markov process jumps into state \( (x, k) \) with \( \tau_0(x, k) = 0 \), and let \( h_n(x, k) \) denote the \( n \)-th holding time in the state \( (x, k) \). Define \( T_n(x, k) := \tau_n(x, k) - \tau_{n-1}(x, k) \) for \( n \geq 1 \). We use \( N_{x,k}(t) \) to denote the number of transitions of the Markov process into state \( (x, k) \) before time \( t \). Then, we have
\[
\int_0^t \mathbb{1}(X(s) = x, K(s) = k) f_{\mu(s)}(x, k) \, ds = \sum_{i=1}^{N_{x,k}(t)-1} h_i(x, k) f_{\mu_i}(x, k) + \left( (t - \tau_{N_{x,k}(t)}) \wedge h_{N_{x,k}(t)}(x, k) \right) f_{\mu_{N_{x,k}(t)}}(x, k) ,
\] (4.11)
where \( \mu_i \), for \( i \in \mathbb{N} \), denotes the service rate during the \( i \)-th holding time in the state \( (x, k) \). Applying the strong Markov property, given \( \{\nu^i: i \in \mathbb{N}\} \), \( \{h_i(x, k): i \in \mathbb{N}\} \) are independent. Let \( h_s(x, k) \) denote the holding time in the state \( (x, k) \) of the Markov process under the optimal service rate control \( \nu_s \). For \( k = 0 \), we have \( \mu_i \equiv 0 \) for \( i \in \mathbb{N} \), and then \( h_i(x, 0) \) are i.i.d., distributed as \( h_s(x, 0) \). For \( k = 1 \), applying (4.7), it follows that
\[
\lim_{i \to \infty} \mathbb{E}[h_i(x, 1) f_{\mu_i}(x, k)] = \mathbb{E}[h_s(x, 1) f_{\nu_s}(x, k)] .
\] (4.12)
Since the service rate is bounded, it is straightforward to check that
\[
\sup_{i \in \mathbb{N}} \mathbb{E}|h_i(x, k) f_{\mu_i}(x, k)| < \infty ,
\]
and
\[
\sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^n \mathbb{P}(|h_i(x, k) f_{\mu_i}(x, k)| > y) \leq \mathbb{P}(|f_0(1, 0)| \Psi > y)
\]
for all \( y \geq 0 \), where \( \Psi \) denotes a random variable having exponential distribution with parameter \( \lambda \). Thus, by Kruglov strong law of large numbers, conditioning on \( \{\nu^i: i \in \mathbb{N}\} \), we have
\[
\frac{1}{n} \sum_{i=1}^n h_i(x, k) f_{\mu_i}(x, k) \to \mathbb{E}_{\nu_s}[h_s(x, k) f_{\nu_s}(x, k)] \quad \text{a.s.}
\] (4.13)
as \( n \to \infty \). For the rejection cost in (4.8), given the state process \( X = N \), the holding time \( h_i(N, k) \) is independent of the arrival process \( A \), and then
\[
\lim_{i \to \infty} \mathbb{E}[A(h_i(N, k))] = \lim_{i \to \infty} \lambda \mathbb{E}[h_i(N, k)] = \lambda \mathbb{E}[h_s(N, k)] .
\] (4.14)
The similar result in (4.13) holds for the rejection cost. By repeating the procedure as above, we have that \( \{T_i(x, k): i \in \mathbb{N}\} \) are independent conditioning on \( \{\nu^i: i \in \mathbb{N}\} \), and
\[
\lim_{i \to \infty} \mathbb{E}[T_i(x, k)] = \mathbb{E}[T_s(x, k)] ,
\]
where \( T_s(x, k) \) denotes the return time to the state \( (x, k) \) of the Markov process under the optimal service rate control policy \( \nu_s \). By applying Kruglov strong law of large numbers and repeating the proof for the elementary renewal theory, it follows that given \( \{\nu^i: i \in \mathbb{N}\} \),
\[
\frac{N_{x,k}(t)}{t} \to \frac{1}{\mathbb{E}_{\nu_s}[T_s(x, k)]} \quad \text{a.s.}
\] (4.15)
as \( t \to \infty \). Thus, it follows by (4.13) and (4.15) that given \( \{ \nu^i : i \in \mathbb{N} \} \),
\[
\frac{1}{t} \sum_{i=1}^{N_{x,k}(t)-1} h_i(x,k) f_{\mu_i}(x,k) = \frac{N_{x,k}(t) - 1}{t} \cdot \frac{1}{N_{x,k}(t) - 1} \sum_{i=1}^{N_{x,k}(t)-1} h_i(x,k) f_{\mu_i}(x,k)
\]  
(4.16)
\[
\rightarrow \frac{E_{\nu_*}[h_*(x,k)] f_{\nu_*}(x,k)}{E_{\nu_*}[T_*(x,k)]} \quad \text{a.s.,}
\]

as \( t \to \infty \). Note that \( \pi_{\nu_*}(x,k) = \frac{E_{\nu_*}[h_*(x,k)]}{E_{\nu_*}[T_*(x,k)]} \). Since the service rate is bounded, we have
\[
\sup_{i \in \mathbb{N}} \frac{E[h_i(x,k)f_{\mu_i}(x,k)]}{t} \to 0.
\]  
(4.17)

Then, by (4.11), (4.16), and (4.17), and applying the dominated convergence theorem, we have
\[
\lim_{t \to \infty} \frac{1}{t} E \left[ \int_0^t 1(X(s) = x, K(s) = k)f_{\mu(s)}(x,k) \, ds \bigg| \{ \nu^i : i \in \mathbb{N} \} \right] = \pi_{\nu_*}(x,k)f_{\nu_*}(x,k). \]  
(4.18)

Similarly, for the rejection cost in (4.8), by using (4.14), we obtain
\[
\lim_{t \to \infty} \frac{1}{t} E \left[ p \sum_{i=1}^{m_i} (A(\tau_i) - A(\tau_{i-1})) \int_{\tau_{i-1}}^{\tau_i} 1(X(s) = N) \, ds \right] = p\lambda(\pi_{\nu_*}(N,1) + \pi_{\nu_*}(N,0)). \]  
(4.19)

Note that the expectation of the fourth term on the LHS of (4.8) is bounded. Therefore, by using (4.8), (4.10), (4.18), and (4.19), we have shown (4.9). This completes the proof. \( \square \)

**Remark 4.1.** To extend the results to the problem with a nonlinear relationship between the breakdown rate and service rate, one may apply the same analysis by replacing (2.17) with a polynomial function, where the coefficients may be initially unknown. One may change the likelihood equations in (4.1) accordingly and use the generalized least square estimation in [26]. On the other hand, instead of assuming the functional form of the relationship between the breakdown and service rates, one may treat the breakdown rate as a general function of the service rate. In this case, some non-parametric approaches in the study of online problems for inventory models may be adopted to study the service rate control problem; see, for example, [45–47] and references therein.

### 5. Numerical examples

In this section, we show the numerical results for the queueing system with finite capacity as in Section 3 and the adaptive control problem as in Section 4. In Section 5.1, we provide the results for the optimal service rate controls under different parameters of the system dynamics and the cost functions. In Section 5.2, we present the simulation study for the convergence of regret under adaptive controls.

We first determine the cost functions and the queueing system for the numerical study. The holding/delay cost function in all examples satisfies \( H(x) = C_h x^2 \) for \( x \geq 0 \), where \( C_h \) is a positive constant. We set the effort cost function \( R(\mu) = C_r \mu^2 \) for \( \mu \in \{0\} \cup U \), where \( C_r \) is a positive constant and \( U = [\mu, \bar{\mu}] \) with \( \mu, \bar{\mu} > 0 \). We consider the cases in which the number of jobs in the system is truncated at \( N \). When there are \( N \) jobs in the system, new arrivals are rejected with the cost \( p \) for a single job. Recall that the arrival rate is denoted by \( \lambda \) and the breakdown rate is assumed to be a linear function of the service rate satisfying \( \beta_\mu(\mu) = \kappa_1 + \kappa_2 \mu \) for \( \mu \in \{0\} \cup U \). \( \beta_d \) is used to represent the maintenance rate and is a positive constant. The maintenance cost is \( C_m \) per unit of time.
5.1. The optimal service rate controls. The parameters for the system dynamics are listed in Table 1. We consider a scenario under different cost settings and compare the results of the optimal service rate controls. As mentioned in Remark 2.2, the ratio \( \beta_d / (\beta_d + \beta_u(\mu)) \) represents the proportion of up times on average for the system under the service rate \( \mu \). The proportion of up times approximately ranges from 73.37\% to 89.05\% on average. As a result, the range of the effective service rate becomes \([3.56, 8.80]\).

In this scenario, we use days as the time unit for the parameters. Then, the server is likely to break down every 12 days on average if the server runs at its lowest rate, while the server is likely to break down every 4 days on average if the server runs at its highest rate. It takes about 1.5 days to repair the server on average. The setting of breakdown and maintenance rates is very close to the real data from [48], where the server corresponds to a coal unloader and the jobs correspond to trainloads waiting to be unloaded. Here we assume that the unloading rate is adjustable and the breakdown rate depends on the unloading rate.

<table>
<thead>
<tr>
<th>( \lambda )</th>
<th>( \mu )</th>
<th>( \bar{\mu} )</th>
<th>( \beta_d )</th>
<th>( \kappa_1 )</th>
<th>( \kappa_2 )</th>
<th>( N )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>4</td>
<td>12</td>
<td>2/3</td>
<td>1/500</td>
<td>1/50</td>
<td>25</td>
</tr>
</tbody>
</table>

In Table 2, we provide the parameters of the cost functions in the scenario. In total, there are 15 parameter combinations for the numerical study of the optimal service rate controls. As shown in Table 2, the cost parameters taking values in the set \( \{1, 10\} \) are permuted to show the impact of costs on the optimal policy.

<table>
<thead>
<tr>
<th>Settings</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_r ) (dollar/hour)</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( C_h ) (dollar/hour)</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( C_m ) (dollar/hour)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p ) (dollar/hour)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Settings</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service Utilization(%)</td>
<td>67.5</td>
<td>89.3</td>
<td>60.8</td>
<td>66.8</td>
<td>66.6</td>
<td>66.6</td>
<td>88.5</td>
<td>89.5</td>
<td>61.0</td>
<td>60.8</td>
<td>67.2</td>
<td>66.7</td>
<td>67.3</td>
<td>88.4</td>
<td>60.6</td>
</tr>
<tr>
<td>Rejection Rate(%)</td>
<td>8.0</td>
<td>55.6</td>
<td>8.5</td>
<td>9.3</td>
<td>8.4</td>
<td>8.7</td>
<td>55.7</td>
<td>54.6</td>
<td>9.2</td>
<td>8.3</td>
<td>8.4</td>
<td>9.0</td>
<td>7.6</td>
<td>54.8</td>
<td>8.3</td>
</tr>
</tbody>
</table>

The rejection rate is the ratio between number of rejected jobs and the total number of arrivals. The rejection rate and the service utilization from implementing the optimal policy are listed in the Table 3, which are obtained via simulating 500,000 events.

For every set of parameters, we compute the optimal service rate policy by solving (3.8)–(3.10). Figure 1 shows the optimal policies for the scenario, where the \( x \)-axis represents the number of jobs in the system and the \( y \)-axis corresponds to the service rate under the optimal policy. We find that the optimal policies may not be monotone in the number of jobs in the system. Specifically, under the cost settings 2,7,8 and 14, the holding cost parameters are relatively low whereas the effort cost parameter is relatively high. In these cost parameter settings, Figure 1 shows that the optimal service rate policies are non-monotone. For all the scenarios under the cost settings 3, 9 and 15, since the effort cost parameters are set as “low” while the holding cost parameters are set
as “high”, the optimal service rates are chosen at the highest value when the number of jobs in the system is large.

In addition, the rejection rates in Table 3 indicate that when the effort cost is high, the system tends to have a high rejection rate (see, for example, cost settings 2, 7, 8, and 14). This is reasonable because the optimal control tends to run the server at a lower service rate when the effort cost is high (see, for example, controls under cost settings 2, 7, 8, and 14 in Figure 1), and the system is likely to be at a high congestion level. Therefore, it is more likely to observe rejections. When the holding cost increases, the rejection rate decreases even when the effort cost is high (see, for example, cost settings 6, 12, and 13). This is justifiable as an increased holding cost could result in optimal control policies that prevent the system from entering a high congestion level.

However, increasing the rejection rate when the effort cost is high does not have a significant impact on the rejection rate (see, for instance, cost settings 8 and 14). From the optimal controls under cost settings 2 and 8 in Figure 1, we can see that an increased rejection penalty does not change the policy significantly when the effort cost is high. Therefore, the system has a high likelihood of entering a high congestion level. Hence, the rejection rate remains high.

In summary, when the effort cost is high, the system controller may choose a lower service rate to decrease the likelihood of unplanned downtime even when the system is near its capacity limit, which results in a non-monotone control policy. When the holding cost is high, the controller may run the server at a relatively high service rate to avoid congestion despite the risk of encountering breakdowns.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Optimal control policies under different cost settings.}
\end{figure}

5.2. The adaptive service rate control problem. We first introduce the procedure used to solve the adaptive service rate control problem. Recall that the parameters $\kappa_1$ and $\kappa_2$ are initially unknown and take values in compact sets. At the beginning of the implementation, we randomly choose the initial parameters $\hat{\kappa}_0^1$ and $\hat{\kappa}_0^2$ from the sets $K_1$ and $K_2$, respectively. In the numerical study, we assume $\beta_u(\mu) \leq \beta_d$ for $\mu \leq \mu \leq \bar{\mu}$, and $K_1$ and $K_2$ are constructed to ensure that the inequality to hold.

We estimate the parameters for every 50 jumps of the process $(X, K)$. The cycle of each estimation is identified as an iteration. At each iteration, we simulate the system under the control policy which is obtained by solving equations (3.8)–(3.10). During the calculation, we use the estimation of parameters $\kappa_1$ and $\kappa_2$ in the equations. The estimation of parameters is updated based on the solution of (4.1) in which we use the data collected from simulations. We project the estimated parameters to the boundaries of $K_1 \times K_2$ if the estimation lies outside the domain $K_1 \times K_2$. We simulate the process $(X, K)$ under the adaptive service rate controls in a finite time horizon. The performance of the algorithm is measured by the average regret

$$ R(n) := \frac{1}{t_n} \mathbb{E}[F(t_n)] - \rho^* $$

(5.1)
for positive integer $n \leq L$, where the cost $F(t)$ is defined in (4.8) and $L$ denotes the number of timestamps in the simulation study.

We set $L$ as 1000. The parameter $t_L$ and the sets $\mathcal{K}_1$ and $\mathcal{K}_2$ for each setting are shown in Table 4. In each setting, $t_L$ is chosen to be sufficiently large such that the average regret is near zero when $n$ is large. In Figure 2, the $x$-axis represents the timestamps in the simulations with the difference between the timestamps equal to $t_L/L$, and the $y$-axis corresponds to the average regret at each timestamp. In each case, the expectation is approximated by the average over the values of 300 trajectories. We conduct experiments for the adaptive service rate control problem under the cost parameter settings 13 and 14 for the scenario shown in Table 1. As shown in Table 3, the rejection rate under cost setting 14 is significantly larger than the one under cost setting 13. The experiments under the other cost parameter settings are similar. Here we focus on two sets of cost parameters for simplicity. As shown in Figure 2, the average regrets converge to 0. This verifies the theoretical results in Theorem 4.1.

### Table 4. The sets $\mathcal{K}_1$ and $\mathcal{K}_2$ and the parameter $t_L$.

<table>
<thead>
<tr>
<th>Cost Setting</th>
<th>$\mathcal{K}_1$</th>
<th>$\mathcal{K}_2$</th>
<th>$t_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>$[10^{-5}, 0.665]$</td>
<td>$[10^{-5}, 0.055]$</td>
<td>2011</td>
</tr>
<tr>
<td>14</td>
<td>$[10^{-5}, 0.665]$</td>
<td>$[10^{-5}, 0.055]$</td>
<td>2357</td>
</tr>
</tbody>
</table>

**Figure 2.** The average regret $R(n)$ under cost settings 13 and 14.

**Acknowledgements**

This work was supported in part by the US National Science Foundation grants DMS-1715875 and DMS-2108683, and Army Research Office grant W911NF-17-1-0019.

**References**


[37] B. Liu, Q. Xie, and E. Modiano, *Reinforcement learning for optimal control of queueing systems*, 2019 57th annual allerton conference on communication, control, and computing (allerton), 2019, pp. 663–670. ↑4


