2018 15th International Conference on
Control, Automation, Robotics and Vision (ICARCV)
Singapore, November 18-21, 2018

# Stochastic Optimal Control of Dynamic Queue Systems: A Probabilistic Perspective

Yulong Gao, Shuang Wu, Karl H. Johansson, Ling Shi, and Lihua Xie

*Abstract*— **Queue overflow of a dynamic queue system gives rise to the information loss (or packet loss) in the communication buffer or the decrease of throughput in the transportation network. This paper investigates a stochastic optimal control problem for dynamic queue systems when imposing probability constraints on queue overflows. We reformulate this problem as a Markov decision process (MDP) with safety constraints. We prove that both finite-horizon and infinite-horizon stochastic optimal control for MDP with such constraints can be transformed as a linear program (LP), respectively. Feasibility conditions are provided for the finite-horizon constrained control problem. Two implementation algorithms are designed under the assumption that only the state (not the state distribution) can be observed at each time instant. Simulation results compare optimal cost and state distribution among different scenarios, and show the probability constraint satisfaction by the proposed algorithms.**

## I. Introduction

The queuing model is a widely studied stochastic dynamic model, which considers decisions about resource allocation for certain services. It has a wide range of applications including business decision-making, industrial management, communication system, computer networks and traffic control [1], [2], [3], [4]. One stream of research on queuing models studies the optimal control policy, which seeks the control action selection to optimize certain performance metrics related to the queue length and control input. If multiple design objectives are involved, the problem becomes a constrained optimization, which aims to find a policy to optimize one of the objectives while guaranteeing the value of the other objectives within a certain range.

It is worth noting that there are a few main approaches to tackle the optimal queue control problem. As the optimal control can be categorized as a sequential decision making problem, a dynamic programming (DP) is applicable for both finite and infinite horizon problems [5], [6], [7]. However, the dynamic programming fails to cope with the optimization with additional constraints on extra performance requirements. Another approach is based on the empirical measure of the state-action pairs, which is commonly used for infinite horizon problems [7], [8]. This approach can deal with the constraints on a design objective which is linear in the empirical measure of the state-action pairs.

In practice, it is desirable to keep the system state inside a safety region with a high probability. For example, limiting probability of queue overflows for a dynamic queue system is of great interest since queue overflows give rise to the information loss (or packet loss) in the communication buffer or the decrease of throughput in the transportation network. The classical approaches mentioned above fail to handle this scenario. The dynamic programming cannot capture the requirement specified by probability constraints, while the empirical measure only deals with a steady state requirement. As the safety region requirement should be satisfied for the whole trajectory, the transient behavior in a finite horizon should also be considered.

In [9], some attempts have been made to deal with the safety constraints for a general sequential decision making problem in a finite horizon regime. They proposed a dynamic program to obtain a policy for the worst case, which can be efficiently computed by a linear program (LP). This result provides a performance guarantee on the lower bound of the system performance.

In this work, we consider the stochastic optimal control problem of dynamic queue systems when imposing some probability constraints on queue overflows. This problem can be transformed as a Markov decision process (MDP) with safety constraints, as in [9]. Different from [9], we tackle the probabilistic constraint by directly solving the problem. Our main continuations are summarized as follows.

(1) We provide an LP reformulation for both finite-horizon and infinite-horizon cases, respectively. Compared with [9], our LP exactly solves the finite-horizon stochastic optimal control problem with probability constraints without any performance loss or conservatism. In addition, we explore the feasibility conditions for the finite-horizon constrained problem.

(2) We propose two implementation algorithms under the assumption that only the state (not the state distribution) can be observed at each time instant. The first algorithm is in a receding-horizon manner according to the state-adaptive optimal policy of the finite-horizon constrained problem while the second one is based on the stationary optimal policy of the infinite-horizon constrained problem.

The remainder of this paper is organized as follows. Section II provides the problem formulation. Section III presents the feasibility conditions and the LP reformulation for the

The work by Y. Gao and K. H. Johansson is supported by the Knut and Alice Wallenberg Foundation, the Swedish Strategic Research Foundation, and the Swedish Research Council. The work by S. Wu and L. Shi is supported by an HKUST-KTH Partnership FP804.

Y. Gao and K. H. Johansson are with ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm 10044, Sweden yulongg@kth.se, kallej@kth.se

S. Wu and L. Shi are with Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong swuak@ust.hk, eesling@ust.hk

L. Xie is with School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore elhxie@ntu.edu.sg

finite-horizon constrained problem. Section IV develops the LP reformulation for the infinite-horizon constrained problem. An example in Section V illustrates the effectiveness of our approach. Section VI concludes this paper.

**Notation.** Let $\mathbb{N}$ denote the set of nonnegative integers and $\mathbb{R}$ denote the set of real numbers. For some $q,s \in \mathbb{N}$ and $q<s$, let $\mathbb{N}_{\geq q}$ and $\mathbb{N}_{[q,s]}$ denote the sets $\{r \in \mathbb{N} \mid r \geq q\}$ and $\{r \in \mathbb{N} \mid q \leq r \leq s\}$, respectively. When $\leq$, $\geq$, $<$, and $>$ are applied to vectors, they are interpreted element-wise. $\mathbf{1}$ denotes a vector of all ones with appropriate dimension. *Pr* denotes the probability measure. $\mathbb{E}$ denotes the expectation of a random variable. For a set $\mathbb{X}$, $\mathscr{P}(\mathbb{X})$ is the collection of the probability distribution on a Borel subset of $\mathbb{X}$.

## II. Problem Formulation

### A. Dynamic queue model

Consider a dynamic queue system

$$q_{t+1} = \min\{\max(q_t - u_t, 0) + a_t, \bar{q}\},\ t \in \mathbb{N}, \tag{1}$$

where $q_t$ is the queue length, $u_t$ is the control input, $a_t$ is the arrival, and $\bar{q} \in \mathbb{N}$ is the size of the queue. The control input is constrained by

$$u_t \in \mathbb{U} \triangleq \mathbb{N}_{[0,\bar{u}]},\ \forall t \in \mathbb{N}, \tag{2}$$

where $\bar{u} \in \mathbb{N}$ is the upper bound on control input. Furthermore, the arrivals $a_t$ are independent and identically distributed (i.i.d.) with probability mass function

$$f(k) = Pr\{a_t = k\},\ \forall t \in \mathbb{N}. \tag{3}$$

### B. Probability constraints

In this work, we are interested in the probability of $q_t = \bar{q}$ at time instant $t$, i.e., the probability that the queue overflows occur. In practice, the queue overflows increase the risk in losing the information or data in the communication buffer, or decreasing the throughput in the transportation network. Here, we impose a probability constraint on $q_t = \bar{q}$ at each time step, i.e.,

$$Pr\{x_t = \bar{q}\} \leq \varepsilon, \forall t \in \mathbb{N}, \tag{4}$$

where $0 \leq \varepsilon \leq 1$ is a prescribed probability level.

### C. Reformulation as a finite MDP

The above dynamic queue system can be formulated as an MDP, denoted by $\mathscr{S} = (\mathbb{X}, \mathbb{U}, T, c)$:

- state space: $\mathbb{X} = \mathbb{N}_{[0,\bar{q}]}$;
- control space: $\mathbb{U} = \mathbb{N}_{[0,\bar{u}]}$;
- transition kernel:

$$T(y|x,u) = \begin{cases} 0, & \text{if } y < (x-u)^+, \\ f(y-(x-u)^+), & \text{if } (x-u)^+ \leq y < \bar{q}, \\ \sum_{k=\bar{q}-(x-u)^+}^{\infty} f(k), & \text{if } y = \bar{q}, \end{cases}$$

- one-stage cost: $c(x,u)$ and terminal cost: $c_f(x)$.

*Remark 2.1:* The transition kernel above is explicitly defined according to the probability mass function of the arrivals and the queue dynamics (1).

*Remark 2.2:* For a dynamic queue model, the queue length captures the so-called "delay" while the control input corresponds to the "control power". In this work, we do not impose any structure on the cost function. It can be a tradeoff between "delay" and "control power".

Denote $\mathbb{U}_x$ as the admissible control set when the state is $x$, i.e., $\mathbb{U}_x = \{u \in \mathbb{U} \mid \exists y \in \mathbb{X}, T(y|x,u) > 0\}$. And let $|\mathbb{U}_x|$ be the cardinality of $\mathbb{U}_x$.

*Assumption 2.1:* For any $x \in \mathbb{X}$, $|\mathbb{U}_x| \geq 1$.

The policy is restricted to the randomized Markovian policy, denoted by $\mu : \mathbb{X} \to \mathscr{P}(\mathbb{U})$. The initial state probability distribution is denoted by $p_0 \in \mathscr{P}(\mathbb{X})$. The objective of this work is to solve the following two problems.

*Problem 2.1:* Given a finite horizon $N$ and an initial state distribution $p_0$, find, if there exists, an optimal policy $\mu^* = (\mu_0^*, \mu_1^*, \ldots, \mu_{N-1}^*)$ solving the following finite-horizon stochastic optimal control problem

$$\min_{\mu}\ \mathbb{E}\{\sum_{t=0}^{N-1}(c(x_t,u_t)) + c_f(x_N)\} \tag{5a}$$

subject to

$$\forall t \in \mathbb{N}_{[1,N]}:\ Pr\{x_t = \bar{q}\} \leq \varepsilon. \tag{5b}$$

*Problem 2.2:* Find, if there exists, an optimal policy $\mu^* = (\bar{\mu}^*, \bar{\mu}^*, \ldots)$ solving the following infinite-horizon stochastic optimal control problem

$$\min_{\mu}\ \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\{\sum_{t=0}^{N-1}(c(x_t,u_t))\} \tag{6a}$$

subject to

$$\forall t \in \mathbb{N}:\ Pr\{x_t = \bar{q}\} \leq \varepsilon. \tag{6b}$$

## III. Finite-horizon Stochastic Optimal Control

This section aims to provide a strategy to solve Problem 2.1. First, we characterize the feasibility conditions. Then, we reformulate the finite-horizon stochastic optimal control problem as an LP.

### A. Feasibility conditions

Given a finite horizon $N$, a policy $\mu = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ can be represented by a sequence of matrices $\mathbf{M} = (M_0, M_1, \ldots, M_{N-1})$ satisfying

$$\forall t \in \mathbb{N}_{[0,N-1]}:\ \begin{cases} M_t \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}, \\ M_t \geq 0, \\ M_t \mathbf{1} = \mathbf{1}. \end{cases} \tag{7}$$

Each row of $M_t$ is a probability distribution over the control space, which corresponds to a randomized Markovian policy. Then, the state distribution $p_t$ evolves as

$$\begin{aligned} p_{t+1}(x) &= \sum_{y\in\mathbb{X}}\sum_{u\in\mathbb{U}} T(x|y,u) M_t(y,u) p_t(y) \\ &= \sum_{y\in\mathbb{X}} H_t(x,y) p_t(y),\ \forall x \in \mathbb{X}, \forall t \in \mathbb{N}_{[0,N-1]}, \end{aligned} \tag{8}$$

with an initial state distribution $p_0$. In (8), the matrix $H_t$ is defined by

$$H_t(x,y) = \sum_{u\in\mathbb{U}} T(x|y,u) M_t(y,u),\ \forall x,y \in \mathbb{X}. \tag{9}$$

The probability constraints (5b) only involve the state $x = \bar{q}$, which can be rewritten with respect to the state distribution $p_t$ as follows

$$p_t \leq \bar{p},\ \bar{p} = \begin{bmatrix} \mathbf{1} \\ \varepsilon \end{bmatrix},\ \forall t \in \mathbb{N}_{[1,N]}. \tag{10}$$

Given an initial state distribution $p_0$, Problem 2.1 is said to be feasible if there exists a sequence of matrices $\mathbf{M} = (M_0, M_1, \ldots, M_{N-1})$ such that

$$\forall t \in \mathbb{N}_{[0,N-1]}: \begin{cases} p_{t+1} = H_t p_t, \\ p_{t+1} \leq \bar{p}, \end{cases} \tag{11}$$

where $H_t$ is a function of $M_t$ defined in (9). The following proposition provides sufficient conditions to ensure the feasibility of Problem 2.1.

*Proposition 3.1:* Given an initial state distribution $p_0$, Problem 2.1 is feasible if there exist $K_1 \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}$, $K_2 \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}$, $L_1 \in \mathbb{R}^{(\bar{q}+1)\times(\bar{q}+1)}$, $L_2 \in \mathbb{R}^{(\bar{q}+1)\times(\bar{q}+1)}$, $S \in \mathbb{R}^{(\bar{q}+1)\times(\bar{q}+1)}$, and $s \in \mathbb{R}^{\bar{q}+1}$ satisfying

$$K_1 \geq 0, K_1\mathbf{1} = \mathbf{1}, \tag{12a}$$
$$L_1(x,y) = \sum_{u\in\mathbb{U}} T(x|y,u)K_1(y,u), \forall x,y \in \mathbb{X}, \tag{12b}$$
$$L_1 p_0 \leq \bar{p}, \tag{12c}$$
$$K_2 \geq 0, K_2\mathbf{1} = \mathbf{1}, \tag{12d}$$
$$L_2(x,y) = \sum_{u\in\mathbb{U}} T(x|y,u)K_2(y,u), \forall x,y \in \mathbb{X}, \tag{12e}$$
$$S \geq 0, \tag{12f}$$
$$L_2 + S + s\mathbf{1}^T \geq 0, s + \bar{p} \geq (L_2 + S + s\mathbf{1}^T)\bar{p}. \tag{12g}$$

*Proof:* Given an initial state distribution $p_0$, the satisfaction of probability constraint (5b) at $t = 1$ is equivalent to the existence of $K_1 \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}$ and $L_1 \in \mathbb{R}^{(\bar{q}+1)\times(\bar{q}+1)}$ satisfying (12a)–(12b) such that the initial distribution $p_0$ can be steered to a new distribution $L_1 p_0 \leq \bar{p}$, i.e., (12c).

Then, we need to find the conditions such that $p_t \leq \bar{p}$ with $p_1 \leq \bar{p}$, $\forall t \in \mathbb{N}_{[2,N]}$. It is sufficient to consider the existence of $K_2 \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}$ and $L_2 \in \mathbb{R}^{(\bar{q}+1)\times(\bar{q}+1)}$ satisfying (12d)–(12e) and $L_2 p \leq \bar{p}$, $\forall p \leq \bar{p}$. And following Lemma 1 in [10], the fact that $L_2 p \leq \bar{p}$, $\forall p \leq \bar{p}$ is equivalent to the existence of $S \in \mathbb{R}^{(\bar{q}+1)\times(\bar{q}+1)}$, and $s \in \mathbb{R}^{\bar{q}+1}$ satisfying (12g)–(12g).

The proof is completed. ∎

*Remark 3.1:* In Problem 2.1, we do not require a feasible initial state distribution, i.e., $p_0 \leq \bar{p}$. Thus, the feasibility conditions consist of two parts: one is to enforce the initial state distribution to be feasible and the other is to guarantee that $p_t \leq \bar{p}$ is controlled invariant.

### B. Reformulation as LP

From (8), we have $p_{t+1} = H_t \cdots H_2 H_1 p_0$ and $H_t$ is a function of decision variables $M_t$. Note that the resulting constraints are non-convex. In [9], it is argued that finding a feasible solution to Problem 2.1 is challenging and DP fails to solve this problem directly. Hence, the authors of [9] propose an approximate method to search the feasible region and provide a modified DP to provide a lower bound on the optimal reward or an upper bound on the optimal cost. Different from [9], we can exactly solve Problem 2.1 without any performance loss or conservatism. The following theorem shows that Problem 2.1 is equivalent to an LP. Denote by $J_N^*(p_0)$ the optimal cost of Problem 2.1.

*Theorem 3.1:* Problem 2.1 can be reformulated as the following LP:

$$\min_{\mathbf{V}=(V_0,V_1,\ldots,V_{N-1},V_N)} \sum_{t=0}^{N-1}\sum_{x\in\mathbb{X}}\sum_{x\in\mathbb{U}} c(x,u)V_t(x,u) + \sum_{x\in\mathbb{X}} c_f(x)V_N(x) \tag{13a}$$

subject to

$$V_1 \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}, V_1 \geq 0, \mathbf{1}^T V_1 \mathbf{1} = 1, V_1\mathbf{1} = p_0, \tag{13b}$$

$\forall t \in \mathbb{N}_{[1,N-1]}$:

$$\begin{cases} V_t \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}, V_t \geq 0, \mathbf{1}^T V_t \mathbf{1} = 1, \\ \sum\limits_{u\in\mathbb{U}} V_t(x,u) - \sum\limits_{y\in\mathbb{X}}\sum\limits_{u\in\mathbb{U}} T(x|y,u)V_{t-1}(y,u) = 0, \forall x \in \mathbb{X}, \\ V_t\mathbf{1} \leq \bar{p}, \end{cases} \tag{13c}$$

$$\begin{cases} V_t \in \mathbb{R}^{\bar{q}+1}, V_N \geq 0, \mathbf{1}^T V_N = 1, \\ V_N(x) - \sum\limits_{y\in\mathbb{X}}\sum\limits_{u\in\mathbb{U}} T(x|y,u)V_{N-1}(y,u) = 0, \forall x \in \mathbb{X}, \\ V_N \leq \bar{p}, \end{cases} \tag{13d}$$

which gives

$$J_N^*(p_0) = \sum_{t=0}^{N-1}\sum_{x\in\mathbb{X}}\sum_{x\in\mathbb{U}} c(x,u)V_t^*(x,u) + \sum_{x\in\mathbb{X}} c_f(x)V_N^*(x), \tag{14}$$

where $\mathbf{V}^* = (V_0^*, V_1^*, \ldots, V_{N-1}^*, V_N^*)$ is the optimal solution of LP (13).

*Proof:* First of all, Problem 2.1 can be rewritten as

$$\min_{\mathbf{M}=(M_0,M_1,\ldots,M_{N-1})} \sum_{t=0}^{N-1}\sum_{x\in\mathbb{X}}\sum_{x\in\mathbb{U}} c(x,u)M_t(x,u)p_t(x) + \sum_{x\in\mathbb{X}} c_f(x)p_N(x) \tag{15a}$$

subject to

$$M_1 \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}, M_1 \geq 0, M_1\mathbf{1} = \mathbf{1}, \tag{15b}$$

$\forall t \in \mathbb{N}_{[1,N-1]}$:

$$\begin{cases} M_t \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}, M_t \geq 0, M_t\mathbf{1} = \mathbf{1}, \\ p_t(x) = \sum\limits_{y\in\mathbb{X}}\sum\limits_{u\in\mathbb{U}} T(x|y,u)M_{t-1}(y,u)p_{t-1}(y), \forall x \in \mathbb{X}, \\ p_t \leq \bar{p}, \end{cases} \tag{15c}$$

$$\begin{cases} p_N(x) = \sum\limits_{y\in\mathbb{X}}\sum\limits_{u\in\mathbb{U}} T(x|y,u)M_{N-1}(y,u)p_{N-1}(y), \forall x \in \mathbb{X}, \\ p_N \leq \bar{p}. \end{cases} \tag{15d}$$

Define $V_t(x,u) = M_t(x,u)p_t(x)$, $\forall x \in \mathbb{X}, u \in \mathbb{U}, \forall t \in \mathbb{N}_{[0,N-1]}$, and $V_N = p_N$. The constraints (15b) and initial state distribution can be reformulated as (13b). And the constraints (15c)–(15d) can be written as (13c)–(13d). Furthermore, the cost function (15a) is equivalent to (18a). Thus, we conclude that Problem 2.1 is equivalent to the LP (13). ∎

*Remark 3.2:* The decision variables $V_t$ in the LP (13) is an occupation measure over $\mathbb{X} \times \mathbb{U}$. The row sum of $V_t$ corresponds the state distribution at time $t$.

*Corollary 3.1:* If Problem 2.1 is feasible and Assumption 2.1 is satisfied, the optimal policy $\mathbf{M}^* = (M_0^*, M_1^*, \ldots, M_{N-1}^*)$ of Problem 2.1 can be characterized as $\forall t \in \mathbb{N}_{[0,N-1]}$,

$$M_t^*(x,u) = \begin{cases} \frac{V_t^*(x,u)}{\sum_{v\in\mathbb{U}} V_t^*(x,v)}, & \text{if } \sum_{v\in\mathbb{U}} V_t^*(x,v) > 0 \\ \frac{1}{|\mathbb{U}_x|}, & \text{if } \sum_{v\in\mathbb{U}} V_t^*(x,v) = 0, u \in \mathbb{U}_x, \\ 0, & \text{otherwise,} \end{cases} \quad (16)$$

where $V_t^*$ is the optimal solution to the LP (13).

*Proof:* This result follows from the definition of $V_t(x,u) = M_t(x,u)p_t(x)$, $\forall x \in \mathbb{X}, u \in \mathbb{U}, \forall t \in \mathbb{N}_{[0,N-1]}$, and $M_1\mathbf{1} = \mathbf{1}$. Note that when $\sum_{v\in\mathbb{U}} V_t^*(x,v) = 0$, any distribution over control space does not change the optimal cost. Here, we choose the uniform distribution over the admissible control set of state $x$. ■

### C. Implementation algorithm

In practice, the state distribution at each time step cannot be observed. Hence, it is more reasonable to design a state-feedback implementation algorithm. The following algorithm provides a receding-horizon implementation strategy for the finite-horizon case. At each time step, we first derive the optimal policy by solving Problem 2.1 initialized from the observed state. Then, the control input is chosen in a random way according to the resulting optimal policy.

**Algorithm 1** Finite-horizon Implementation Algorithm

1: Initialize $t = 0$.
2: Observe the current queue length $q_t$ and set the initial distribution $p_0 = e_{q_t+1}$ where $e_k$ is a vector with $k$th element being one and other elements being zero.
3: Solve the LP (13) and obtain a sequence of matrices $\mathbf{M}^* = (M_0^*, M_1^*, \ldots, M_{N-1}^*)$ based on (16).
4: Select the $(q_t+1)$th row of $M_0^*$ as $\mu_t^*$, i.e., the optimal policy when the state is $q_t$.
5: Randomly choose one control input following the distribution $\mu_t^*$ and implement it.
6: Set $t = t+1$ and go to step 2.

## IV. Infinite-horizon Stochastic Control

In this section, we focus on handling Problem 2.2. We will first investigate an LP reformulation for infinite-horizon MDP with probability constraints. Then, we will design the implementation algorithm for this problem.

The stationary policy $\mu = (\bar{\mu}, \bar{\mu}, \ldots)$ can be represented by a matrix $F$ satisfying

$$\begin{cases} F \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}, \\ F \geq 0, \\ F\mathbf{1} = \mathbf{1}. \end{cases} \quad (17)$$

### A. Reformulation as LP

The following theorem provides an LP formulation for Problem 2.2.

*Theorem 4.1:* Problem 2.2 can be reformulated as the following LP:

$$\min_G \sum_{x\in\mathbb{X}}\sum_{x\in\mathbb{U}} c(x,u)G(x,u) \quad (18a)$$

subject to

$$G \in \mathbb{R}^{(\bar{q}+1)\times(\bar{u}+1)}, G \geq 0, \mathbf{1}^T G\mathbf{1} = 1, \quad (18b)$$

$$\sum_{u\in\mathbb{U}} G(x,u) - \sum_{y\in\mathbb{X}}\sum_{u\in\mathbb{U}} T(x|y,u)G(y,u) = 0, \forall x \in \mathbb{X}, \quad (18c)$$

$$G\mathbf{1} \leq \bar{p}. \quad (18d)$$

*Proof:* Please refer to [8] for a detailed proof. The difference from [8] is the constraint (18d). ■

*Corollary 4.1:* If Problem 2.2 is feasible, the optimal policy $F^*$ of Problem 2.2 can be recovered by

$$F^*(x,u) = \begin{cases} \frac{G^*(x,u)}{\sum_{v\in\mathbb{U}} G^*(x,v)}, & \text{if } \sum_{v\in\mathbb{U}} G^*(x,v) > 0 \\ \frac{1}{|\mathbb{U}_x|}, & \text{if } \sum_{v\in\mathbb{U}} G^*(x,v) = 0, u \in \mathbb{U}_x, \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

*Proof:* The proof is similar to that of Corollary 3.1. ■

*Remark 4.1:* The feasible set of Problem 2.2 is constructed by the constraints (18b)–(18d). Thus, we do not impose feasibility conditions for the infinite-horizon case. In addition, the decision variable $G$ in the LP (18) is also an occupation measure over $\mathbb{X} \times \mathbb{U}$. The row sum of $G$ corresponds the stationary state distribution.

### B. Implementation algorithm

For the infinite-horizon case, we provide an implementation strategy in Algorithm 2. The difference from Algorithm 1 is that the infinite-horizon problem is solved once to obtain a stationary policy at the beginning. The control input at each time instant is chosen according to this policy and the observed state. The state-feedback of Algorithm 2 is static and "uniform" while the state-feedback of Algorithm 1 is state-adaptive.

**Algorithm 2** Infinite-horizon Implementation Algorithm

1: Solve the LP (18) and obtain the optimal policy $F^*$ based on (19).
2: Initialize $t = 0$.
3: Observe the current queue length $q_t$ and select the $(q_t+1)$th row of $F^*$ as $\mu_t^*$, i.e., the optimal policy when the state is $q_t$.
4: Randomly choose one control input following the distribution $\mu_t^*$ and implement it.
5: Set $t = t+1$ and go to step 2.

## V. Simulations

In this section, we will simulate a dynamic queue system with parameters as follows

$$\bar{q} = 6,\ \bar{u} = 5,\ \text{and } a(t) \sim poiss(\lambda), \quad (20)$$

where $poiss(\lambda)$ denotes the poisson distribution with average rate $\lambda$. Set $\lambda = 2$. The cost is defined by

$$c(x,u) = \alpha x + \beta u,\ c_f(x) = \gamma x, \quad (21)$$

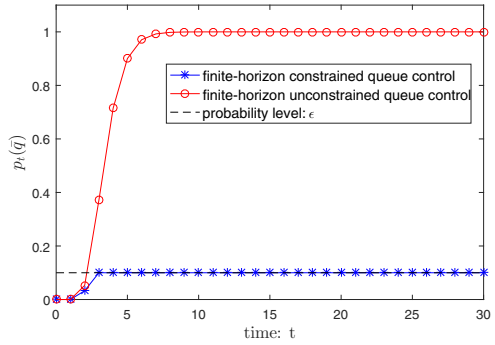



Fig. 1. The probability of $q_t = \bar{q}$ under finite-horizon constrained queue control and finite-horizon unconstrained queue control.

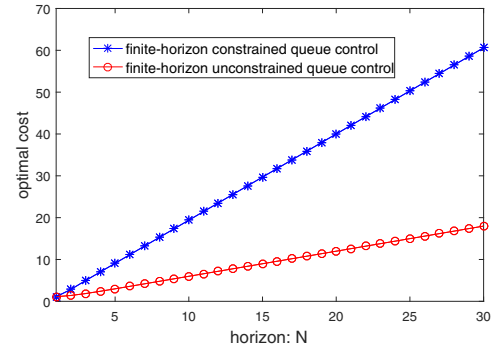



Fig. 2. The optimal cost under finite-horizon constrained queue control and finite-horizon unconstrained queue control for different horizons.

where the weight parameters $\alpha, \beta$ and $\gamma$ are nonnegative constants. We always choose the initial queue length $q_0 = 1$, which is a feasible starting point.

### A. Comparisons and Analysis

We first clarify some notions used in the following. The notion "finite-horizon constrained queue control" corresponds to Problem 2.1, the notion "finite-horizon unconstrained queue control" corresponds to only minimizing the objective function of Problem 2.1 without probability constraints, and the notion "infinite-horizon constrained queue control" refers to Problem 2.2.

*Constraint satisfaction*: Choose $\alpha = 0.05$, $\beta = 1$, $\gamma = 1$, $N = 30$, and $\varepsilon = 0.1$. In Fig. 1, we compare the probability of $q_t = \bar{q}$ under finite-horizon unconstrained queue control and finite-horizon constrained queue control. It shows that our LP approach can generate a distribution satisfying the constraint (5b) at each time instant while the unconstrained queue control cannot guarantee the constraint satisfaction even though the starting point is feasible.

*Cost vs. horizon*: Choose $\alpha = 0.05$, $\beta = 1$, $\gamma = 1$, and $\varepsilon = 0.1$. In Fig. 2, we compare the optimal cost under finite-horizon unconstrained queue control and finite-horizon unconstrained queue control for different horizons $N$. The cost of the constrained queue control is much higher than that of the unconstrained queue control. This is the price for the constraint satisfaction.

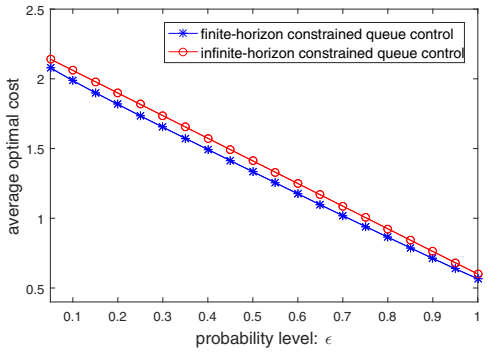



Fig. 3. The average optimal cost under finite-horizon constrained queue control and infinite-horizon constrained queue control for different probability level $\varepsilon$.

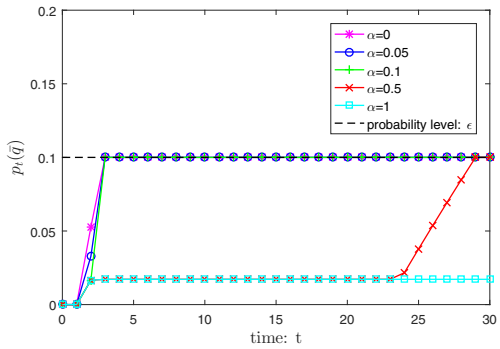
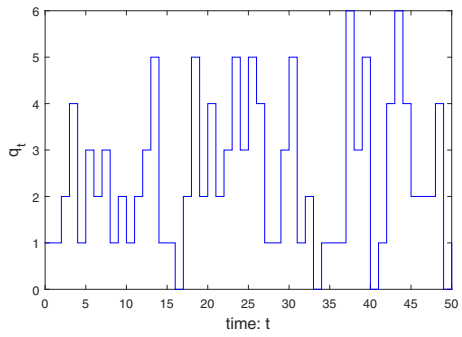


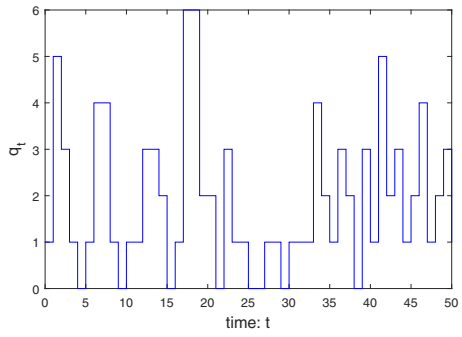

Fig. 4. The probability of $q_t = \bar{q}$ under finite-horizon constrained queue control for different $\alpha$.

*Average cost vs. probability level*: Choose $\alpha = 0.05$, $\beta = 1$, $\gamma = 0$, and $N = 30$. Setting $\gamma = 0$ is to provide a fair comparison between the finite-horizon case and the infinite-horizon case. In Fig. 3, we present the average optimal cost under the finite-horizon constrained queue control and the infinite-horizon constrained queue control for different probability levels $\varepsilon$'s. It further explains the relation between the optimal cost and the constraint: the tighter the constraint is, the higher optimal cost will be. Another observation is that the average optimal cost under the infinite-horizon constrained queue control is always higher than that under the finite-horizon constrained queue control. One interpretation is that the optimal policy in the infinite-horizon case is stationary and "uniform" while the optimal policy in the finite-horizon case is state-adaptive according to the initial state.

*Probability distribution vs. cost weight*: Choose $\beta = 1$, $\gamma = 1$, $N = 30$, and $\varepsilon = 0.1$. In Fig. 4, we compare the probability of $q_t = \bar{q}$ under finite-horizon constrained queue control for different weight parameters $\alpha$s. The intuitive meaning of $a$ and $\beta$ is the tradeoff between the "delay" (i.e., the queue length $q_t$) and the "control power" (i.e., the control input $u_t$). Fig. 4 shows that the increase of $\alpha$ compresses the probability distributions of $q_t = \bar{q}$ in order to achieve a shorter "delay".

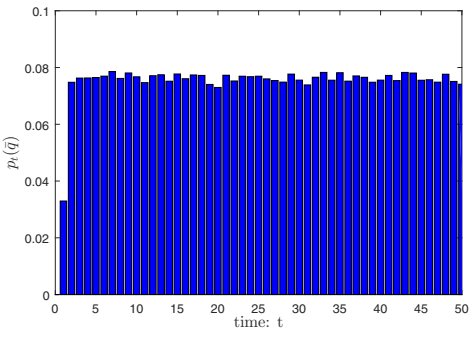



(a) Queue length evolutions under Algorithm 1

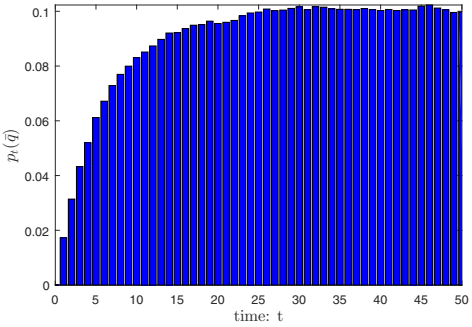



(b) Queue length evolutions under Algorithm 2

Fig. 5. Queue length evolutions under Algorithms 1 and 2 for one realization.




(a) The probability of $q_t = \bar{q}$ under Algorithm 1




(b) The probability of $q_t = \bar{q}$ under Algorithm 2

Fig. 6. The probability of $q_t = \bar{q}$ under Algorithms 1 and 2 over 50000 realizations.

### B. Implementations

Choose $\alpha = 0.05$, $\beta = 1$, $\gamma = 1$, $N = 30$, and $\varepsilon = 0.1$. By implementing Algorithms 1 and 2, we run 50000 realizations for 50 time instants from the same feasible starting point $q_0 = 1$, respectively. One realization of the queue length $q_t$ is shown in Fig. 5 under Algorithms 1 and 2, respectively. Fig. 6 presents the corresponding probability of $q_t = \bar{q}$ over 50000 realizations. As shown in subfigure (a), the probability of $q_t = \bar{q}$ generated by Algorithm 1, which uses the optimal policy of Problem 2.1 in the receding-horizon fashion, is always lower than the bound $\varepsilon = 0.1$ at each time instant. In subfigure (b), Algorithm 2 can steer the probability of $q_t = \bar{q}$ close to the the bound $\varepsilon = 0.1$ by utilizing the optimal policy of Problem 2.2, which is consistent with the resulting stationary optimal distribution, i.e., $p = [0.1218\ 0.2457\ 0.2479\ 0.1667\ 0.0840\ 0.0339\ 0.1000]^T$. In comparison with the infinite-horizon case, one explantation that the probability of $q_t = \bar{q}$ is always lower than $\varepsilon = 0.1$ in the finite-horizon case is the state-adaptive policy resulting from the receding-horizon manner.

## VI. Conclusion

This paper investigated the stochastic optimal control problem of dynamic queue systems when imposing probability constraints on queue overflows. This problem was reformulated as an MDP with state distribution constraints. Both the finite-horizon and the infinite-horizon stochastic optimal control for the MDP with probability constraints were exactly transformed as an LP, respectively. Feasibility conditions were explored for the finite-horizon case. Two implementation algorithms were designed under the assumption that only the state (not the state distribution) can be observed at each time instant. Simulation results illustrate the effectiveness of our approach.

## References

[1] L. Kleinrock, "*Queueing systems*," John Wiley & Sons, 1976.
[2] F. Gebali, "*Analysis of computer and communication networks*," Springer, 2008.
[3] Y. Cui, Q. Q. Huang, V. K. Lau, "Queue-aware dynamic clustering and power allocation for network MIMO systems via distributed stochastic learning," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 12291238, 2011.
[4] Y. L. Gao, M. Jafarian, K. H. Johansson, L. H. Xie, "Distributed freeway ramp metering: Optimization on flow speed," *In Proceedings of IEEE 56th Conference on Decision and Control*, 2017, pp. 5654-5659.
[5] L. I. Sennott, "*Stochastic dynamic programming and the control of queueing systems*," John Wiley & Sons, 2009.
[6] D. P. Bertsekas, "*Dynamic programming and optimal control*," Athena Scientific, 1995.
[7] O. Hernández-Lerma, J. B. Lasserre, "*Discrete-time markov control processes: basic optimality criteria*," Springer, 1996.
[8] E. Altman, "*Constrained markov decision processes*," CRC Press, 1999.
[9] M. E. Chamie, Y. Yu, B. Açıkmeşe, "Convex synthesis of randomized policies for controlled markov chains with density safety upper bound constraints," *In Proceedings of American Control Conference*, 2016, pp. 6290-6295.
[10] N. Demirer, M. E. Chamie, B. Açıkmeşe, "Safe markov chains for on/off density control with observed transitions," *IEEE Transactions on Automatic Control*, vol. 63, no. 5, pp. 1442-1449, 2018.