Robust Control Policy for Closed Queuing Networks with Uncertain Routing Probabilities[1)]

TANG Hao[1,2] XI Hong-Sheng[2] HAN Jiang-Hong[1] YUAN Ji-Bin[1]

[1](*School of Computer and Information, Hefei University of Technology, Hefei* 230009)

[2](*Department of Automation, University of Science and Technology of China, Hefei* 230026)

(E-mail: tangh@ustc.edu/htang@hfut.edu.cn)

Abstract The paper is concerned with the robust control problems for exponential controlled closed queuing networks (CCQNs) under uncertain routing probabilities. As the rows of some parameter matrices such as infinitesimal generators may be dependent, we first transform the objective vector under discounted-cost criteria into a weighed-average cost. Through the solution to Poisson equation, *i.e.*, Markov performance potentials, we then unify both discounted-cost and average-cost problems to study, and derive the gradient formula of the new objective function with respect to the routing probabilities. Some solution techniques are related for searching the optimal robust control policy. Finally, a numerical example is presented and analyzed.

Key words Controlled closed queuing networks, routing probabilities, robust control, Markov performance potentials

1 Introduction

Many real-world discrete event dynamic systems (DEDSs) may be modeled as controlled closed queuing networks (CCQNs), and can be solved by using Markov performance theory if the routing probabilities are all deterministic and known[1,2]. For large-scale practical systems, the routing probabilities may be unavailable or even slow-varying, and the decision-maker only knows their range. Then, the optimization problem is to calculate the robust control policy, that is, to find the optimal policy under the worst case. For Markov decision problems (MDPs) with independent uncertain transition rates, a policy iteration approach for solving the robust control policy was introduced in [3], and the potential-based policy iteration was developed in [4]. In CCQNs, the uncertainty of routing probabilities will lead to the correlation between rows of the infinitesimal generator, and the policy iteration will be inapplicable. In this paper, we will provide a uniform framework to solve the robust decision problems for both discounted and average criteria by some suitable transformation of the discounted-cost vector.

2 Problem formulation

Consider a CCQN with M servers and N customs[2]. The routing matrix is $q=[q_{ij}]$, and the state space is $\Phi=\{n=(n_1,n_2,\cdots,n_M):\sum_{i=1}^{M}n_i=N\}$ with n_i denoting the number of the customs at the i-th server. A stationary policy is $v=(v(1),\cdots,v(K))$ with $v(n)=(\mu_{1,n},\mu_{2,n},\cdots,\mu_{M,n})\in D(n)$, where $\mu_{i,n}$ is the serving rate of server i at state n, and $D(n)$ is a feasible action space. Let Ω_s be the set of all stationary policies, and $N(t)$ the state process. Define $\mu(n)=\sum_{i=1}^{M}\mu_{i,n}$ and $\lambda^v=\max_n\{\mu(n)\}$. Under policy v, the transition matrix of the embedded Markov chain $P^v(q)$ and the infinitesimal generator $A^v(q)$ satisfy $A^v(q)=\Lambda^v(P^v(q)-I)$ with $\Lambda=\mathrm{diag}(\mu(1),\cdots,\mu(K))$. Let the performance vector be $f^v=(f(1,v(1)),\cdots,f(K,v(K)))^\tau$. Similar to [5], define the discounted-cost

1) Supported by National Nature Science Foundation of P. R. China (60404009, 60274012), the Natural Science Foundation of Anhui Province (050420303), and by the Sustentation Project of Hefei University of Technology for the Science and Technology-innovation Groups

Received November 11, 2003; in revised form September 6, 2004

criteria as $\eta_\alpha^v(i,q) = E\{\int_0^{+\infty} \alpha e^{-\alpha t} f(N(t), v(N(t)))\mathrm{d}t | X(0) = i\}, i \in \Phi$. Here, $\alpha > 0$ is a discount factor. Let $\eta_\alpha^v(q) = (\eta_\alpha^v(1,q), \cdots, \eta_\alpha^v(K,q))^\tau$; then[5]

$$\eta_\alpha^v(q) = \alpha(\alpha I - A^v(q))^{-1} f^v \tag{1}$$

For a deterministic CCQN, the optimal policy can be obtained by potential-based policy iteration and value iteration[6,7]. But in some uncertain practical systems, the decision-maker only knows the range of q_{ij}. Let $q_{ij} \in \Theta_{ij}^v$ with Θ_{ij}^v being compact, and $\Theta^v = \{q = [q_{ij}] : q_{ij} \in \Theta_{ij}^v, \sum_j q_{ij} = 1\}$. Then, our goal is to find a policy v^* satisfying $v^* \in \arg\min_{v\in\Omega} \max_{q\in\Theta^v} \eta_\alpha^v(q)$ under the "worse" choice of q. Notice that each q_{ij} does not appear in unique row of $A^v(q)$, so that the rows of $A^v(q)$ are dependent. Therefore, there is usually none solution to the min-max problem. Now, redefine a weighted-average discounted-cost as $\bar{\eta}_\alpha^v(q) = \omega\eta_\alpha^v(q)$, where $\omega = (\omega_1, \omega_2, \cdots, \omega_K)$ with $\omega_i \in [0,1]$ and $\sum_i \omega_i = 1$. Then, the optimal robust control problem is to find a policy

$$v^* \in \arg\min_{v\in\Omega_s}\max_{q\in\Theta^v} \bar{\eta}_\alpha^v(q) \tag{2}$$

Here, ω may be a constant vector, or dependent on v, q and α. Suppose ω is continuously differentiable with respect to (q, α) on $\Theta^v \times [0, \infty)$ for any v.

3 Potential-based solution of robust control policy

Suppose the stochastic process is irreducible for any $v \in \Omega_s$ and $q \in \Theta^v$. The steady-distribution $\pi^v(q)$ satisfies $\pi^v(q)e = 1, A^v(q)e = 0, \pi^v(q) = 0$, where e is an all-one vector. Then, average-cost $\eta^v(q) = \pi^v(q)f^v$. For any $\alpha \geqslant 0$, suppose the performance potential vector $g_\alpha^v(q)$ is the unique solution to Poisson equation $(\alpha I - A^v(q) + \lambda^v e\pi^v(q))g_\alpha^v(q) = f^v$. Let $\eta_0^v(q) = \lim_{\alpha\to0^+}\eta_a^v(q), \bar{\eta}_0^v(q) = \lim_{\alpha\to0^+}\bar{\eta}_\alpha^v(q)$; then we have the following lemma.

Lemma 1. a) $\eta_0^v(q) = e\eta^v(q)$, $\bar{\eta}_0^v(q) = \eta^v(q)$;

b) For any $\alpha > 0$, $\pi^v(q)\alpha(\alpha I - A^v(q))^{-1} = \pi^v(q)$;

c) $\lim_{\alpha\to0^+}\alpha(\alpha I - A^v(q))^{-1} = e\pi^v(q)$;

d) If $\omega = \pi^v(q)$, then for any $\alpha \geqslant 0$, $\bar{\eta}_\alpha^v(q) = \bar{\eta}_0^v(q) = \eta^v(q)$.

Proof. a) First, it is easy to prove the following equation[6].

$$(\alpha I - A^v(q) + \lambda^v e\pi^v(q))^{-1} = (\alpha I - A^v(q))^{-1} - \lambda^v e\pi^v(q)/[\alpha(\lambda^v + \alpha)] \tag{3}$$

Right-multiplying both sides of the above equation by αf^v yields

$$\eta_\alpha^v(q) = \alpha g_\alpha^v(q) + \lambda^v e\eta^v(q)/(\lambda^v + \alpha) \tag{4}$$

which implies $\lim_{\alpha\to0^+}\eta_\alpha^v(q) = e\eta^v(q)$. Obviously, $\lim_{\alpha\to0^+}\omega e = 1$, then

$$\bar{\eta}_0^v(q) = \lim_{\alpha\to0^+}\omega\eta_\alpha^v(q) = \lim_{\alpha\to0^+}\omega \lim_{\alpha\to0^+}\eta_\alpha^v(q) = \lim_{\alpha\to0^+}\omega e\eta^v(q) = \eta^v(q)$$

b) Since $\pi^v(q)A^v(q) = 0$, $\pi^v(q)(\alpha I - A^v(q)) = \alpha\pi^v(q)$. Right-multiplying both sides of this equation by $(\alpha I - A^v(q))^{-1}$ leads to the desired result.

c) The similar result has appeared in [5]. In fact, from (3), we directly have

$$\lim_{\alpha\to0^+}\alpha(\alpha I - A^v(q))^{-1} = \lim_{\alpha\to0^+}\{\alpha(\alpha I - A^v(q) + \lambda^v e\pi^v(q))^{-1} + \lambda^v e\pi^v(q)/(\lambda^v + \alpha)\} = e\pi^v(q)$$

d) If $\alpha > 0$, from (b) we obtain

$$\bar{\eta}_\alpha^v(q) = \pi^v(q)\eta_\alpha^v(q) = \pi^v(q)\alpha(\alpha I - A^v(q))^{-1}f^v = \pi^v(q)f^v = \eta^v(q)$$

which, combined with (a) as $a = 0$, implies the desired results for any $\alpha \geqslant 0$. □

Lemma 1 a) or Lemma 1 d) show that the uncertain average-cost problem is a special form of the weighted-average discounted-cost problem. If $\omega = e_i$, where e_i is a unit coordinate vector with the i-th element equal to 1, then $\bar{\eta}_\alpha^v(q) = e_i \eta_\alpha^v(q) = \eta_\alpha^v(i, q)$, and problem (2) becomes to find the optimal robust policy for discounted-cost $\eta_\alpha^v(i, q)$.

The solution of (2) may be divided into two steps. First, for any given policy v, select the worst routing matrix q^v satisfying $q^v \in \arg\max_{q \in \Theta^v} \bar{\eta}_\alpha^v(q)$. Second, choose a policy v^* satisfying $v^* \in \arg\min_{v \in \Omega_s} \bar{\eta}_\alpha^v(q^v)$.

Theorem 1. For any $v \in \Omega_s$ and $\alpha > 0$, we have the following gradient formula

$$\nabla \bar{\eta}_\alpha^v(q) = \nabla \omega \alpha g_\alpha^v(q) + \omega \alpha (\alpha I - A^v(q))^{-1} \nabla A^v(q) g_\alpha^v(q) \tag{5}$$

Here, "∇" denotes taking gradient with respect to q.

Proof. From (1), $(\alpha I - A^v(q)) \eta_\alpha^v(q) = \alpha f^v$. So $(\alpha I - A^v(q)) \nabla \eta_\alpha^v(q) - \nabla A^v(q) \eta_\alpha^v(q) = 0$. $A^v(q)e = 0$ leads to $\nabla A^v(q)e = 0$. Then, from (4),

$$(\alpha I - A^v(q)) \nabla \eta_\alpha^v(q) = \nabla A^v(q) \eta_\alpha^v(q) = \alpha \nabla A^v(q) g_\alpha^v(q)$$

that is, $\nabla \eta_\alpha^v(q) = \alpha(\alpha I - A^v(q))^{-1} \nabla A^v(q) g_\alpha^v(q)$. $\omega e = 1$ implies $\nabla \omega e = 0$, which together with (4) leads to $\nabla \omega \eta_\alpha^v(q) = \nabla \omega \alpha g_\alpha^v(q)$. Therefore,

$$\nabla \bar{\eta}_\alpha^v(q) = \nabla \omega \eta_\alpha^v(q) + \omega \nabla \eta_\alpha^v(q) = \nabla \omega \alpha g_\alpha^v(q) + \omega \alpha (\alpha I - A^v(q))^{-1} \nabla A^v(q) g^v(q)$$

and the desired result is obtained. □

Now denote $g_0^v(q)$ to be $g^v(q)$. Then we have the following corollary.

Corollary 1. a) If $\omega = \pi^v(q)$, $\alpha > 0$, then $\nabla \bar{\eta}_\alpha^v(q) = \nabla \eta^v(q) = \pi^v(q) \nabla A^v(q) g^v(q)$.

b) $\lim_{\alpha \to 0^+} \nabla \bar{\eta}_\alpha^v(q) = \nabla \bar{\eta}_0^v(q) = \nabla \eta^v(q)$

Proof. a) Lemma 1 d) implies $\nabla \bar{\eta}_\alpha^v(q) = \nabla \eta^v(q)$. From (5) and Lemma 1 b),

$$\nabla \bar{\eta}_\alpha^v(q) = \nabla \pi^v(q) \alpha g_\alpha^v(q) + \pi^v(q) \nabla A^v(q) g_\alpha^v(q) \tag{6}$$

From the Poisson equation, $\nabla \pi^v(q)(\alpha I - A^v(q) + \lambda^v e \pi^v(q)) g_\alpha^v(q) = \nabla \pi^v(q) f^v$, and $\nabla \pi^v(q)(-A^v(q) + \lambda^v e \pi^v(q)) g^v(q) f^v$. It is easy to know $\nabla \pi^v(q) e = 0$, thus $\nabla \pi^v(q)(\alpha I - A^v(q)) g_\alpha^v(q) = -\nabla \pi^v(q) A^v(q) g^v(q)$, that is,

$$\nabla \pi^v(q) \alpha g_\alpha^v(q) = \nabla \pi^v(q) A^v(q) g_\alpha^v(q) - \nabla \pi^v(q) A^v(q) g^v(q)$$

From $\pi^v(q) A^v(q) = 0$, we obtain $\nabla \pi^v(q) A^v(q) + \pi^v(q) \nabla A^v(q) = 0$. Thus

$$\nabla \pi^v(q) \alpha g_\alpha^v(q) = -\pi^v(q) \nabla A^v(q) g_\alpha^v(q) + \pi^v(q) \nabla A^v(q) g^v(q)$$

Substituting the above equation into (6), we obtain $\nabla \bar{\eta}_\alpha^v(q) = \pi^v(q) \nabla A^v(q) g^v(q)$, which is similar to the gradient formula of average-cost η^v with respect to v[8].

b) Lemma 1 a) implies $\nabla \bar{\eta}_0^v(q) = \nabla \eta^v(q)$. Since $\nabla \omega$ is bounded and $\lim_{\alpha \to 0^+} g_\alpha^v(q) = g_0^v(q) = g^v(q)$, from (5) and Lemma 1 (c), we have

$$\lim_{\alpha \to 0^+} \nabla \bar{\eta}_\alpha^v(q) = \omega e \pi^v(q) \nabla A^v(q) g_0^v(q) = \pi^v(q) \nabla A^v(q) g^v(q) = \nabla \eta^v(q) = \nabla \bar{\eta}_0^v(q)$$

and derive the desired result. □

Theorem 2. For any $\alpha > 0$, if ω is a constant vector independent of v and q, then for any v', $v \in \Omega_s$ and $q' \in \Theta^{v'}, q \in \Theta^v$

$$\bar{\eta}_\alpha^{v'}(q') - \bar{\eta}_\alpha^v(q) = \omega \alpha (\alpha I - A^v(q))^{-1} [(f^{v'} + A^{v'}(q') g_\alpha^{v'}(q')) - (f^v + A^v(q) g_\alpha^{v'}(q'))] \tag{7}$$

Proof. From (1), $\alpha \eta_\alpha^v(q) = \alpha f^v + A^v(q) \eta_\alpha^v(q)$. Then

$$(\alpha I - A^v(q))(\eta_\alpha^{v'}(q') - \eta_\alpha^v(q)) = (\alpha f^{v'} + \alpha A^{v'}(q') \eta_\alpha^{v'}(q')) - (\alpha f^v + A^v(q) \eta_\alpha^{v'}(q'))$$

which together with (4) yields

$$(\alpha I - A^v(q))(\eta_\alpha^{v'}(q') - \eta_\alpha^v(q)) = (\alpha f^{v'} + \alpha A^{v'}(q')g_\alpha^{v'}(q')) - (\alpha f^v + \alpha A^v(q)g_\alpha^{v'}(q'))$$

Then

$$\eta_\alpha^{v'}(q') - \eta_\alpha^v(q) = \alpha(\alpha I - A^v(q))^{-1}[(f^{v'} + A^{v'}(q')g_\alpha^{v'}(q')) - (f^v + A^v(q)g_\alpha^{v'}(q'))]$$

Since ω is a constant vector, we have $\bar{\eta}_\alpha^{v'}(q') - \bar{\eta}_\alpha^v(q) = \omega(\eta_\alpha^{v'}(q') - \eta_\alpha^v(q))$, which together with the foregoing equation leads to the desired result. □

From Theorem 1, we can look for the "worse" routing probabilities for a given policy by standard gradient method or other gradient-like approaches such as Newton method and Quasi-Newton method. Using (7), we may compare the performance of arbitrary two policies to search the optimal robust control policy. If $\mathrm{argmin}_{v\in\Omega_s}\{f^v + A^v(q^v)g_\alpha^{\tilde{v}}(q^{\tilde{v}})\}$ is nonempty for any $\tilde{v}$, potential-based policy iteration and value iteration may be applied[7]. In addition, some global optimization techniques such as simulation annealing and evolutionary algorithms may be necessary and effective in searching q^v and the robust control policy v^*.

For ergodic Markov processes, every element of $\alpha(\alpha I - A^v(q))^{-1}$ is positive. Then, we have the following theorem.

Theorem 3. If ω is a constant vector, a policy v^* is an optimal robust control policy if

$$f^{v^*} + A^{v^*}(q^{v^*})g_\alpha^{v^*}(q^{v^*}) \leqslant f^v + A^v(q^v)g_\alpha^{v^*}(q^{v^*}), \quad \forall v \in \Omega_s \tag{8}$$

In fact, (8) is a sufficient and necessary condition for a policy to be optimal in deterministic case, and is similar to the optimality condition in Theorem 2 of [6].

4 A numerical example

Consider an uncertain exponential CCQN with $M = 3$, $N = 4$, $q_{ij} \in [1/2M, 3/2M]$, $\mu_{i,n} \in [0.5, 1.2], n_i \neq 0; \mu_{i,n} = 0, n_i = 0$ and $f(n, v(n)) = \sum_{i=1}^{M} f_i(n, v(n))$, where

$$f_i(n, v(n)) = \ln(1 + n_i/N) \cdot \mu_{i,n} + \sqrt{n_i}/2N\mu_{i,n}, n_i \neq 0; f_i(n, v(n)) = 0, n_i = 0$$

Select $\lambda^v = 3.6$ and $\alpha = 0.9$. By Quasi-Newton method, we have the following results for $\omega = e_1, e/K, \pi^v(q)$ shown in Table 1.

Table 1 Computation results corresponding to different weight vectors

ω	q^v	$\bar{\eta}_\alpha^v(q^v)$
$\omega = e_1$	[0.16667, 0.37491, 0.45843; 0.16667, 0.5, 0.33333; 0.16667, 0.33333, 0.5]	0.90204
$\omega = e/K$	[0.17758, 0.42938, 0.39305; 0.21626, 0.5, 0.28374; 0.37681, 0.29402, 0.32917]	1.00563
$\omega = \pi^v(q)$	[0.39533, 0.40281, 0.20186; 0.49148, 0.21014, 0.29838; 0.16667, 0.45247, 0.38087]	0.97812

In our simulation, we noticed that with the increasing of the customers or servers, the optimization would be more and more time-consuming. Letting $\omega = \pi^v(q)$, the computation time is 58 seconds if $M = 2$, $N = 3$, and is 57 minutes for $M = 3$, $N = 3$, but it will reach 6.7 hours if $M = 3$, $N = 4$. Thus, for large-scale systems, the computation-based method may be infeasible. There are some ways to overcome the computation complexity such as potential-based parallel computing and simulation methods. Especially, potential-based reinforcement learning and neuro-dynamic programming may be effective in practice.

5 Conclusions

We see that the study of both average– and discounted-cost robust control problems for uncertain CCQNs can be unified by using the concept of potential. The obtained results may be extended to other queuing networks such as phase-type queuing networks, and to semi-Markov decision processes (SMDPs) with uncertain parameters through the potential theory of semi-Markov processes[5].

Acknowledgement. The first author would like to thank Suresh Kalyanasundaram for his presentation of the extended version of his CDC paper.

References

1 Cao X R, Chen H F. Perturbation realization, potentials and sensitivity analysis of Markov processes. *IEEE Transactions on Automatic Control*, 1997, **42**(10): 1382∼1393

2 Zhou Y P, Yin B Q, Xi H S, Tang H. An optimality algorithm for a class of closed queuing network by using Oblivion factor. *Journal of System Simulation*, 2003, **15**(3): 386∼392

3 Kalyanasundaram S, Chong E K P, Shroff N B. Markov decision processes with uncertain transition rates: Sensitivity and robust control. In: Proceedings of the 41st IEEE Conference on Decision and Control, Las Vegas, USA: Nevada, 2002. **4**: 3799∼3804

4 Tang H, Han J H, Gao J. Optimal robust control policy for continuous-time Markov control processes with average-cost criteria. *Journal of University of Science and Technology of China*, 2003, **34**(2): 219∼225

5 Cao X R. Semi-Markov decision problems and performance sensitivity analysis. *Transactions on Automatic Control*, 2003, **48**(5): 758∼769

6 Xi H S, Tang H, Yin B Q. Optimal policies for a continuous time MCP with compact action set. *Acta Automatica Sinica*, 2003, **29**(2): 206∼211

7 Tang H, Xi H S, Yin B Q. Iteration optimization algorithms for Markov control processes with compact action set. *Control and Decision*, 2003, **18**(3): 267∼271

8 Tang H, Xi H S, Yin B Q. A simulation optimization algorithm for CTMDPs based on randomized stationary policies. *Acta Automatica Sinica*, 2004, **30**(2): 225∼229

TANG Hao Received his master degree in nuclear-energy science and engineering from Institute of Plasma Physics, Chinese Academy of Sciences in 1998, and Ph.D degree in pattern recognition and intelligent system from University of Science and Technology of China (USTC) in 2002. He is currently an associate professor of Hefei University of Technology. His research interests include DEDS (MDP, SMDP, and queuning networks), reinforcement learning, and the methodology of neuro-dynamic programming.

XI Hong-Sheng Received his bchelor and master degrees in mathematics from USTC in 1977 and 1985, respectively. He is currently a professor of USTC. His research interests include DEDS and hybrid systems.

HAN Jiang-Hong Received his master degree in computer application from Hefei University of Technology (HFUT) in 1987. He is currently a professor of HFUT, and his research interests include intelligent control, embedded systems, and computer information systems.

YUAN Ji-Bin Master student at Hefei University of Technology. His research interests include computer control technology and artificial intelligence.