About Service Intensity Bounds for a Queuing Model

I.Kovalev[1], Y.Satin[1] and A.Zeifman[2,a)]

[1] *Vologda State University*
[2] *Vologda State University, Institute of Informatics Problems of the Federal Research Center "Informatics and Control", RAS; Vologda Research Center, RAS*

[a)] Corresponding author: a_zeifman@mail.ru

Abstract. We estimate the server capacity, at which the average number of customers in the system does not exceed a given number. Using perturbation bounds, we estimate the boundaries of the service intensity so that the limiting mean of the length of the queue remains within the specified boundaries.

INTRODUCTION

The initial description of the model and the first studies one can find in [1]. Ergodicity bounds for corresponding non-stationary model have been obtained in [3].

Let $X(t)$ be an inhomogeneous continuous-time Markov chain describing the evolution of the number of customers in the system with a countable state space $\mathcal{X} = \{0, 1, 2, 3 \ldots\}$, and the correspionding intensity matrix has the form

$$Q(t) = \begin{pmatrix} -\lambda(t) & \lambda(t)b_1 & \lambda(t)b_2 & \cdots \\ \mu(t) & -(\mu(t) + \lambda(t)B_2) & \lambda(t)b_2 & \cdots \\ 0 & \mu(t) & -(\mu(t) + \lambda(t)B_3) & \cdots \\ 0 & 0 & \mu(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where $\lambda(t)$ is general arrival intensity. $\mu(t)$ is the service intensity. $\lambda(t)b_k$ is the arrival intensity of a group of customers is such that the total number of them in the system is equal k. If the system already has $k - j$ of customers, then this is the intensity of the simultaneous receipt of a group of j of customers. At the same time $B_k = \sum_{n=k}^{\infty} b_n$ for all $k \geq 1$ and $B_1 = 1$. All functions that determine the intensities are assumed to be non-negative and locally integrable on $[0, \infty)$. In addition, it is assumed that the following conditions are met $b_k \leq Cq^k$, where $C > 0$ and $0 < q < 1$.

Applying our standard approach we assume that $\sup_i |q_{ii}(t)| \leq L < \infty$, for almost all $t \geq 0$.

Denote by $\|\cdot\|$ the l_1−norm of the vector, $\|x\| = \sum |x_i|$, $\|B\| = \sup_j \sum_i |b_{ij}|$, if $B = (b_{ij})_{i,j=0}^{\infty}$, and we denote by Ω the set of all vectors from l_1 with non-negative coordinates and unit norm.

THEORETICAL BOUNDS

Let $\mathbf{p}(t) = (p_0(t), p_1(t), \ldots)^T$ be the probability distribution vector at time t. Given any proper initial condition $\mathbf{p}(0)$, the probabilistic dynamics of the Markov chain $X(t)$ is described by the forward Kolmogorov system of differential equations

$$\frac{d}{dt}\mathbf{p}(t) = A(t)\mathbf{p}(t), \tag{1}$$

where $A(t) = Q^T(t)$ is the transposed intensity matrix.

Using the normalization condition $p_0(t) = 1 - \sum_{i \geq 1, i \in \mathcal{X}} p_i(t)$ it can be checked that the system (1) can be rewritten as follows:

$$\frac{d}{dt}\mathbf{z}(t) = B(t)\mathbf{z}(t) + \mathbf{f}(t), \tag{2}$$

where $B(t) = \left(b_{ij}(t)\right)_{i,j=1}^{\infty}$, $b_{ij}(t) = a_{ij}(t) - a_{i0}(t)$, $\mathbf{f}(t) = (\lambda(t)b_1, \lambda(t)b_2, \dots)^T$ and $\mathbf{z}(t) = (p_1(t), p_2(t), \dots)^T$.
Note that the matrix $B(t)$ has no probabilistic meaning. Let $\mathbf{z}^*(t)$ and $\mathbf{z}^{**}(t)$ be the solutions of (2) corresponding to (different) initial conditions $\mathbf{z}^*(0)$ and $\mathbf{z}^{**}(0)$. Then for the vector $\mathbf{y}(t) = \mathbf{z}^*(t) - \mathbf{z}^{**}(t) = (y_1(t), y_2(t), \dots)^T$, which has coordinates of arbitrary signs, we have

$$\frac{d}{dt}\mathbf{y}(t) = B(t)\mathbf{y}(t). \tag{3}$$

It is more convenient to study the rate of convergence using the transformed version $B^*(t)$ of $B(t)$ given by $B^*(t) = DB(t)D^{-1}$, where D is the upper triangular matrix of the form

$$D = \begin{pmatrix} d_1 & d_1 & d_1 & \cdots \\ 0 & d_2 & d_2 & \cdots \\ 0 & 0 & d_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

where $d_i,\ i \geq 1$ be a sequence of positive numbers such that $1 = d_1 \leq d_2 \leq \dots$. Let $\delta < 1$ be a positive number, and $d_{k+1} = \delta^{-k}$, $k \geq 1$.

Then

$$\|B(t)\|_{1D} = \|DB(t)D\| = \|B^*(t)\| \leq \lambda(t) + (1 + \delta)\mu(t) \tag{4}$$

and

$$\|\mathbf{f}^{**}(t)\| = \|DT\mathbf{f}(t)\| \leq \frac{C|\lambda(t)|\delta q}{(\delta - q)(1 - q)} \leq C^* L, \tag{5}$$

for $\delta > q$.

Then one can write the following bounds

$$\|\mathbf{y}\,(t)\,\|_{1D} \leq e^{-\int_0^t (1-\delta)\mu(u)\,du}\|\mathbf{y}(0)\|_{1D},$$

for any $\delta \in (q, 1)$.

The Markov chain $X(t)$ has a limiting mean, say $\phi(t)$, and the following rate of convergence bounds hold:

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \leq 4e^{-\int_0^t (1-\delta)\mu(u)\,du}\|\mathbf{y}(0)\|_{1D}, \tag{6}$$

$$|E(t,k) - \phi(t)| \leq \frac{4}{W}e^{-\int_0^t (1-\delta)\mu(u)\,du}\|\mathbf{y}(0)\|_{1D}, \tag{7}$$

where $W = \inf_{k \geq 1} \frac{d_k}{k} > 0$ and $E(t,k) = \sum_{n \in \mathcal{X}} np_n(t)$ is the mean number of customers in the system at time t, given that initially there where k customers in the system i.e. $p_k(0) = 1$.

Firstly we consider the situation with known upper bound of the limiting mathematical expectattion of the $X(t)$. Assume that the service intensity is given by the expression $\mu(t) = \mu g(t)$, where $g(t)$ is known, $e^{-\int_\tau^t g(u)du} \leq He^{-\nu(t-\tau)}$, and we can control μ (server power). Consider the estimating the possible power of the server μ, assuming that the other parameters are set, and the limiting mean does not exceed N^*, i.e. $\limsup_{t\to\infty} E(t,0) \leq N^*$.

We have $E(t,0) = \sum_{n \in \mathcal{X}} np_n(t)$. Since all $p_i(t)$ are non-negative, we have

$$\|\mathbf{z}\|_{1D} = \sum_{n \geq 1} p_n \sum_{k \geq 1} d_k \geq \sum_{n \geq 1} d_n p_n \geq W \sum_{n \geq 1} np_n, \tag{8}$$

where $W = \inf_{k \geq 1} \frac{d_k}{k} > 0$. Hence $E(t,0) \leq \frac{1}{W}\|\mathbf{z}(t)\|_{1D}$.

On the other hand, we get the following inequality:

$$\|\mathbf{z}\|_{1D} \leq \|V(t)\|_{1D}\|\mathbf{z}(0)\|_{1D} + \int_0^t \|V(t,\tau)\|_{1D}\|\mathbf{f}(\tau)\|_{1D}d\tau \leq$$

$$\leq 4e^{-(1-\delta)\mu\int_0^t g(u)du}\|\mathbf{z}(0)\|_{1D} + 4C^*L\int_0^t e^{-(1-\delta)\mu\int_\tau^t g(u)du}d\tau \leq$$

$$\leq 4\Big(He^{-\nu t}\Big)^{(1-\delta)\mu}\|\mathbf{z}(0)\|_{1D} + 4C^*L\int_0^t \Big(He^{-\nu(t-\tau)}\Big)^{(1-\delta)\mu}d\tau \leq \frac{4C^*LH^{(1-\delta)\mu}}{\nu(1-\delta)\mu}$$

Hence $\limsup_{t\to\infty} E(t,0) \leq \frac{4C^*LH^{(1-\delta)\mu}}{W\nu(1-\delta)\mu} \leq N^*$ and we obtain the following statement

Theorem 1. Let service intensity $\mu(t) = \mu g(t)$, and let $\lambda(t)$, $g(t)$, and $\{b_k\}$ be known.Then for the inequality

$$\limsup_{t\to\infty} E(t,0) \leq N^* \tag{9}$$

to be valid, it suffices to

$$\frac{H^{(1-\delta)\mu}}{\mu} \leq \frac{\nu(1-\delta)N^*W}{4C^*L}. \tag{10}$$

Consider now possible perturbations for a service intensity function.

Let $\bar{X}(t), t \geq 0$ be a "perturbed" process with an infinitesimal matrix $\bar{Q}(t)$ and the corresponding transposed matrix $\bar{A}(t)$, where the perturbation matrix $\hat{A}(t) = A(t) - \bar{A}(t)$ is small in a sense.

We consider stability estimates of the main characteristics of the $X(t)$ process under such perturbations, assuming additionally that the $X(t)$ process is exponentially ergodic, that is, that for some positive c, b and all s, t, $0 \leq s \leq t$ the inequality is satisfied

$$4e^{-\int_0^t (1-\delta)\mu(u)\,du} \leq Me^{-a(t-s)}. \tag{11}$$

Let the perturbed process $\bar{X}(t)$ is such that the corresponding intensity matrix $\bar{Q}(t)$ has the same structure.

Let the perturbed intensities be $\bar{\lambda}(t), \bar{\mu}(t)$ such that $\lambda(t) = \bar{\lambda}(t)$, $|\mu(t)-\bar{\mu}(t)| \leq \hat{\epsilon}_\mu$ for almost all $t \geq 0$. Then, applying the corresponding results of [2] we obtain the following statement.

Theorem 2. Let (11) hold. Then $X(t)$ and $\bar{X}(t)$ are $1D$-exponentially weakly ergodic, and the following perturbation bounds hold:

$$\lim_{t\to\infty}\sup \|p(t) - \bar{p}(t)\|_{1D} \leq \frac{M^2\hat{\epsilon}_\mu(1+\delta)C^*L}{a(a - M\hat{\epsilon}_\mu(1+\delta))}, \tag{12}$$

(for any non-decresing $\{d_k\}$),

$$\lim_{t\to\infty}\sup |\phi(t) - \bar{\phi}(t)| \leq \frac{M^2\hat{\epsilon}_\mu(1+\delta)C^*L}{Wa(a - M\hat{\epsilon}_\mu(1+\delta))}, \tag{13}$$

(if $W = \inf_{k\geq 1}\frac{\delta^{-k}}{k} > 0$).

Corollary 1. Let

$$\limsup_{t\to\infty} |E(t,0) - \bar{E}(t,0)| \leq h, \tag{14}$$

for some positive h. Then the service intensity can be estimated as follows:

$$|\mu(t) - \bar{\mu}(t)| \leq \frac{Wa^2h}{(1+\delta)(C^*LM^2 + MhWa)}. \tag{15}$$

EXAMPLE

Consider example 1 from [3].

Let both the arrival and service intensities be periodic and equal to $\lambda(t) = 1 + \sin(2\pi t)$ and $\mu(t) = 1 + \cos(2\pi t)$. Let $q = \frac{2}{3}$, i.e. the batch size distribution be $b_k = \frac{2^{k-1}}{3^k}$, $k \geq 1$, i.e. the mean batch size $\sum_{k=1}^{\infty} k b_k$ is 3. Put $\delta = \frac{5}{6}$ and it follows that $e^{-\int_s^t \alpha^*(u)\,du} = e^{-\frac{1}{6}\int_s^t (1+\cos(2\pi u))\,du} \leq 2e^{-\frac{1}{6}(t-s)}$. We can put $a = \frac{1}{6}$, $M = 2$, $L = 2$, $C = \frac{1}{3}$ and $C^* = \frac{10}{3}$. Thus $W = \inf_{k\geq 1} \frac{d_k}{k} = \frac{3}{4}$.

Inequality $E(t,0) \leq \frac{4C^* L H^{(1-\delta)\mu}}{W\nu(1-\delta)\mu}$ ensures that the average number of requirements in the system does not exceed 1437. It follows from Theorem 1 that in order for the average number of requirements in the system not to exceed, for example, 500, it is enough to increase the power of the server by 5 times.

Let $|E(t,0) - \bar{E}(t,0)| \leq h = 3$. Then, from the corollary of Theorem 2, the bounds for the service intensity will have the form $|\mu(t) - \bar{\mu}(t)| \leq \hat{\epsilon}_\mu \leq \frac{3}{4(320+3h)} \leq 10^{-2}$.

Figure 1 shows the behavior of the average value of the mean number of customers in the system when $\mu(t) = 0.99(1 + \cos(2\pi t))$:

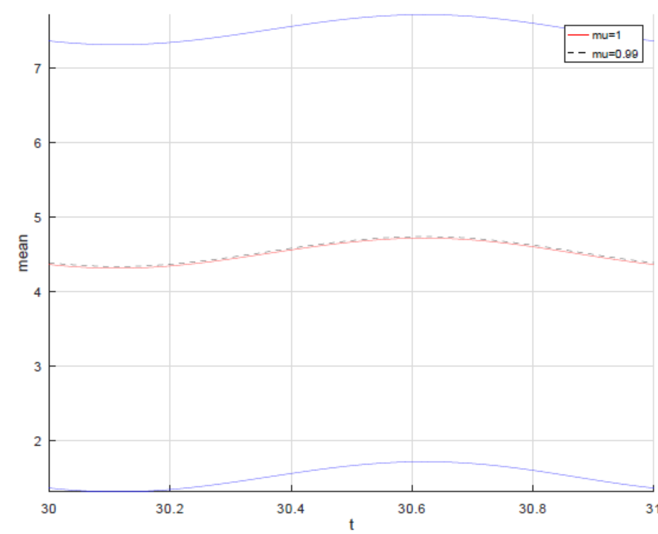

FIGURE 1. The limiting mean $E(t;0)$ with bounds for $\mu = 1$ and $\mu = 0.99$ for $t \in [30,31]$.

ACKNOWLEDGMENTS

This work was supported by Russian Science Foundation under grant 19-11-00020.

REFERENCES

[1] Marin, A. and S. Rossi, A Queueing Model that Works Only on the Biggest Jobs. *Lecture Notes in Computer Science book series (LNCS)* 12039: 118-132 (2020).

[2] Zeifman, A., Korolev, V., Satin, Y., Two approaches to the construction of perturbation bounds for continuous-time Markov chains. *Mathematics*, **8(2)**, 253 (2020).

[3] Zeifman, A. I., R. V. Razumchik, Y. A. Satin, and I. A. Kovalev. Ergodicity bounds for the Markovian queue with time-varying transition intensities, batch arrivals and one queue skipping policy. *Applied Mathematics and Computation* 395:125846 (2021).