Control and Cybernetics

vol. **22** (1993) No. 1/2

Cost characteristics of system service with input controlled by Markov chain

by

Marian Chudy

Military University of Technology
Warsaw
Poland

The paper presents the definitions and calculations related to the limit cost characteristic for system consisting of N computers serving computational tasks. The tasks arrive at random intervals of time. These intervals have the same probability distribution. The calculation times are also random with the same probability distribution for separete computers.

Markov chain is the mathematical model of the process of allocation of computational tasks to computers.

The total cost of service per unit time and an estimation of the total cost of waiting for the computational tasks per unit are presented.

These characteristics can be used to formulate and solve some optimization problems.

1. Preface

We consider a system consisting of N computers which serve computational tasks. A common input stream of computational tasks is divided into N parts by some control element (CE). A computational task allocated by CE to n–th computer takes place in a FIFO queue for this computer.

Markov chain (Z_i) is the mathematical model of (CE), in which:
Z_i is the random variable being the index of the computer to which i–th computational task was allocated.

Let us introduce the following notations:

t_i – random variable denoting the instance of arrival of i–th computational task at the CE,

$t_{i+1} - t_i = x_i$ – interarrival time between $(i+1)$–st and i–th tasks.

We asssume that:

1. $\{x_i\}$ is the squence of independent random variables with the same distribution and finite time a.

Other notations:

$N(t)$ – total number of tasks that arrive at CE until t,

t_k^n – random variable corresponding to the instant of arrival of the k–th computational task at the n–th computer,

$t_{k+1}^n - t_k^n = x_k^n$ – interarrival times between $(k+1)$–st and k–th tasks.

We assume also that:

2. The Markov chain $\{Z_i\}$ is homogeneous.

We will now try to develop the expression describing x_k^n.

We denote by

i_k^n – the index of the time instant t_i at which Markov chain $\{Z_i\}$ returns for the k–th time to the state n.

Then

$i_{k+1}^n - i_k^n = y_{k+1}^n$ is the number of the time steps between two consecutive instants when the Markov chain is in state n.

Taking Assumption 2 and the initial condition $Z_1 = n$ we obtain that the sequence of random variables $\{y_n^k\}$ is the sequence of independent random variables with the same distribution, having all moments (Borovkov, 1972).

We can present random variable x_k^n in the form

$$x_k^n = \sum_{i=i_k^n+1}^{i_{k+1}^n} x_i \tag{1}$$

This is the sum of y_{k+1}^n independent random variables with the same distribution. Hence, the sequence $\{x_k^n\}$ is the sequence of independent random variables with the same distribution. The mathematical expectation of x_k^n is

$$Ex_k^n = E\sum_{i=i_k^n+1}^{i_{k+1}^n} x_i = aEy_{k+1}^n = ah^n \tag{2}$$

and equals

$$Ex_k^n = \frac{a}{p_n} \tag{3}$$

when the Markov chain is, additionally, aperiodic, and where

$$h^n = Ey_{k+1}^n,$$

p_n – limit probability for the Markov chain $\{Z_i\}$ of being in the state n.

Denote by

η_k^n – random variable for the value of service time of k–th task with n–th computer.

We assume that
3. $\{\eta_k^n\}$ – the sequence of independent random variables with the same distribution and finite mean time b^n, and that we know the following quantities:
C_o^n – cost of unit time of computation performed by n–th computer,
C_w^n – cost of unit time of waiting for computation at n-th computer.

2. Cost characteristics of the system

The quantities introduced give us the possibility of defining the following characteristics:
– total payment for computation per unit time

$$\lim_{t\to\infty} \frac{K_o(t,N)}{t} = \lim_{t\to\infty} \frac{1}{t} \sum_{n=1}^{N} \sum_{k=1}^{N^n(t)} C_o^n \eta_k^n \tag{4}$$

– total cost of waiting per unit time

$$\lim_{t\to\infty} \frac{K_w(t,N)}{t} = \lim_{t\to\infty} \frac{1}{t} \sum_{n=1}^{N} \sum_{k=1}^{N^n(t)} C_w^n W_k^n \tag{5}$$

where
$N^n(t)$ – number of tasks having arrived at n–th computer
W_k^n – waiting time of the k–th task at the n–th computer.

By Assumption 1 and the basic theorem of renewal theory we obtain, with probability one,

$$\lim_{t\to\infty} \frac{N^n(t)}{t} \stackrel{1}{=} \frac{1}{ah^n} \tag{6}$$

or

$$\lim_{t\to\infty} \frac{N^n(t)}{t} \stackrel{1}{=} \frac{p^n}{a} \tag{7}$$

when the Markov chain is aperiodic.

By Assumption 3 and the strong law of large numbers (Doob, 1953) we obtain with probability one

$$\lim_{t\to\infty} \frac{C_o^n}{N^n(t)} \sum_{k=1}^{N^n(t)} \eta_k^n \stackrel{1}{=} C_o^n b^n \tag{8}$$

Using (8), (10) and the theorem related to multiplication of the limits of stochastic processes we can write

$$\lim_{t\to\infty} \frac{K_0(t,N)}{t} \stackrel{1}{=} \sum_{n=1}^{N} \frac{C_0^n b^n}{ah^n} \tag{9}$$

or, in the case of aperiodic chain,

$$\lim_{t\to\infty} \frac{K_0(t,N)}{t} \stackrel{1}{=} \sum_{n=1}^{N} \frac{p_n C_0^n b^n}{a} \tag{10}$$

Now we fix our attention on limit estimation for expression (5). The waiting time of $(k+1)$–st task on n–th computer can be expressed in the form

$$W_{k+1}^n = \max\{0, W_k^n + \eta_k^n - x_k^n\} = \max\{0, W_k^n + \xi_k^n\} \tag{11}$$

where

$$\xi_k^n = \eta_k^n - x_k^n.$$

Assuming that $W_1^n = 0$, $n \in N = \{1, 2, \dots, N\}$ one can show that

$$W_k^n = \max\{0, \xi_k^n, \xi_k^n + \xi_{k-1}^n, \dots, \xi_k^n + \dots + \xi_2^n\} \tag{12}$$

The sequence $\{W_k^n\}$ is not stationary, so that we can not apply ergodic theorem. It follows from Kolmogorov theorem (Borovkov, 1972) that the sequence $\{\xi_k^n\}$, $n \in N$ can be completed to the form

$$\{\xi_k^n\}, \quad -\infty < k < \infty, \quad n \in N. \tag{13}$$

Let us consider the sequence of random variables $\{\bar{W}_k^n\}$ where

$$\bar{W}_k^n = \{0, \xi_k^n, \xi_k^n + \xi_{k-1}^n, \xi_k^n + \xi_{k-1}^n + \xi_{k-2}^n, \dots\} \quad 1 \le k < \infty, \quad n \in N. \tag{14}$$

One can observe that for every elementary event the following condition is true

$$W_k^n \le \bar{W}_k^n, \quad k \ge 1, \quad n \in N. \tag{15}$$

This property is useful for our estimation.

As it is shown in (Borovkov, 1972a) the sequences $\{\bar{W}_k^n\}$, $n \in N$ are stationary and if $E\xi_k^n < 0$ then $\bar{W}_k^n$ is finite with probability one. Hence we can apply ergodic theorem (Loeve, 1955) to the sequence $\{\bar{W}_k^n\}$, $n \in N$ and obtain

$$\lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} \bar{W}_k^n \overset{1}{=} \bar{W}^n(n), \quad E\bar{W}^n(N) = E\bar{W}_1^n, \quad n \in N. \tag{16}$$

We emphasize dependence of $\bar{W}^n(N)$ on N as the number of states of the chain $\{Z_i\}$, which determines x_k^n.

If we present (5) in the form

$$\lim_{t \to \infty} \sum_{n=1}^{N} \frac{N^n(t)}{t} \frac{C_w^n}{N^n(t)} \sum_{k=1}^{N^n(t)} W_k^n$$

and use the basic theorem of renewal theory and the theorem on multiplication of limits of stochastic processes we can write, considering (5) and (15)

$$\begin{aligned} \lim_{t \to \infty} \frac{K_w(t, N)}{t} &= \lim_{t \to \infty} \sum_{n=1}^{N} \frac{N^n(t)}{t} \frac{C_w^n}{N^n(t)} \sum_{k=1}^{N^n(t)} W_k^n \le \\ &\le \lim_{t \to \infty} \sum_{n=1}^{N} \frac{N^n(t)}{t} \frac{C_w^n}{N^n(t)} \sum_{k=1}^{N^n(t)} \bar{W}_k^n \overset{1}{=} \sum_{n=1}^{N} \frac{C_w^n \bar{W}^n(N)}{ah^n} \end{aligned}$$

or

$$\lim_{t\to\infty} \frac{K_w(t,N)}{t} \leq \sum_{n=1}^{N} \frac{C_w^n \bar{W}^n(N) p_n}{a}$$

when the chain $\{Z_i\}$ is aperiodic.

Hence, the respective mathematical expectations can be estimated as

$$E \lim_{t\to\infty} \frac{K_w(t,N)}{t} \leq \sum_{n=1}^{N} \frac{C_w^n E\bar{W}_1^n}{ah^n} \tag{17}$$

or

$$E \lim_{t\to\infty} \frac{K_w(t,N)}{t} \leq \sum_{n=1}^{N} \frac{C_w^n E\bar{W}_1^n p_n}{a} \tag{18}$$

We know from Borovkov (1972a) that the distribution of random variables W_k^n, $n \in N$ with $k \to \infty$ tend to the distributions of random variables $\bar{W}_1^n$, $n \in N$ respectively.

Let us denote by $W^n(N)$, $n \in N$ random variables which satisfy the following conditions

$$P\{W^n(N) < x\} = \lim_{k\to\infty} P\{W_k^n < x\} \quad n \in N \tag{19}$$

Taking into account the above condition and remarks we can compute $E\bar{W}_1^n$ as

$$E\bar{W}_1^n = EW^n(N) \tag{20}$$

where it is possible to compute $EW^n(N)$ as the limit characteristic of the queueing system.

This characteristic can be expressed sometimes in analytical form. Thus, in particular, for homogeneous Poisson process $N(t)$ and for deterministic service time with $b = 1$, $EW^n(N)$ equals

$$EW^n(N) = \sum_{k=1}^{N-1} \frac{1}{S_k} + \frac{1}{2}\frac{a^2 - N^2 + N}{a(N-a)}$$

where S_k, $k = 1, \ldots, N-1$, $S_0 = 0$ are the roots of the equation

$$a^N e^{-S} - (a-S)^n = 0.$$

3. Remarks and Conclusions

The results obtained depend strongly on the kind of Markov chain $\{Z_i\}$ considered because it determines x_k^n and therefore also its Ex_k^n. Thus, it is possible to investigate the quality of allocation of computational tasks to computers depending up on the form of Markov chain. The estimate of profit per unit time, i.e.

$$\sum_{n=1}^{N}\left(\frac{C_0^n b^n - C_w^n E\bar{W}_1^n}{ah^n} - C^n\right) \tag{21}$$

in which C^n – performance cost of n-th computer per unit time, can be taken as objective function.

In case when computers are homogeneous i.e.

$$C_0^n = C_0, \quad b^n = b, \quad C_w^n = C_w, \quad n \in N$$

we obtain

$$\frac{C_0 b}{a}\sum_{n=1}^{N}\frac{1}{h^n} - \frac{C_w}{a}\sum_{n=1}^{N}\frac{E\bar{W}_1^n}{h^n} - NC \tag{22}$$

Let the matrix of transition probabilities be

$$P = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ . & . & . & & . \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix} \tag{23}$$

If we assume that the chain $\{Z_i\}$ is given by (23), then we can formulate the optimization problem for determination of the number of computers in the system i.e:

$$\max\left(\frac{C_0 b}{a} - \frac{C_w EW^n(N)}{a} - NC\right) \tag{24}$$

subject to

$$b - N \cdot a < 0 \tag{25}$$

$$N \geq 0, \quad N - \text{integer} \tag{26}$$

Objective function corresponding in this case profit per unit time. Constraint (25) follows from condition $E\xi^n < 0$.

Reference

BERTSIMAS, D., (1990) An Analytic Approach to a General Class G/G/S Queueing Systems. *Operations Research*, **38**, 139–155.

BOROVKOV, A., A., (1972) Lecture on Probability. Moscow (in Russian).

BOROVKOV, A., A., (1972A) Stochastic Process in Queueing Theory. Moscow (in Russian).

BURMAN, D., Y., SMITH, D., R., (1986) An asymptotic Analysis of a Queueing System with Markov-Modulated Arrivals. *Operations Research*, **34**, 105–119.

DOOB, J., L., (1953) Stochastic Process. John Wiley, New York.

KLEINROCK, L. (1975) Queueing Systems; Vol 1: Theory, John Wiley, New York.

LOEVE, M. (1955) Probability Theory. D. van Nostrand Company, New York.