REGENERATIVE SIMULATION OF
TIME-SHARING QUEUEING SYSTEM IN
RANDOM ENVIRONMENT

A.V. ZORINE
N.I. Lobachevsky State University of Nizhni Novgorod
Nizhni Novgorod, RUSSIA
e-mail: andrei.zorine@itmm.unn.ru

Abstract

Time-sharing queueing systems in random environments allow only a limited-
depth analytical study. In particular, the coinciding necessary and sufficient con-
ditions for the stationary probability distribution existence and an optimization
problem for mean total sojourn cost per working tact can be solved only under
much stricter assumptions about the time structure of input flows. In this talk a
regenerative method is discussed in its application to Monte-Carlo simulation of
the class of queueing systems.

Keywords: data science, random environment, regenerative simulation

1 Introduction

Time-sharing queueing systems [1] is a family of multi-class queueing system with
Bernoulli feedback. In the more modern queueing theory language they can be de-
scribed as polling systems with state-dependent routing and Bernoulli feedback. Two
main questions are of fundamental interest about such queueing systems, namely the
stability conditions and the optimal control policy. In the seminal works [1, 2] it was
assumed that the input flows are Poisson. Such assumption, typical for an early stage
of study of almost any queueing model, made it feasible to obtain a simple condition,
both necessary and sufficient at the same time, for the existence of a stationary proba-
bility distribution of the queueing process. The authors employed different techniques
to obtain stationarity conditions. In [1] it was a combination of the the mean-drift test
due to Moustafa and a limit theorem for regenerative processes. In [2], only an em-
bded ded chain was considered, and an original iterative-dominating approach was applied.
Furthermore, in the cited papers two different optimization objectives were considered:
a long-run time-average for sojourn cost of all jobs, and the stationary mean sojourn
cost of all customers per one operational tact of the server. But the so-called Klimov’s
priority indices were shown to minimize both objectives.

In [3], a time-sharing queueing system was considered whose input flows were mod-
ulated by a two-state Markov chain. Some stationarity conditions were found, some
of them were sufficient, and the others were necessary, and the stationary mean so-
journ cost of all customers was chosen as an objective for the optimization problem.
A rigorous proof for the optimality of Klimov’s indices was possible only under certain
stricter assumptions on the batch sizes distributions together with service times and
setup times distribution functions. So, a computer-aided simulation was needed to
investigate the optimality properties of Klimov’s indices. An approach with burn-in period and a quasi-stationary period was used that time. The purpose of the present research is to redo experimental study using the regenerative simulation.

2 Time-sharing system in random environment and optimal control problem

There are \( m < \infty \) input flows (of jobs) \( \Pi_1, \Pi_2, \ldots, \Pi_m \) in the system. Jobs from these flows are called primary jobs. The flows are modulated by a random environment with two states, \( e^{(1)} \) and \( e^{(2)} \). During the time the environment is in \( e^{(k)} \) the flow \( \Pi_j \) (\( j = 1, 2, \ldots, m \)) is a Poisson flow of batches, with the intensity \( \lambda^{(k)} \) of batch arrivals, and the probability \( p_x^{(j,k)} \) of batch size \( x = 1, 2, \ldots \) jobs from \( \Pi_j \) join an infinite capacity FIFO buffer \( O_j \). The server has \( n = m + 1 \) states: \( \Gamma^{(1)}, \Gamma^{(2)}, \ldots, \Gamma^{(n)} \). In the state \( \Gamma^{(r)} \) only a job from \( O_r \) is serviced, \( r = 1, 2, \ldots, m \). In the state \( \Gamma^{(n)} \) the server doesn’t serve jobs but performs inner setup. The server selects a queue for service according to some ‘switching function’ \( h(\cdot) \) as follows: if the queue sizes make a nonzero vector \( x = (x_1, x_2, \ldots, x_m) \) at the decision moment, then the service starts for the queue \( O_j \) with \( j = h(x) \). Here \( h(\cdot) \) denotes a mapping of the \( m \)-dimensional non-negative integer lattice \( X = \{0, 1, \ldots \}^m \) onto \( \{1, 2, \ldots, n\} \), and \( h(x) = j \) implies \( x_j > 0 \) for \( j = 1, 2, \ldots, m \). Only the zero vector \( \hat{0} = (0, 0, 0, \ldots) \) in \( X \) is mapped to \( n \). But if the queues are empty at the decision moment, the first arriving job will occupy the server. Jobs are serviced without interruption. A service time for a job from queue \( O_j \) has a probability distribution function \( B_j(t) \) \( (B_j(0) = 0) \). Besides ordinary service, the server needs some sort of setup times after each service. No job is serviced during setup intervals. The probability distribution function for a setup time after a queue \( O_j \) was serviced is \( B_j(t) \) \( (B(0) = 0) \). Service durations and setup durations are mutually independent and independent of the input flows. After service, the job from queue \( O_j \) can be instantly transferred to a queue \( O_r \) with probability \( p_{j,r} \), thus forming a secondary flow, or leave the system with probability \( p_{j,n} = 1 - \sum_{r=1}^{m} p_{j,r} \), where \( (n = m+1) \). So, both primary and secondary flows enter the system. The external random environment is synchronous with the service-and-setup process. It can change states only at service termination epochs and setup termination epochs. Let the environment state be a homogeneous irreducible aperiodic Markov chain. The probability for this Markov chain to go from the state \( e^{(k)} \) to the state \( e^{(l)} \) in one step is \( a_{k,l} \).

Let \( I_m \) be the \( m \times m \) identity matrix, \( Q = (p_{j,r})_{j,r} = \overline{1,m} \). It will be assumed throughout this paper that:

1) a matrix \((I_m - Q)\) is invertible;

2) the moments \( \beta_{j,s} = \int_{0}^{\infty} t^s dB_j(t), \overline{\beta}_{j,s} = \int_{0}^{\infty} t^s dB_j(t) \) are finite for \( s = 1, 2 \);

3) batch size moments \( \mu_{j,x}^{(k)} = \sum_{x=1}^{\infty} x^s p_x^{(j,k)} \) are finite for \( s = 1, 2 \).

Let \( \tau_0 = 0 \) and \( \tau_i \) be the end of the \( i \)-th working tact (these are both service periods and setup periods). Denote by \( \Gamma_0 \in \Gamma = \{\Gamma^{(1)}, \Gamma^{(2)}, \ldots, \Gamma^{(n)}\} \) the initial server
state at time $\tau_0$, by $\Gamma_j \in \Gamma$ the server state during the time interval $(\tau_{i-1}, \tau_i]$, by $\kappa_i = (\kappa_{1,i}, \kappa_{2,i}, \ldots, \kappa_{m,i})$ the vector of queues sizes at time instant $\tau_i$, by $\chi_i$ the random environment state during the time interval $(\tau_i, \tau_{i+1}]$. Then, under the model assumptions, the stochastic sequence $\{(\Gamma_j, \kappa_j, \chi_j), j = 0, 1, \ldots\}$ is a period homogeneous irreducible Markov chain.

Denote by $Q^{(r,k)}(x)$ the stationary probability of the event $\{\Gamma_j = \Gamma^{(r)}, \kappa_j = x, \chi_j = e^{(k)}\}$. It was also proven by direct computation, that the stationary probabilities

$$\sum_{x \in \mathcal{X}} (Q^{(r,1)}(x) + Q^{(r,2)}(x)), \ r = 1, 2, \ldots, m,$$

and

$$\sum_{k=1}^{2} Q^{(n,k)}(\tilde{\mathcal{U}}) \left( \lambda_1^{(k)} + \ldots + \lambda_m^{(k)} \right)^{-1}$$

are independent of the switching function $h(\cdot)$ under either of the two conditions: i) $a_{1,2} = a_{2,2}$, or ii) $a_{1,2} \neq a_{2,2}$ and $\lambda_j^{(1)} \mu_{1,j}^{(1)} = \lambda_j^{(2)} \mu_{1,j}^{(2)}$ for all $j = 1, 2, \ldots, m$.

Now, let $\zeta_{j,i}$ be the total time sojourn time of all jobs in the queue $O_j$ during the time interval $(\tau_i, \tau_{i+1}]$. Further, let the sojourn cost $c_j$ per time unit be given for the queue $O_j$, $j = 1, 2, \ldots, m$. We consider the mean sojourn cost

$$J_i(h) = \sum_{j=1}^{m} (c_j \zeta_{j,i})$$

as a measure of the control quality. In the stationary case $J_i(h) = J(h)$. Again, it was proved that Klimov’s priority indices are optimal with respect to the objective function $J(h)$ when either i) $a_{1,2} = a_{2,2}$, or ii) $a_{1,2} \neq a_{2,2}$, $\lambda_j^{(1)} \mu_{1,j}^{(1)} = \lambda_j^{(2)} \mu_{1,j}^{(2)}$ and $\lambda_j^{(1)} \mu_{2,j}^{(1)} = \lambda_j^{(2)} \mu_{2,j}^{(2)}$ for all $j = 1, 2, \ldots, m$.

3 Regenerative simulation of time-sharing queueing system

A computer-aided simulation model was constructed using the cybernetic approach [5]. Simulation takes place in discrete time scale $\{\tau_i, i = 0, 1, \ldots\}$. A multivariate stochastic sequence $\{Y_i, i = 1, 2, \ldots\}$ with $Y_i = (\tilde{\tau}_i - \tau_{i-1}, \Delta_i, \Gamma_i, \chi_i, \kappa_i, \tilde{\eta}_{i-1}i; \tilde{\zeta}_{i-1}i, \ldots, \tilde{\zeta}_{m,i-1}i)$ is generated where $\Delta_i$ is the server idle time during the tie interval $(\tau_{i-1}, \tau_i]$, $\tilde{\eta}_i = (\tilde{\eta}_{1,i}, \ldots, \tilde{\eta}_{m,i})$, $\tilde{\eta}_{j,i} = 1$ if a job from $O_j$, $\Gamma^{(\cdot)} = \Gamma_{i}$, is placed into $O_J$ after having been serviced in the time interval $(\tau_{i-1}, \tau_i]$, otherwise $\tilde{\eta}_{j,i} = 0$, $\tilde{\zeta}_{j,i} = (\tilde{\zeta}_{j,i}(1), \ldots, \tilde{\zeta}_{j,i}(\kappa_{j,i}))$, $\tilde{\zeta}_{j,i}(x)$ is the total amount of time spent in the queueing system by the $x$-th job in the queue $O_j$ by the time $\tau_i$. Then

$$\tilde{\zeta}_{j,i} = \sum_{x=1}^{\kappa_{j,i}} \min\{\tau_i - \tau_{i-1}, \tilde{\zeta}_{j,i}\}$$

and

$$J_i(h) = (f(Y_i)) \text{ with } f(\cdot) \text{ defined by } (1) \text{ and } (2).$$

The sequence $\{Y_i, i = 0, 1, \ldots\}$ is a general Markov chain and at the same time a discrete-time regenerative process with a corresponding sequence $T_i = \min\{i \geq 0: \kappa = \tilde{\Gamma}, \Gamma_i = \Gamma^{(n)}, \chi_i = e^{(1)}\}$, $T_{i+1} = \min\{i \geq T_i: \kappa = \tilde{\Gamma}, \Gamma_i = \Gamma^{(n)}, \chi_i = e^{(1)}\}$, $s = 1, 2, \ldots$ of regeneration epochs [4]. Let $Y_\infty$ be a random vector with the same probability
distribution as the stationary probability distribution of \( \{Y_i, i = 1, 2, \ldots \} \). In ergodic case, \( N^{-1}(f(Y_1) + f(Y_2) + \ldots + f(Y_N)) \rightarrow f(Y\infty) \) a.s. The regenerative method consists in using \( \left( \sum_{i=T_s}^{T_{s+1}-1} f(Y_i) \right) \left( (T_{s+1} - T_s)^{-1} \right) \) to evaluate \( f(Y\infty) \). This approach also allows to obtain a confidence interval for \( f(Y\infty) \) based on a sequence \( (V_1, \alpha_1), (V_2, \alpha_2), \ldots, (V_N, \alpha_N) \) of i.i.d. vectors, where \( V_s = \sum_{i=T_s}^{T_{s+1}-1} f(Y_i) \) and \( \alpha_s = T_{s+1} - T_s \) is the regenerative cycle length.

Let us consider now the load \( \rho \) of the queuing system, i.e. the fraction of time the queuing system is non-empty. Since in ergodic case \( 1 - \rho \) equals

\[
\lim_{N \to \infty} \left( \sum_{i=1}^{N} g_1(Y_i) \right) \left( \sum_{i=1}^{N} g_2(Y_i) \right)^{-1} = \frac{g_1(Y\infty)}{g_2(Y\infty)} = \left( \left[ \sum_{i=T_s}^{T_{s+1}-1} g_1(Y_i) \right] \right) \left( \left[ \sum_{i=T_s}^{T_{s+1}-1} g_2(Y_i) \right] \right)^{-1}
\]

where the mapping \( g_1(t, d, \Gamma^{(e)}, e, x, w; z_1, z_2, \ldots, z_m) \) for \( x = (x_1, \ldots, x_m) \), \( z_j = (z_j(1), \ldots, z_j(x_j)) \) equals \( t - d \) if \( r \neq n \), it equals 0 otherwise, and the mapping \( g_2(t, \Gamma^{(e)}, e, x, w; z_1, \ldots, z_m) = t \), the regenerative method is also applicable to evaluate the ratio of means along a cycle and to obtain a confidence interval for the ratio estimator when one sets \( V_s = \sum_{i=T_s}^{T_{s+1}-1} g_1(Y_i) \) and \( \alpha_s = \sum_{i=T_s}^{T_{s+1}-1} g_2(Y_i) \) instead of the cycle length.

Results of the computational experiments will be included in the talk.

References


