CALCULATION OF WAITING TIME DISTRIBUTION IN MULTI-CHANNEL NON-MARKOVIAN QUEUING SYSTEMS WITH “COOLING” AND “HEAT-UP”

Anatoliy D. Khomonenko,
PhD, professor, head of the Department of Information and computing systems of Emperor Alexander I St. Petersburg state transport university, professor at the Department of mathematical and software of Military space academy of A.F. Mozhaysky, St-Petersburg, Russia, khomonenko@pgups.ru

Vladimir A. Lokhvitskiy,
PhD, doctoral candidate at the Department of mathematical and software of Military space academy of A.F. Mozhaysky, St-Petersburg, Russia, lokhv_va@mail.ru

Maad M. Khalil,
postgraduate student of Emperor Alexander I St. Petersburg state transport university, St-Petersburg, Russia, maadalomar@gmail.com

ABSTRACT

The approach to calculation of distribution of waiting time of requests in the multi-channel non-Markov queuing systems with “cooling”, “heat-up” and distributions of phase type is developed. The characteristic of methods of calculation of distribution of waiting time for non-Markov queuing systems is given. At the heart of calculation of probable characteristics of multi-channel non-Markov queuing systems with approximating distributions of phase type (the hyperexponential, the Erlang, Cox) the Takakhashi-Takami's iterative method is lies. When calculating distribution of waiting time of multi-channel QS with “cooling” and “heat-up” transitions between microstates are considered. The main result is a method for calculating the waiting time distribution in a multichannel non-Markov queuing system. The method is based on the weighted convolution of the Laplace-Stieltjes transformations of the transit times of each of the components of the exponential phases of the microstates of the system. The knowledge of the Laplace-Stieltjes transformation of the waiting time distribution of the application in the queue makes it possible, by numerically differentiating this transformation at the point s = 0, to calculate the initial moments of the required distribution, from which to construct the approximation of the distribution function. The example of calculation of waiting time in queue for model of multi-channel QS M/M/E2/n with the Poisson input flow, exponentially distributed duration of service and with the "cooling" distributed under the generalized law of the Erlang of the 2nd order is considered. From the presented results it follows that with an increase in the average “cooling” duration, the average waiting time increases. In addition, the duration of “cooling” has a very significant effect on the average waiting time. The Kolmogorov distance for the distributions of the number of requests received by the numerical method and using the simulation model was {0.0014; 0.0013; 0.0038; 0.0024; 0.012} for a different intensity of “cooling”, respectively, which indicates the correctness of the analytical model. Thus, a generalization of the classical Little’s formula is achieved. The proposed approach can be useful in probabilistic modeling of nodes of distributed data processing centers, modeling and justification of the architecture of cloud computing systems with the Web interface, and evaluating the impact of costs on updating the context.

Keywords: multi-channel systems of service; “cooling”; “heat-up”; Laplace-Stieltjes conversion; distribution of waiting time; distribution of the Erlang; Little’s formula.

For citation: Khomonenko A. D., Lokhvitskiy V. A., Khalil M. M. Calculation of waiting time distribution in multi-channel non-markovian queuing systems with “cooling” and “heat-up”. *H&ES Research*. 2017. Vol. 9. No. 4. Pp. 88–94.

INTRODUCTION

In the analysis of different types of QS the following types of wait time of the request in queue are used:

the *unconditional* waiting time in W queue determined by length of an interval between the moments of arrival of the request system and its arrivals on service;

the *conditional* waiting time in W_q queues, a determined is similar to W, but in case of the additional assumption that the request surely waits ($W_q \neq 0$);

the *virtual* waiting time of W_V differing from W those that counting of length of an interval of waiting is carried not from the moment of arrival of the request in queue, but from some arbitrary time point.

From line items of applied researches searching of characteristics of time W waiting of the request in queue is of the greatest interest. Characteristics of the conditional and virtual wait time are often necessary as a monitoring aid of the results received by means of new approaches.

Among the approaches used for calculation of distribution of wait time of requests in queue to multi-channel not Markov QS it is possible to select the following three methods [1]: integral method; a method on the basis of a convolution of the Laplace-Stieltjes conversions (LSC) for QS with distributions of phase type; a count method (on the basis of invariants of the relations).

1. INTEGRAL METHOD

For M/G/n/R ≤ ∞ QS the initial moments of distributions of duration of waiting of the request in queue (stay in system) can be calculated by means of Little [2] result and his generalizations [3]. The required moments express through stationary distributions of probabilities of states on the basis of the conservation law of queue here.

We will bring correlation between stationary distribution of number of requests into queues at the time of t-0 of arrival of the request and the initial moments of distribution of waiting time in queue to the GI/G/n/R ≤ ∞ system. The generalized Little's formula and her analog for GI_2/G/n/R ≤ ∞ QS follows from it.

We will designate: z — probability that the next request is not labeled; $a(s)$ — LSC of distribution of length of an interval between adjacent requests of a recurrent flow; $a_z(s)$ — the same for interval length from arbitrary event of the entering flow to the next labeled request; $L^{-1}(\bullet)$ — operator of inverse transformation of Laplace. It is obvious that

$$a_z(s) = \sum_{k=1}^{\infty}(1-z)z^{k-1}a^k(s) = (1-z)\frac{a(s)}{1-za(s)}.$$

Respectively, $A_z(t) = L^{-1}(a_z(s)/s)$ — an interval length distribution function to the next labeled request, a $1 - A_z(t)$ — probability that any labeled request in time t won't come.

Let for the request arriving to completely busy system:

$w_c(t)$ — the conditional frequency curve of waiting time of the beginning of service (frequency curve of the conditional waiting time);

$\prod_A^-(z) = \sum_{i=0}^{\infty} z^i \pi_{A_i}^-$ — generating function of distribution of number of requests in queue at the time of $t-0$ of arrival of the new request. This function defines probability that at the time of arrival of the new request all requests in system will be not labeled.

According to the conservation law of probabilities of statuses of queue, equality shall be observed:

$$\prod_A^-(z) = \int_0^{\infty}\left[1 - A_z(t)\right] w_c(t)\,dt. \quad (1)$$

In case of the system M/G/n we have $a(s) = \lambda/(\lambda + s)$ where λ — intensity of the Poisson entering flow. At the same time

$$1 - A_z(t) = 1 - L^{-1}(\lambda(1-z)/([s(\lambda(1-z)+s)]) = e^{-\lambda(1-zt)},$$

And, equality (1) takes a form

$$\prod_A^-(z) = \int_0^{\infty} e^{-\lambda(1-z)t} w_c(t)dt = w_c(\lambda(1-z)). \quad (2)$$

Here in the right part — LSC of the conditional distribution of wait time of the request in queue with conversion parameter $\lambda(1-z)$. On the basis of conversions (2) come to the generalized Little's formula [3]:

$$q_{[k]} = \lambda^k w_k,\ k = 1,2,\ \ldots\ . \quad (3)$$

Here: w_k — k-y the initial moment of distribution of wait time of the request in queue; $q_{[k]} = \sum_{r=k}^{\infty} r(r-1)...(r-k+1)\pi_{Ar}^{-1}$ — k-y the factorial moment of distribution of number of requests in queue at the time of t-0 of arrival of the request. In case of k=1 the formula (3) is called Little's formula [2].

Yu. I. Ryzhikov in [4] offered approach to calculation of time response characteristics for QS of more general E_2/G/n and H_2/G/n type. The solution of the 1st kind Fredholm integral equation is its cornerstone:

$$\prod_A^-(z) = \int_0^{\infty}\left(F_1 e^{-ft} - F_2 e^{-gt}\right) w_c(t)dt.$$

Here F_1, F_2, f and g — the known functions z and distribution parameters of length of an interval between adjacent requests of the 2nd order of phase type.

2. METHOD OF A CONVOLUTION OF LSC OF COMPONENT PHASE DISTRIBUTIONS

In case of numerical calculation of systems of a general type GI/G/n usually not Markov distributions replace with approximations in the form of a compound of exponentially distributed phases. A method of calculation of distribution of waiting time in system GI_q/H_k/n/R ≤ ∞ offered in [5]. At the same time distribution of lengths of intervals between requests of an input flow is one of distributions of the phase type order of q: An Erlang, the hyperexponential, Cox or PH, and distribution of duration of service is the hyperexponential order of k.

An alternative to phase-type distributions is the Delta function. In particular, in the article Smagin V.A. [6] an example of applying the complex delta function for finding the stationary value of an alternating random accumulation process with the informational income and expenditure is presented.

We will explain an approach essence on the example of the M/M/n system. We will call the j-request the request which after arrival finds in queue exactly of *j* of requests. Waiting time in queue of the *j*-request represents the amount of *j* of independent random variables with LSC of components $b_n(s) = n\mu/(n\mu + s)$. Therefore LSC of the conditional distribution of wait time of the *j*-request is

$$w_j(s) = b_n^j(s) = \left(\frac{n\mu}{n\mu + s}\right)^j, \ j = 1, 2...$$

We will designate $\pi^+_{A,j}$ probability to find in system j of requests right after arrival of the request. As $w(s) = \sum_{j=1}^{\infty} \pi^+_{A,j+n} w_j(s)$ LSC of distribution of waiting time of requests in queue is:

$$w(s) = \sum_{j=1}^{\infty} \pi^+_{A,j+n} \left(\frac{n\mu}{n\mu + s}\right)^j. \qquad (4)$$

In article [5] the solution variant which generalizes a formula (4) for finding of LSC of distribution of waiting time of the request in queue in relation to QS GIq/Hk/n/R ≤ ∞ is offered. Such approach yields more general results in comparison with [4] regarding validity of q≥2 value of an order of distribution of phase type of duration of intervals between requests of an input flow.

3. COUNT METHOD

It is known that the conditional (in case of zero waiting) distribution of waiting time of the request in the GI/M/n system is subordinate to the exponential law. This circumstance allows consider that the type of distribution of waiting time doesn't depend on distribution of intervals between adjacent requests and number of channels and is defined only by distribution of holding time. In [1] the method of the setting of a type of distribution and a method of its use is offered.

Difference of arbitrary distribution from the normal, having the general with it two initial moments, it is accepted to determine by two dimensionless coefficients, (asymmetry and an excess) which are equal to zero for normal distributions. As difference of distributions from demonstrative is of interest (Markov), in [1] it is offered to enter coefficients of not Markov behavior of distribution on a formula

$$\kappa_k = w_k / w_1^k - k!, \quad k = 2, 3, ...,$$

where w_k — *k*-y the initial moment.

With the help w_1 and a set $\{\kappa_k\}$ it is possible to define the appropriate number of the higher moments [1]:

$$w_k = w_1^k(\kappa_k + k!).$$

Value w_1 can be set by means of Little's formula and probability π_0 of immediate reception of the request on service:

$$w_1 = f_1 \, a_1 / (1 - \pi_0),$$

where f_1 — average length of queue.

The marked dependence of distribution only on distribution of duration of service allows to take as a basis for determination $\{\kappa_k\}$ the M/G/1 queuing system [3]. Actually such approach represents rather evident form of use of invariants of the relations.

4. APPROACHES TO CALCULATION OF MULTI–CHANNEL QS WITH “COOLING”

Calculation of multi-channel non-Markov multi-channel QS with “cooling” and distributions of phase type based on the iterative method of Takakhasi-Takami [7]. This method successfully used when calculating non-Markov multi-channel QS of the most general classical $GI_q/G_k/n/R \leq \infty$ type. At the same time calculation of QS with “cooling” (as well as multi-channel QS with “heat-up”) is a little more difficult in comparison with classical models in case of similar assumptions of other model parameters.

QS models with “heat-up” and “cooling” have been studied in a relatively small number of publications, we will mention some of them. For example, one of the early papers [8] explores a single-channel system with “heat-up”. The paper [9] investigates a multichannel QS of the MAP/PH/n type with “heat-up” and broadband service discipline. In article [10], an QS is investigated with a fixed delay before the start of the service.

In [11] the model of multi-channel $M/M/H_2/n$ QS with hyperexponential distributed duration of "cooling" is considered. The papers [12–13] consider multichannel QS with "cooling" and Erlang distribution of the second order.

We will consider multi-channel $GI_q/G_k/n/R \leq \infty$ QS with arbitrary distributions of phase type (H_k or C_k). A state of this system it is representable a tuple $\{j; r; m_1 .... m_k\}$, where j — total number of requests in system ($j = 0,1...$), r — the current exponential branch (phase) of arrival of the request $(r = \overline{1, q})$, m_i — number of the requests undergoing service on *i*-y to an exponential branch (phase), $i = \overline{1, k}$. We will designate H_j a set of possible statuses of QS in the presence in it exactly *j* of requests, h_j — a number of elements of this set.

We will designate $\gamma_j = \left[\gamma_{j,1}, \gamma_{j,2}, ..., \gamma_{j,h_j}\right]$ vectors lines of probabilities of finding of QS in tier *j*-go microstates. We will write the vector-matrix equations of balance of transitions between the states specified on the chart.

$$\gamma_0 D_0 = \gamma_1 B_1 + \gamma_0 !_{\ 0},$$
$$\gamma_j D_j = \gamma_{j-1} A_{j-1} + \gamma_j C_j + \gamma_{j+1} B_{j+1}, \quad j = \overline{1, R},$$

where R — number of the shorthanded chart tiers.

Here A_j, B_j, C_j and D_j an essence of a matrix of the size $h_j \times h_{j+1}$, $h_j \times h_{j-1}$, $h_j \times h_j$, $h_j \times h_j$ of the conditional intensity of transitions down arrival of requests, transitions up completion of service of requests, transitions within *j*-go of a tier and goings from tier *j*-go states, respectively. For GIq/Gk/n/R ≤ ∞ QS elements of the listed matrixes can be created only algorithmically according to the principles [11].

For the solution of this system of equations, we will use an iterative method [4]. As result of calculation, we will receive the relations of adjacent probabilities numbers of applications in system $x_j = p_{j+1}/p_j, \ j = \overline{1, R}$.

After the end of iterations, using x_j values, we will realize transition to probabilities of states of system on the following algorithm:

Probability of the free state (p_0) to suppose equal 1.

To calculate $p_{j+1} = p_j x_j, \ j = \overline{1, R}$.

To calculate the amount $S = \sum_{j=0}^{R} p_j$.

To normalize the received values: $p_j = p_j/S,\ j=\overline{0,R}$.

Calculation is possible also for systems without restriction of the buffer. In this case, supposed that "tail" of distribution of number of requests represents infinitely decreasing geometrical progression which denominator is equal to the relation of two last calculated probabilities.

5. FEATURES OF CALCULATION OF WAITING TIME IN QS WITH "COOLING" AND "HEAT-UP"

Calculation of time response characteristics is based on the way, the offered A. D. Khomonenko in article [5]. Unlike the multi-channel non-Markov QS considered in article [4] in systems with "cooling" (and "heat-up") it is necessary to consider expenses of time for transitions between microstates of one tier (with fixed j). We will consider technology of the accounting of these expenses when calculating distribution of waiting time of the request in QS queue with "cooling".

Waiting time of again arrived request is defined by a system microstates right after its arrival. We will enter for each tier of the chart a row vector $\pi_j = \left[\pi_{j,1}, \pi_{j,2}, ..., \pi_{j,h_j}\right], j=\overline{1,R}$, final distribution of probabilities of microstates of system right after arrival of the next request. As the input flow not Poisson, conditions of the theorem of PASTA (Poisson Arrival See Time Average) [4] aren't satisfied, and final distribution π_j doesn't match stationary distribution γ_j.

Vector components π_j represent the relative numbers of arrivals of requests with which arrival the system passed into the appropriate microstates:

$$\pi_j = \gamma_{j-1} A_{j-1} \Big/ \sum_{i=0}^{R} \gamma_i A_i \mathbf{1}_{i+1},$$

where $\mathbf{1}_i$ — the single column vector the $h_i \times 1$size.

We will define $B_{n+1}(s)$ as a matrix of the conditional LSC of exponential distributions of lengths of intervals before transition of QS from the microstates $(j+n,i),\ i=\overline{1,h_{n+1}}$ on completion of service increased by system transition probability in one of $(j+n-1,l),\ l=\overline{1,h_{n+1}}$ microstates. The matrix has the dimension $h_{n+1} \times h_{n+1}$, its elements are calculated according to

$$b_{n+1,i,l}(s) = \frac{b_{n+1,i,l}}{\sum_{r=1}^{h_{n+1}} b_{n+1,i,r} + s},\ i,l=\overline{1,h_{n+1}}. \qquad (5)$$

We will similarly define a matrix of $C_{n+1}(s)$ LSC of distributions of duration of transitions between QS microstates on one tier (in case of the fixed number of requests in system), caused by the phases "cooling" ("heat-up"). The matrix of $C(s)$ has dimensionality$h_{n+1} \times h_{n+1}$, its elements are calculated according to:

$$A_{n+1,i,l}(s) = \frac{A_{n+1,i,l}}{\sum_{r=1}^{h_{n+1}} A_{n+1,i,r} + s},\ i,l=\overline{1,h_{n+1}}\ . \qquad (6)$$

We will call the k-request the request right after which arrival the system appears in a microstate $(k+n,i),\ j=1,2,...,\ i=\overline{1,h_{n+1}}$ (in queue there is exactly k of requests). Waiting time in queue of the k-request represents the amount of durations of k of advances of queue on completion of service plus "cooling" time if the system is in the appropriate mode.

Based on the accepted designations, LSC of waiting time of the k-request:

$$W_k(s) = \pi_{k+n} B_n^k(s) \sum_{i=0}^{r} C_{n+1}^i(s), \qquad (7)$$

where r — number of sequential phases of "cooling". We will explain a physical sense of a formula (7). Depending on in what microstate there was a system with arrival of the k-request, time of its waiting will make $i=\overline{1,r}$ of the stages "cooling" plus k of completions of service. As distribution of the amounts of random variables represents the multiplication of their LSC, the formula takes a form (7).

Then expression for LSC of distribution of waiting time of the request in queue taking into account "cooling" is:

$$W(s) = \sum_{k=1}^{R} \pi_{k+n} B_n^k(s) \sum_{i=0}^{r} C_{n+1}^i(s) = \sum_{k=1}^{R} \pi_{k+n} W_k(s). \qquad (8)$$

Notes.

1. It is possible to show that the formula (8) is fair also for multi-channel non-Markov QS with "heat-up", with "heat-up" and "cooling".

2. The knowledge of the Laplace-Stieltjes transformation of the waiting time distribution of the application in the queue makes it possible, by numerically differentiating this transformation at the point $s = 0$, to calculate the initial moments of the required distribution, from which to construct the approximation of the distribution function.

6. AN EXAMPLE OF THE CALCULATION OF A QS M/M/E$_2$/n WITH "COOLING"

We will consider model of multi-channel QS M/M/E$_2$/n with the Poisson input flow, exponentially distributed duration of service and with the "cooling" distributed under the generalized law of the Erlang of the 2nd order. The diagram of transitions between states of this system is shown in fig. 1.

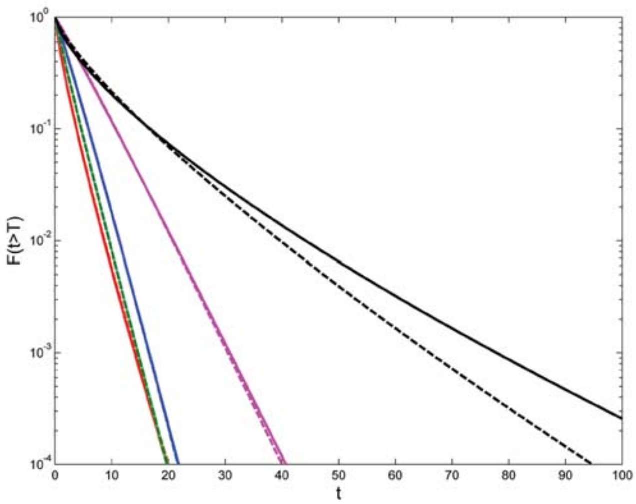

Fig. 1. Diagram for M/M/E$_2$/n QS with "cooling"

For $M/M/E_2/n$ QS the matrixes of the intensity of transitions has following view:

$$A_j = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}, j = \overline{0, N-1};$$

$$B_1 = \begin{bmatrix} 0 & \mu & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B_j = \begin{bmatrix} \min(n, j)\mu & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, j = \overline{2, N};$$

$$C_j = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \mu_c \\ \mu_c & 0 & 0 \end{bmatrix}, j = \overline{0, N};$$

$$D_j = \begin{bmatrix} \lambda + \min(j, n)\mu & 0 & 0 \\ 0 & \lambda + \mu_c & 0 \\ 0 & 0 & \lambda + \mu_c \end{bmatrix}, \; j = \overline{0, N}.$$

We will calculate the waiting time for the following initial data: $\lambda = 2.5$, $\mu = 1$, $\mu_c = \{1.0; 0.5; 0.3; 0.1; 0.05\}$. Waiting time calculating results are presented in the table.

The Kolmogorov distance for the distributions of the number of applications received by the numerical method and using the simulation model was {0.0014; 0.0013; 0.0038; 0.0024; 0.012} for a different intensity of "cooling", respectively, which indicates the correctness of the analytical model.

We calculate the additional distribution function (ADF) of the waiting time for applications in the system. To do this, we represent it in the form of a Weibull ADF with a correction polynomial [4] and construct a plot of the approximation obtained (Fig. 2).

From the presented results it follows that with an increase in the average "cooling" duration, the average waiting time increases. In addition, the duration of "cooling" has a very significant effect on the average waiting time.

CONCLUSION

The proposed approach to calculating the LSC of the waiting time distribution of the request in the queue to the QS with approximating distributions of the phase type generalizes the results of [1–4] to the case of multichannel non-Markov systems "cooling" and "warming up". The proposed approach can be useful in probabilistic modeling of nodes of distributed data processing centers, modeling and justification of the architecture of cloud computing systems with the Web interface, and evaluating the impact of costs on updating the context. Examples of this application of models of multichannel non-Markov QS with «cooling» and «heating» are given in [14–17]. Further research is advisable to continue in the directions of research systems with prioritized services [18], optimal deterministic algorithms and adaptive heuristics for energy and performance efficiency [19].

References

1. Ryzhikov Y. I. Three methods for calculating the time characteristics of open queuing systems. *Automation and Remote Control*. 1993. 54:2. Pp. 127–133. (In Russian)

2. John D. C. Little. A Proof for the Queuing Formula: $L = \lambda W$. *Operation Research*. No. 9 (3): Pp. 383–387.

3. Brumelle S. L. A generalization of $L = \lambda W$ to moments of queue length and waiting times. *Oper. Res*. 1972. Vol. 20. No. 6. Pp. 1127–1136.

4. Ryzhikov Y. I. *Teorija ocheredej i upravlenie zapasami* [Queuing theory and inventory management]. St.-Petersburg: Piter. 2001. 384 p. (In Russian)

5. Khomonenko A. D. The distribution of the waiting time in queuing systems of the type GIq/Hk/n/R ≤∞. *Automation and Remote Control*. 1990. No. 51:8. Pp. 91–98. (In Russian).

6. Smagin V. A. Complex Delta Function and its Information Application. *Automatic Control and Computer Sciences*. 2014. Vol. 48. No. 1. Pp. 10–16.

7. Takahashi Y., Takami Y. A numerical method for the steady-state probabilities of a GI/G/c queueing system in a general class. *J. of the Operat. Res. Soc. of Japan*. 1976. Vol. 19. No. 2. Pp. 147–157.

8. Kreinin Ya. Single-channel queueing system with warm up. *Automation and Remote Control*. 1980. No. 41:6. Pp. 771–776.

9. Bin Sun, Dudin A. N. The MAP/PH/N multi-server queuing system with broadcasting service discipline and server heating. *Automatic Control and Computer Sciences*. 2013. Vol. 47. Issue 4. Pp. 173–182.

Table

Initial moments of the distribution of the waiting time for requests

μ_c	w_1		w_2		w_3	
	Numerical	*Simulation*	*Numerical*	*Simulation*	*Numerical*	*Simulation*
1,00	1.534E+00	1,542E+00	5,747E+00	5,824E+00	2,987E+01	3,057E+01
0,50	1.968E+00	1,974E+00	8,143E+00	8,169E+00	4,482E+01	4,489E+01
0,30	2.624E+00	2,646E+00	1,302E+01	1,314E+01	8,329E+01	8,403E+01
0,10	4.696E+00	4,758E+00	4,268E+01	4,311E+01	6,127E+02	6,169E+02
0,05	6.473E+00	6,625E+00	1,272E+02	1,144E+02	4,133E+03	3,766E+03

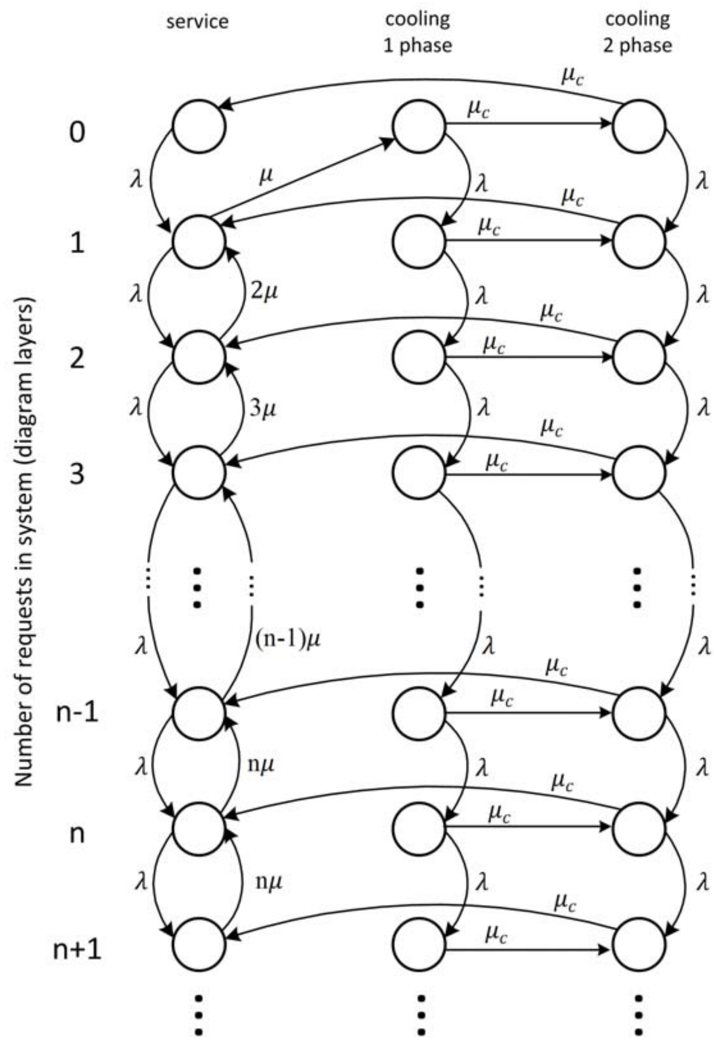

Fig. 2. ADF of the waiting time in a QS M/M/E_2/n with "cooling"

10. Eremin A. S. A Queueing System with Determined Delay in Starting the Service. *Intellectual Technologies on Transport*. 2015. No. 4. Pp. 23–26.

11. Lokhvitskii V.A., Ulanov A.V. Numerical analysis of queuing systems with Hyper exponential "cooling. *Tomsk State University Journal of Control and Computer Science*. 2016. No. 4(37). Pp. 36–43. (In Russian)

12. Khomonenko A. D., Gindin S. I., Khalil M. M. A cloud computing model using multi-channel queuing system with cooling. 2016 XIX IEEE International Conference on Soft Computing and Measurements (SCM). 2016. Pp. 103–106.

13. Khalil M. M., Andruk A. A. Testing of Software for Calculating a Multichannel Queuing System with "Cooling" and E2-approximation. *Intellectual Technologies on Transport*. 2016. No. 4 (8). Pp. 22–28.

14. Khomonenko A. D., Khalil M. M., Kassymova D. T. Probabilistic Models for Evaluating the Performance of Cloud Computing Systems with Web Interface. *SPIIRAS Proceedings*. 2016. No. 6 (49). Pp. 49–65.

15. Murugesan R, Elango C., Kannan S. Resource Allocation in Cloud Computing with M/G/s Queueing System. *International Journal of Advanced Research in Computer Science and Software Engineering.* 2014. Vol. 4. Issue 9. Pp. 443–447.

16. Kaczynski W. H. Transient Queueing Analysis. *INFORMS Journal on Computing*. 2012. Vol. 24. No. 1. Pp. 10–28.

17. Vilaplana J., Solsona F., Teixid´o I. A performance model for scalable cloud computing. Proceedings of the 13th Australasian Symposium on Parallel and Distributed Computing (AusPDC2015), Sydney, Australia, 27–30 January 2015. Pp. 51–60.

18. Muliukha V., Ilyashenko A., Zayats O., Zaborovsky V. Preemptive Queueing System with Randomized Push-Out Mechanism. *Communications in Nonlinear Science and Numerical Simulation*. 2015. Vol. 21. No. 1–3. Pp. 147–158.

19. Beloglazov A., Buyya R. Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers. *Concurrency and Computation: Practice and Experience*. 2012. Vol. 24. Pp. 1397–1420.

РАСЧЕТ РАСПРЕДЕЛЕНИЯ ВРЕМЕНИ ОЖИДАНИЯ В ОЧЕРЕДИ МНОГОКАНАЛЬНЫХ НЕМАРКОВСКИХ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ С «ОХЛАЖДЕНИЕМ» И «РАЗОГРЕВОМ»

Хомоненко Анатолий Дмитриевич,
д.т.н., профессор, заведующий кафедрой информационных и вычислительных систем Петербургского государственного университета путей сообщения Императора Александра I профессор кафедры математического и программного обеспечения Военно-космической академии имени А.Ф.Можайского г. Санкт-Петербург, Россия, khomonenko@pgups.ru

Лохвицкий Владимир Александрович,
к.т.н., докторант кафедры математического и программного обеспечения Военно-космической академии имени А.Ф.Можайского, г. Санкт-Петербург, Россия, lokhv_va@mail.ru

Халил Маад Модер,
аспирант кафедры информационных и вычислительных систем Петербургского государственного университета путей сообщения Императора Александра I, г. Санкт-Петербург, Россия, maadalomar@gmail.com

АННОТАЦИЯ

Предлагается подход к вычислению распределения времени ожидания запросов в многоканальных немарковских системах массового обслуживания с «охлаждением», «разогревом» и распределениями фазового типа. Дана характеристика методов расчета распределения времени ожидания для немарковских систем массового обслуживания. В основе расчета вероятностных характеристик многоканальных немарковских систем массового обслуживания с аппроксимирующими распределениями фазового типа (гиперэкспоненциальное, Эрланга, Кокса) лежит итерационный метод Такахаши-Таками. При расчете распределения времени ожидания в многоканальной системе массового обслуживания с «охлаждением» и «разогревом» учитываются переходы между микросостояниями одного уровня. Основной результат – метод расчета распределения времени ожидания в многоканальной немарковской системе массового обслуживания с «охлаждением» и/или «разогревом». Метод основан на взвешенной свертке преобразований Лапласа-Стилтьеса времен прохождения каждой из составляющих экспоненциальных фаз микросостояний системы. Знание преобразования Лапласа-Стилтьеса распределения времени ожидания заявки в очереди позволяет путем численного дифференцирования этого преобразования в точке s=0 рассчитать начальные моменты искомого распределения, по ним построить аппроксимацию функции распределения. Рассмотрен пример численного расчета распределения времени ожидания заявки в очереди модели многоканальной системы массового обслуживания типа $M/M/E_2/n$ – с пуассоновским входным потоком, экспоненциально распределенной продолжительностью обслуживания и с «охлаждением», распределенной по обобщенному закону Эрланга 2-го порядка. Из представленных результатов следует, что с увеличением средней продолжительности «охлаждения» среднее время ожидания увеличивается. Кроме того, длительность «охлаждения» оказывает значительное влияние на среднее время ожидания. Расстояние Колмогорова для распределений числа заявок, полученных численным методом и с помощью имитационной модели, составило {0.0014; 0.0013; 0.0038; 0.0024; 0.012} при различной интенсивности «охлаждения» соответственно, что свидетельствует о корректности аналитической модели. Таким образом, достигнуто обобщение классической формулы Литтла. Предложенный подход может быть полезен при вероятностном моделировании узлов центров распределенной обработки данных, моделировании и обосновании построения архитектуры систем облачных вычислений с Web-интерфейсом, оценке влияния затрат на актуализацию контекста.

Ключевые слова: многоканальные системы обслуживания; «охлаждение»; «разогрев»; преобразование Лапласа-Стилтьеса; распределение времени ожидания; распределение Эрланга; формула Литтла.

Для цитирования: *Хомоненко А.Д., Лохвицкий В.А., Халил М.М.* Расчет распределения времени ожидания в очереди многоканальных немарковских систем массового обслуживания с «охлаждением» и «разогревом» // Наукоемкие технологии в космических исследованиях Земли. 2017. Т. 9. № 4. С. 88–94.