MULTI-SERVER QUEUEING SYSTEM OF THE MAP/PH/N TYPE WITH SERVERS HEATING AND BROADCASTING SERVICE DISCIPLINE

Bin Sun, Alexander Dudin

Laboratory of Applied Probabilistic Analysis
Department of Applied Mathematics and Computer Science
Belarusian State University
4, Nezavisimosti Ave., 20030 Minsk, Belarus
telephone: + (37517) 2095486; fax: + (37517) 2265548
e-mail: sunbin_bsu@hotmail.com, dudin@bsu.by

We consider the multi-server queueing system of the MAP/PH/N type with preliminary heating of the servers that are idle at a customer arrival epoch and broadcasting service discipline which assumes that the customer is served by all free servers. The key performance measures of the system and steady state distribution of the underlying multidimensional Markov chain are computed.

Key words – Broadcasting, heating times, MAP (Markovian Arrival Process), PH (phase type)

1 INTRODUCTION

Multi-server queueing systems adequately describe operation of fragments of many communication networks and have got a lot of attention in literature. However, such systems are investigated in much less extent comparing to the single server systems although they represent much more reach source of different interesting mathematical models.

A typical assumption in analysis of the multi-server queueing systems is that the service time distributions of all homogeneous customers are identical. However, in some real-life systems service time distribution can be different in servers, which start the service a customer from a buffer immediately after completion of the service of a previous customer, and customers, which arrive when this server is empty. E.g., service of the first request of a user, which arrives to the empty channel may require additional time to activating the channel and connection initializing. Following [1], we call this additional time as heating time.

The single server queueing systems of such a type are already well investigated, see, e.g., early work [1] and a lot of another works. Analysis of the single server queueing systems with heating the servers is very close to investtigation of the systems with vacations and is not very difficult from the mathematical point of view because behavior of the queue has here peculiarity only at one state when the number of customers in the system is equal to zero.

The standard assumption in analysis of multi-server queues is classical service discipline: each customer is served by one server. In this paper, we investigate the case that heating of all free servers is started at the customer arrival epoch and the customer gets a service after the finish of the heating of all free servers. Such a service discipline creates some redundancy, but it can help to decrease the average delivering times of a first copy of the broadcasted customer, see [2]-[4].

Study of the multi-server queueing systems with heating servers and broadcasting service discipline is more complicated because, along with counting the number of customers currently presenting in the system, one has to keep track on how many servers are processing the customers and how many servers are heated currently. This is a reason why, to the best of our knowledge, the multi-server queueing systems with heating servers and broadcasting service discipline still are not investigated in literature except the paper [5]. In that paper such a system was investigated under the simplest assumption about the service process. Service time is assumed to exponentially distributed and service discipline is assumed to be classical discipline. Account of heating time is made by including it to service time and assuming another service intensity. Here we consider the model under much general assumptions. We assume that the service and heating times have PH (phase type) distributions and broadcasting service discipline.

The rest of the paper is organized as follows. In section 2, the model is described and the steady state joint distribution of the number of customers in the system and servers states is analyzed in section 3.

2 MATHEMATICAL MODEL

We consider an N-server queueing system. The servers are assumed to be identical and independent of each other. The customers arrive to the system according to the MAP. Customers arrival in the MAP is directed by an irreducible continuous time Markov chain $\nu_t, t \geq 0$, with the finite state space $\{0,1,\ldots,W\}$. Sojourn time of the Markov chain $\nu_t, t \geq 0$, in the state ν has an exponential distribution with parameter $\lambda_\nu, \nu = \overline{0,W}$. After this sojourn time expires, with probability $p_k(\nu,\nu')$, the process $\nu_t, t \geq 0$, transits to the state ν', and k customers, $k = 0,1$, arrive into the system. The intensities of jumps from one state into another, which are accompanied by an arrival of k customers, are combined into the matrices $D_k, k = 0,1$, of size $(W+1)\times(W+1)$. The matrix generating function of these matrices is $D(z) = D_0 + D_1 z, |z| \le 1$. The matrix $D(1)$ is the infinitesimal generator of the process $\nu_t, t \ge 0$. The stationary distribution vector $\boldsymbol{\theta}$ of this process satisfies the equations $\boldsymbol{\theta} D(1) = \mathbf{0}, \boldsymbol{\theta}\mathbf{e} = 1$. Here and in the sequel $\mathbf{0}$ is the zero row vector and $\mathbf{e}$ is the column vector of appropriate size consisting of 1's.

The average intensity λ (fundamental rate) of the MAP is defined as $\lambda = \boldsymbol{\theta} D_1 \mathbf{e}$.

For more information about the MAP, its special cases and properties and related research see, e.g., [6], [7] and the survey paper by S. Chakravarthy [8]. Usefulness of the MAP in modeling modern telecommunication systems is discussed in [9], [10].

If an arriving customer meets one or more servers being free, it occupies these servers. Heating of the all free servers are started. Heating time is random and has PH type distribution. It means the following. Server heating time is governed by the underlying process $n_t, t \ge 0$, which is continuous time Markov chain with state space $\{1, ..., M_2\}$. The initial state of the process $n_t, t \ge 0$, at the epoch of starting the heating is determined by the probabilistic row-vector $\boldsymbol{\beta}_2 = (\beta_2^{(1)}, ..., \beta_2^{(M_2)})$. The transitions of the process $n_t, t \ge 0$, that do not lead to heating completion, are defined by the irreducible matrix S_2 of size $M_2 \times M_2$. The intensities of transitions, which lead to heating completion, are defined by the column vector $\mathbf{S}_2^{(0)} = -S_2 \mathbf{e}$. The heating time distribution function has the form $B_2(x) = 1 - \boldsymbol{\beta}_2 e^{S_2 x}\mathbf{e}$.. The average heating time is given by $b_2^{(1)} = \boldsymbol{\beta}_2 (-S_2)^{-1}\mathbf{e}$. The more detailed description of the PH-type distribution and its partial cases can be found e.g. in the book [11].

After the finish of the heating, the service of a customer begins. Service time also has PH type distribution. It is governed by the underlying process $m_t, t \ge 0$, which is continuous time Markov chain with state space $\{1, ..., M_1\}$. The initial state of the process $m_t, t \ge 0$, at the epoch of starting the service is determined by the probabilistic row-vector $\boldsymbol{\beta}_1 = (\beta_1^{(1)}, ..., \beta_1^{(M_1)})$. The transitions of the process $m_t, t \ge 0$, that do not lead to service completion, are defined by the irreducible matrix S_1 of size $M_1 \times M_1$. The intensities of transitions, which lead to service completion, are defined by the vector $\mathbf{S}_1^{(0)} = -S_1 \mathbf{e}$. The service time distribution function has the form $B_1(x) = 1 - \boldsymbol{\beta}_1 e^{S_1 x}\mathbf{e}$. The average service time is given by $b_1^{(1)} = \boldsymbol{\beta}_1 (-S_1)^{-1}\mathbf{e}$.

If an arriving customer meets all servers busy, it is placed to the buffer having infinite capacity. It will be picked up from the buffer for the service according to the FIFO discipline. Service time has the same distribution as the service time in the heated server.

3 STEADY-STATE DISTRIBUTION OF THE SYSTEM STATES

Let:

- $i_t, i_t \ge 0$, be the number of customers in the system;
- $k_t, 0 \le k_t \le \min\{i_t, N\}$, be the number of heated servers;
- $\nu_t, \nu_t = \overline{0, W}$, be the state of the underlying process of the MAP;
- $m_t^{(j)}, m_t^{(j)} = \overline{1, M_1}$, be the state of the underlying process of the service in the j th server providing the service, $0 \le j \le \min\{N, i\} - k_i$;
- $n_t^{(l)}, n_t^{(l)} = \overline{1, M_2}$, be the state of the underlying process of the heating process in the l th heated server, $0 \le l \le k_i$,

at the moment $t, t \ge 0$. Here we assume that the busy and heated servers are numerated in order of their occupying, i.e. the server, which begins the service or heating, is appointed the maximal number among all busy or heated servers; when some server finishes the service or heating, the servers are correspondingly enumerated.

It is easy to see that the process

$$\xi_t = (i_t, k_t, \nu_t, \{m_t^{(j)}\}, \{n_t^{(l)}\}), t \ge 0,$$

is the continuous time multi-dimensional Markov chain. It can be verified that this Markov chain is ergodic if and only if the following inequality is fulfilled:

$$\rho = \lambda b_1^{(1)} < 1. \quad (3.1)$$

In what follows we assume that this condition is fulfilled. Then the following steady-state (stationary) probabilities exist:

$$\pi(i, k, \nu, \{m^{(j)}\}, \{n^{(l)}\}) =$$
$$= \lim_{t \to \infty} P\{i_t = i, k_t = k, \nu_t = \nu, \{m_t^{(j)} = m^{(j)}\}, \{n_t^{(l)} = n^{(l)}\}\},$$
$$m^{(j)} = \overline{1, M_1}, n^{(l)} = \overline{1, M_2},$$
$$0 \le j \le \min\{N, i\} - k, 0 \le l \le k, k = \overline{0, i}, i \ge 0.$$

Let us enumerate the states of the Markov chain $\xi_t, t \ge 0$, in the lexicographic order and combine steady-state probabilities of the states having the value (i, k) of the components (i_t, k_t) of the Markov chain (we call these states as the macro-state (i, k)) into the row-vectors $\boldsymbol{\pi}(i, k)$ and then introduce the vectors

$$\boldsymbol{\pi}_i = (\boldsymbol{\pi}(i, 0), ..., \boldsymbol{\pi}(i, \min\{N, i\})), i \ge 0.$$

For use in the sequel, we introduce the following denotations.

- I is an identity matrix. If the dimension of the matrix is not clear from context, it will be indicated by the suffix. E.g., $I_{\overline{W}}$ is the identity matrix of dimension $\overline{W} = W+1$. O is zero square matrix. $\oplus$ and $\otimes$ are symbols of the Kronecker sum and product of matrices, see, e.g., [12].
- $\boldsymbol{\beta}^{\otimes l} \stackrel{def}{=} \underbrace{\boldsymbol{\beta}\otimes\cdots\otimes\boldsymbol{\beta}}_{l}, l \ge 1, S_1^{\oplus l} \stackrel{def}{=} \underbrace{S_1 \oplus \cdots \oplus S_1}_{l}, l \ge 1$.
- $\mathbf{S}_1^{(0)\oplus l} \stackrel{def}{=} \sum_{m=0}^{l-1} I_{M_1^M} \otimes \mathbf{S}_1^{(0)} \otimes I_{M_1^{l-m-1}}, l \ge 1.$

Let Q be the generator of the Markov chain $\xi_t, t \ge 0$, with blocks $Q_{i,j}$ consisting of intensities $(Q_{i,j})_{k,k'}$ of the Markov chain $\xi_t, t \ge 0$, transitions from the macro-state (i,k) into the macro-state $(j,k'), k,k' = \overline{0,K}$. The diagonal entries of the matrix $Q_{i,j}$ are negative and the modulus of the diagonal entry of $(Q_{i,j})_{k,k}$ defines the total intensity of leaving the corresponding state of the Markov chain.

Lemma 1. The generator Q of the Markov chain $\xi_t, t \ge 0$, is given by

$$Q = \qquad (3.2)$$

$$\begin{pmatrix}
\text{✌}_0 & O & \cdots & O & \text{☞}_{0,N} & O & O & \cdots \\
\text{💣}_1 & \text{✌}_1 & \cdots & O & \text{☞}_{1,N} & O & O & \cdots \\
O & \text{💣}_2 & \cdots & O & \text{☞}_{2,N} & O & O & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
O & O & \cdots & \text{✌}_{N-1} & \text{☞}_{N-1,N} & O & O & \cdots \\
O & O & \cdots & \text{💣}_N & \text{✌}_N & \text{☞}_{N,N} & O & \cdots \\
O & O & \cdots & O & \text{💣}_{N+1} & \text{✌}_N & \text{☞}_{N,N} & \cdots \\
O & O & \cdots & O & O & \text{💣}_{N+1} & \text{✌}_N & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{pmatrix} \mathrm{w}$$

here the non-zero blocks $(\text{✌}_i)_{r,r'}$ of the matrices $\text{✌}_i, 0 \le i \le N$, are defined by:

$D_0 \oplus S_1^{\oplus i}$, if $r = r' = 0$,

$D_0 \oplus S_1^{\oplus(i-r)} \oplus S_2^{\oplus r}$, if $r = r', r = \overline{1,i}$,

$I_{\overline{W}} \otimes I_{M_1^{i-r}} \otimes \boldsymbol{\beta}_1 \otimes \mathbf{S}_2^{(0)\oplus r}$, if $r' = r-1, r = \overline{1,i}$,

the non-zero blocks $(\text{💣}_i)_{r,r'}$ of the matrices $\text{💣}_i, 1 \le i \le N$, are defined by:

$$I_{\overline{W}} \otimes \mathbf{S}_1^{(0)\oplus(i-r)} \otimes I_{M_2^r}, r = r', r = \overline{0,i},$$

the non-zero blocks $(\text{💣}_{N+1})_{r,r'}$ of the matrix 💣_{N+1} are defined by:

$$I_{\overline{W}} \otimes (\mathbf{S}_1^{(0)} \otimes \boldsymbol{\beta}_1)^{\oplus(N-r)} \otimes I_{M_2^r}, r = r', r = \overline{0,N},$$

the non-zero blocks $(\text{☞}_{i,N})_{r,r'}$ of the matrices $\text{☞}_{i,N}$, are defined by:

$$D_1 \otimes I_{M_1^{i-r}} \otimes I_{M_2^r} \otimes \boldsymbol{\beta}_2^{\otimes(N-i)}, 0 \le i < N, r = \overline{0,i}, r' = \overline{N-i,N}$$

the non-zero blocks $(\text{☞}_{N,N})_{r,r'}$ of the matrix $\text{☞}_{N,N}$ are defined by:

$$D_1 \otimes I_{M_1^{N-r}} \otimes I_{M_2^r}, r' = r, r = \overline{0,N}.$$

Proof. Proof of the lemma is implemented by means of analysis of the probabilities of the Markov chain $\xi_t, t \ge 0$, transitions during an infinitesimal time interval. The matrices ✌_i define intensities of possible transitions which do not imply the change of the value i of the first component of the Markov chain. The matrix $S_1^{\oplus i}$ defines intensity of transition of an underlying service process in one of i busy servers that does not cause service termination. The matrix $S_2^{\oplus r}$ defines intensity of transition of an underlying heating process in one of r heated servers that does not imply the heating termination. The vector $\mathbf{S}_2^{(0)\oplus r}$ defines intensity of transition of an underlying heating process in one of r heated servers that causes the heating termination. In this case the number of the heated servers is decreased by 1 and the service of a customer begins. The vector $\boldsymbol{\beta}_1$ defines the initial state of the underlying process for the service in this server. The matrices 💣_i define intensities of possible transitions which imply the change of the value i of the first component of the Markov chain to the value $i-1$ of this component, namely, intensity of the service completion in one of the busy servers. In the case $i \ge N+1$, there is a queue in the system. So, one customer immediately occupies the server, which finishes the service, and the vector $\boldsymbol{\beta}_1$ defines the initial state of the underlying process for the service in this server. The matrices $\text{☞}_{i,N}$ define intensities of possible transitions which imply the change of the value i of the first component of the Markov chain to the value $i+1$ of this component, new customer arrival. If $i < N$, this customer immediately initializes the heating process in one or more free servers. The vector $\boldsymbol{\beta}_2$ defines the initial state of the underlying process for this heating. If $i \ge N+1$, this customer moves to the buffer without initializing any process.

Proof is based on careful use of specifics of the generator Q in the first $N + 1$ block columns with combination with technique by M. Neuts it is possible to verify that the stationary probability vectors $\boldsymbol{\pi}_i, i \geq 0$, are defined by the following statement.

Theorem 1. The stationary probability vectors $\boldsymbol{\pi}_i, i \geq 0$, are defined by

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_{N+1} ❄_i, i \leq N, \qquad (3.3)$$

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_{N+1} ☼^{i-N-1}, i > N, \qquad (3.4)$$

where

$$❄_i = -💣_{i+1} ✌_i^{-1}, i \leq N-1, \qquad (3.5)$$

$$❄_N = -💣_{N+1} (\sum_{i=0}^{N-1} ✍_i ☞_{i,N} + ✌_☠)^{-1}, \qquad (3.6)$$

$$✍_i = \prod_{j=i}^{N-1} ❄_j, i = \overline{0, N-1}, ✍_N = ❄_N. \qquad (3.7)$$

The matrix ☼ is the minimal non-negative solution to the equation

$$☼^{▣} 💣_{N+1} + ☼ ✌_☠ + ☞_{N,N} = O, \qquad (3.8)$$

and the vector $\boldsymbol{\pi}_{N+1}$ is defined as the unique solution to the system

$$\boldsymbol{\pi}_{N+1}(☼ 💣_{N+1} + ✌_☠ + ☞_{N,N} ❄_N) = \mathbf{0}, \qquad (3.9)$$

$$\boldsymbol{\pi}_{N+1}(\sum_{i=0}^{N} ✍_i ♏ + (I - ☼)^{-1} \mathbf{e}) = 1. \qquad (3.10)$$

Having calculated the stationary probability vectors $\boldsymbol{\pi}_i, i \geq 0$, it is possible to compute different performance measures of the system.

Average number L of customers in the system is computed by:

$$\mathbf{L} = \sum_{i=0}^{\infty} i \boldsymbol{\pi}_i \mathbf{e} =$$

$$= \boldsymbol{\pi}_{N+1}(\sum_{i=0}^{N} ♓ ✍_i + (N+1)(I - ☼)^{-1} + ☼(I - ☼)^{-2})\mathbf{e}.$$

Probability that an arbitrary customer will be served by the server, which requires heating, is computed by:

$$P_{heat} = \sum_{i=0}^{N-1} \sum_{k=0}^{i} \boldsymbol{\pi}(i,k) \frac{(D_1 \otimes I_{M_1^{i-k}} \otimes I_{M_2^k})\mathbf{e}}{\lambda}.$$

Average number of busy servers at arbitrary time is computed by:

$$N_{busy} = \sum_{i=0}^{N} \sum_{k=0}^{i} (i\text{-}k) \boldsymbol{\pi}(i,k)\mathbf{e} + \sum_{i=N+1}^{\infty} \sum_{k=0}^{N} (N\text{-}k) \boldsymbol{\pi}(i,k)\mathbf{e}.$$

Average number of heating servers at arbitrary time is computed by:

$$N_{heat} = \sum_{i=0}^{N} \sum_{k=0}^{i} k \boldsymbol{\pi}(i,k)\mathbf{e} + \sum_{i=N+1}^{\infty} \sum_{k=0}^{N} k \boldsymbol{\pi}(i,k)\mathbf{e}.$$

Average number of free servers at arbitrary time is computed by:

$$N_{free} = \sum_{i=0}^{N} (N - i) \boldsymbol{\pi}_i \mathbf{e}.$$

REFERENCES

[1] A queueing system with warm-up and queue dependent service time / Shakhbazov A.A. // Izv. Acad. Sci. AsSSR Ser Phys.-Tech. Math – 1974 – P. 32-35.

[2] The SM/M/N queueing system with broadcasting service / Lee M.H., Dudin A.N., Klimenok V.I. // Mathematical Problems in Engineering. – 2006. – V.2006. – Article ID 98171. – 18 p.

[3] Multi-server queueing model with broadcasting service / Lee M.H., Dudin A., Klimenok V., Kim C. J. // IEEE Communications Letters. – 2007. –V.11. – P. 546-548.

[4] Multi-server queueing systems with cooperation of the servers / Kim C.S., Lee M.H., Dudin A.N., Klimenok V.I. // Annals of Operations Research. – 2008. – V.162. – P. 57-68.

[5] The MAP/M/N queueing model with exceptional service of fresh customers / M.H. Lee, A. Dudin, V. Klimenok, S. Chernov, O. Kostyukova // Proceedings of the 6th International Workshop on Multimedia Signal Processing and Transmission, Jeonju, Korea. – 2006. – P. 37-46.

[6] The Markov-modulated Poisson process (MMPP) cookbook / W. Fisher, K.S. Meier-Hellstern // Performance Evaluation 18 – 1993. P. 149-171.

[7] New results on the single server queue with a batch Markovian arrival process / D.M. Lucantoni // Communications in Statistics-Stochastic Models 7 – 1991.– P. 1-46.

[8] The batch Markovian arrival process: a review and future work / S.R. Chakravarthy // in: A. Krishnamoorthy, et~al. (Eds.), Advances in Probability Theory and Stochastic Process: Proc., Notable Publications, NJ 2001. P. 21-49.

[9] Modelling multiple IP traffic streams with rate limits / D.P. Heyman, D. Lucantoni // IEEE/ACM Transactions on Networking 11 – 2003. – P. 948-958.

[10] Modelling IP traffic using the batch Markovian arrival process / A. Klemm, C. Lindermann, M. Lohmann // Performance Evaluation 54 – 2003. – P.149-173.

[11] Matrix-geometric solutions in stochastic models / M.F. Neuts // The Johns Hopkins University Press, Baltimore. 1981.

[12] Kronecker products and matrix calculus with applications / A. Graham A. // Chichester: Ellis Horwood Ltd. – 1981.