STATIONARY DISTRIBUTION OF THE TWO-PHASE QUEUE WITH INTERMEDIATE BUFFER

O. Taramin
Belarusian State University
Minsk, Belarus
taramin@bsu.by

We consider a tandem queueing system consisting of two phases and having an intermediate buffer. The first phase is represented by the $BMAP/G/1$ queue. After service at the first phase a customer proceeds to the second phase which is described by a multi-server queue with a finite buffer. A customer leaves the system forever or waits service at the second phase in the buffer if it completes the service at the first phase and meets the buffer being full. The waiting period is accompanied by blocking the first phase server operation. The stationary distribution of the system states at an arbitrary time is derived. The dependence of the system performance measures on the correlation in the input flow and buffer capacity is numerically illustrated.

Keywords: Tandem queue, intermediate buffer, quasi-Toeplitz Markov chain.

1. INTRODUCTION

Tandem queueing systems are good models of many real-world two-node networks as well as fragments of general computer and communication systems. They are widely used in capacity planning and performance evaluation of telecommunication networks, service centers, and manufacturing systems. In this paper, we deal with a tandem queue under the assumption that customers arrive according to a batch Markovian arrival process (*BMAP*) which is an ideal to model correlated and bursty traffic in modern telecommunication networks. The queue under consideration has a multi-server second phase and an intermediate buffer providing losses of customers and blocking the first phase server after the service completion at the first phase. Previously such a system has been considered in [1] where the results concerning the condition for the stable operation of the system and the stationary distribution of the system states at embedded epochs were obtained. Here the steady state distribution of the system states at an arbitrary times is investigate.

2. MATHEMATICAL MODEL

We consider a tandem queueing system consisting of two phases. The first phase is represented by the $BMAP/G/1$ queue. The *BMAP* input is defined by means of the underlying process ν_t, $t \geq 0$, which is an irreducible continuous-time Markov chain with the finite state space $\{0, \ldots, W\}$ where W is some finite integer. Arrivals occur only at the epochs of the process ν_t, $t \geq 0$, transitions. The intensities of transitions which are accompanied by an arrival of a batch consisting of k customers are combined to the matrices $D_k, k \geq 0$, of size $(W+1) \times (W+1)$. The matrix generating function of these matrices is $D(z) = \sum_{k=0}^{\infty} D_k z^k, |z| \leq 1$. The matrix $D(1)$ is an infinitesimal generator of the process $\nu_t, t \geq 0$. The stationary distribution vector $\boldsymbol{\theta}$ of this process satisfies the equations $\boldsymbol{\theta}D(1) = \mathbf{0}, \boldsymbol{\theta}\mathbf{e} = 1$, where $\mathbf{e}$ is a column vector consisting of 1's, $\mathbf{0}$ is a row vector consisting of 0's. The main characteristics of the $BMAP$ are calculated by: the average intensity $\lambda = \boldsymbol{\theta}D'(z)|_{z=1}\mathbf{e}$; the average intensity of group arrivals $\lambda_b = \boldsymbol{\theta}(-D_0)\mathbf{e}$; the variation coefficient of intervals between successive group arrivals $c_{var}^2 = 2\lambda_b\boldsymbol{\theta}(-D_0)^{-1}\mathbf{e} - 1$; the correlation coefficient of the successive intervals between group arrivals $c_{cor} = (\lambda_b\boldsymbol{\theta}(-D_0)^{-1}(D(1) - D_0)(-D_0)^{-1}\mathbf{e} - 1)/c_{var}^2$.

The successive service times of customers at the first phase are independent random variables with general distribution $B(t)$ and finite first moment $b_1 = \int_0^\infty t dB(t)$. After receiving service at the first phase a customer proceeds to the second phase which is represented by N independent identical servers. The service time by a server is exponentially distributed with the parameter $\mu > 0$. The second phase has a finite buffer of capacity $R < \infty$. If a customer completes the service at the first phase and meets the buffer being full, it leaves the system forever with probability γ, $0 \leq \gamma \leq 1$. With probability $1 - \gamma$ the customer waits its service at the second phase in the buffer. The waiting period is accompanied by blocking the first phase server operating.

3. THE STATIONARY DISTRIBUTION AT AN ARBITRARY TIME

The process of the system states at an arbitrary time is defined as $\xi_t = \{i_t, r_t, \nu_t\}$, $t \geq 0$, where i_t, $i_t \geq 0$, is the number of customers at the first phase, r_t, $r_t = \overline{0, N+R}$, is the number of customers at the second phase, ν_t, $\nu_t = \overline{0, W}$, is the state of the $BMAP$ at time t, $t \geq 0$. The process ξ_t, $t \geq 0$, is non-Markovian but it can be shown that ξ_t, $t \geq 0$, is a semi-regenerative process (see [2]) with the embedded Markov renewal process $\{\xi_n, t_n\}$, $n \geq 1$. Here t_n is the nth service completion at the first phase, $\xi_n = \{i_n, r_n, \nu_n\}$, $n \geq 1$, is an embedded Markov chain, where i_n is the number of customers at the first phase at the epoch $t_n + 0$, $i_n \geq 0$; r_n is the number of customers at the second phase at the epoch $t_n - 0$, $r_n = \overline{0, N+R}$; ν_n is the state of the arrival directing process ν_t at the epoch t_n, $\nu_n = \overline{0, W}$.

Using the theory of semi-regenerative processes, the stationary distribution of the process ξ_t, $t \geq 0$, can be related to the stationary distribution of the embedded Markov chain ξ_n, $n \geq 1$ investigated in [1]. In this paper the condition for the stable operation was derived and the stationary distribution of the chain ξ_n, $n \geq 1$, was calculated by means of the use of results for quasi-Toeplitz Markov chains [3]. The sufficient condition for existence of the process ξ_t, $t \geq 0$, stationary distribution coincides with the necessary and sufficient condition for ergodicity of the chain ξ_n, $n \geq 1$. So, we will use the results [1] to obtain the stationary distribution of the process ξ_t, $t \geq 0$.

For further use in the sequel, we introduce the following notation:

- I is an identity matrix of appropriate dimension. When needed the dimension of the matrix is identified with a suffix;
- $\otimes$ is the sign of the Kronecker product of matrices, see, e.g, [4];
- $diag\{a_l,\ l=\overline{1,L}\}$ is a diagonal matrix with diagonal entries a_l;
- $\tilde{D}_k = I_{N+R+1} \otimes D_k,\ k \geq 0,\ \tilde{D}(z) = \sum\limits_{k=0}^{\infty} \tilde{D}_k z^k,\ |z| \leq 1;$
- $P(n,t)$ is a matrix whose (ν,ν')th entry defines the probability that n customers arrive during the interval $(0,t]$ and the state of the process ν_t at the epoch t is ν', given $\nu_0 = \nu$. The matrices $P(n,t)$ are defined by the expansion $\sum\limits_{n=0}^{\infty} P(n,t)z^n = e^{D(z)t}$;
- $\delta_{r,r'}(t)$ is the probability that in the time interval of the length t the number of customers at the second phase is decreased from r to r':

$$\delta_{r,r'}(t) = \begin{cases} 0,\ r < r'; \\ \begin{pmatrix} r \\ r' \end{pmatrix} e^{-\mu r' t}(1-e^{-\mu t})^{r-r'},\ 0 \leq r' \leq r \leq N; \\ \int\limits_0^t \frac{N\mu(N\mu\tau)^{r-N-1}}{(r-N-1)!} e^{-N\mu\tau} \begin{pmatrix} N \\ r' \end{pmatrix} e^{-\mu r'(t-\tau)} \\ \times(1-e^{-\mu(t-\tau)})^{N-r'},\ 0 \leq r' < N < r \leq N+R; \\ \frac{(N\mu t)^{r-r'}}{(r-r')!} e^{-N\mu t},\ N < r \leq N+R,\ N \leq r' \leq N+R, \end{cases}$$

- $\delta(t) = (\delta_{r,r'}(t))_{r,r'=\overline{0,N+R}}, \qquad \Delta = (\Delta_{r,r'})_{r,r'=\overline{0,N+R}}, \qquad \Delta_{r,r'} = \int\limits_0^{\infty} \delta_{r,r'}(t)e^{D_0 t}dt;$
- Q_1, Q_2 are square matrices of dimension $N+R+1$:

$$Q_1 = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix},\ Q_2 = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix};$$

- $\bar{W} = W + 1;$
- $\tilde{Q}_m = Q_m \otimes I_{\bar{W}},\ m = 1,2;$
- $\hat{Q} = \tilde{Q}_1 + \gamma\tilde{Q}_2 + (1-\gamma)Q_2 \otimes (N\mu(-D_0 + N\mu I)^{-1});$
- $\Omega_n = \int\limits_0^{\infty} \delta(t) \otimes P(n,t)(1-B(t))dt,\ n \geq 0,\ \Omega(z) = \sum\limits_{n=0}^{\infty} \Omega_n z^n;$
- R_1, R_2, R_3 are diagonal matrices of dimension $N+R+1$: $R_1 = diag\{r,\ r = \overline{0,N+R}\},\ R_2 = diag\{0,\dots,N,0,\dots,0\},\ R_3 = diag\{0,\dots,0,1,\dots,R\}$.

Denote the stationary state probabilities of the Markov chain $\xi_n = \{i_n, r_n, \nu_n\}$ by $\pi(i,r,\nu),\ i \geq 0,\ r = \overline{0,N+R},\ \nu = \overline{0,W}$. Form the row vectors $\boldsymbol{\pi}_i,\ i \geq 0$, of these probabilities listed in the lexicographic order and let $\Pi(z) = \sum\limits_{i=0}^{\infty} \boldsymbol{\pi}_i z^i,\ |z| \leq 1$, be the generating function of the stationary state probabilities.

Denote the stationary state probabilities of the process $\xi_t,\ t \geq 0$, as $p(i,r,\nu) = \lim\limits_{t\to\infty} P\{\xi_t = (i,r,\nu)\},\ i \geq 0, r = \overline{0,N+R}, \nu = \overline{0,W}$. Form the row vectors $\boldsymbol{p}(i,r) =$

$(p(i,r,0), p(i,r,1), \dots, p(i,r,W))$, $p_i = (p(i,0), p(i,1), \dots, p(i,N+R))$, $i \ge 0$, of these probabilities and introduce the generating function $P(z) = \sum_{i=0}^{\infty} p_i z^i$, $|z| \le 1$.

Theorem 1. *The stationary probability vectors p_i, $i \ge 0$, of the process $\xi_t = \{i_t, r_t, \nu_t\}$, $t \ge 0$, are related to the stationary probability vectors π_i, $i \ge 0$, of the embedded Markov chain ξ_n, $n \ge 1$, as follows:*

$$p_0 = \tau^{-1} \pi_0 \hat{Q} \Delta,$$

$$p_i = \tau^{-1}\{\pi_0 \sum_{k=1}^{i} [\hat{Q}\Delta\tilde{D}_k + (1-\gamma)Q_2 \otimes \int_0^\infty P(k,t) N\mu e^{-N\mu t} dt]\tilde{\Omega}_{i-k}$$

$$+(1-\gamma)\pi_0 Q_2 \otimes \int_0^\infty P(i-1,t) e^{-N\mu t} dt$$

$$+\sum_{l=0}^{i-1} \pi_l [(\tilde{Q}_1 + \gamma\tilde{Q}_2)\tilde{\Omega}_{i-l} + (1-\gamma)\sum_{k=0}^{i-1} Q_2 \otimes \int_0^\infty P(k,t) N\mu e^{-N\mu t} dt \tilde{\Omega}_{i-k-l}]$$

$$+(1-\gamma)\sum_{l=0}^{i-1} \pi_l Q_2 \otimes \int_0^\infty P(i-l-1,t) e^{-N\mu t} dt\}, \ i \ge 1,$$

where τ is the mean inter-departure time at the first phase

$$\tau = \pi_0 \hat{Q}(-D_0)^{-1} e + \Pi(1)[(\tilde{Q}_1 + \tilde{Q}_2) b_1 + (1-\gamma)\tilde{Q}_2 (N\mu)^{-1}] e.$$

Corollary 1. *The generating function $P(z)$ is expressed in the following way:*

$$P(z) = p_0 + \tau^{-1}\{\pi_0[\hat{Q}\Delta(\tilde{D}(z) - \tilde{D}_0)$$

$$+(1-\gamma)Q_2 \otimes N\mu((-D(z) + N\mu I)^{-1} - (-D_0 + N\mu I)^{-1})]\tilde{\Omega}(z)$$

$$+(1-\gamma) z \Pi(z) Q_2 \otimes (-D(z) + N\mu I)^{-1}$$

$$+(\Pi(z) - \pi_0)[\tilde{Q}_1 + \gamma\tilde{Q}_2 + (1-\gamma)Q_2 \otimes N\mu(-D(z) + N\mu I)^{-1}]\tilde{\Omega}(z)\}.$$

Based on the stationary distribution at an arbitrary time, we can calculate variuos performance measures of the system at an arbitrary time as follows:

- The mean number of customers at the first phase $\bar{L}_1 = P'(1)e$.
- The mean number of customers at the second phase $\bar{L}_2 = P(1)(I_{N+R+1} \otimes e_{\bar{W}}) R_1 e$.
- The mean number of busy servers at the second phase $\bar{N}_{busy} = P(1)(I_{N+R+1} \otimes e_{\bar{W}}) R_2 e$.
- The mean number of customers in the buffer $\bar{N}_{buffer} = P(1)(I_{N+R+1} \otimes e_{\bar{W}}) R_3 e$.
- The probability that the server of the first phase is idle $P_{idle} = p_0 e$.
- The probability that the server of the first phase processes a customer $P_{serve} = \tau^{-1} b_1$.
- The probability that the server of the first phase is blocked $P_{block}^{(server)} = 1 - P_{idle} - P_{serve}$.

4. NUMERICAL EXAMPLES

The goal of the numerical examples is to analyze the influence of the correlation in the *BMAP* and buffer capacity on performance measures of the system.

We consider four *BMAP*s with maximal group size $k = 3$. The $BMAP_1$ is a group Poisson process. It has $c_{cor} = 0$ and $c_{var} = 1$. $BMAP_2$, $BMAP_3$ and $BMAP_4$ have the same fundamental rate $\lambda = 2.5$, variation coefficients $c_{var} = 2$, and different correlation coefficients $c_{cor} = 0.1$, $c_{cor} = 0.2$, and $c_{cor} = 0.3$ respectively.

The service time at the first phase is deterministic, $b_1 = 0.2$. The number of servers at the second phase $N = 3$, the probability of blocking after first phase $\gamma = 0.5$, the mean service rate $\mu = 1$.

Figures 1-3 illustrate the dependence of key system performance measures on the buffer size R for the *BMAP*s with different correlation coefficient.

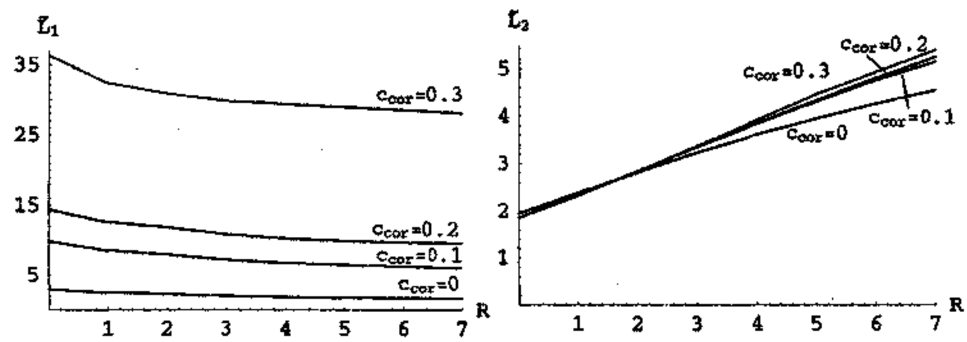


Fig. 1. The mean number of customers at the first phase and at the second phase as functions of the buffer size for the BMAPs with different correlation

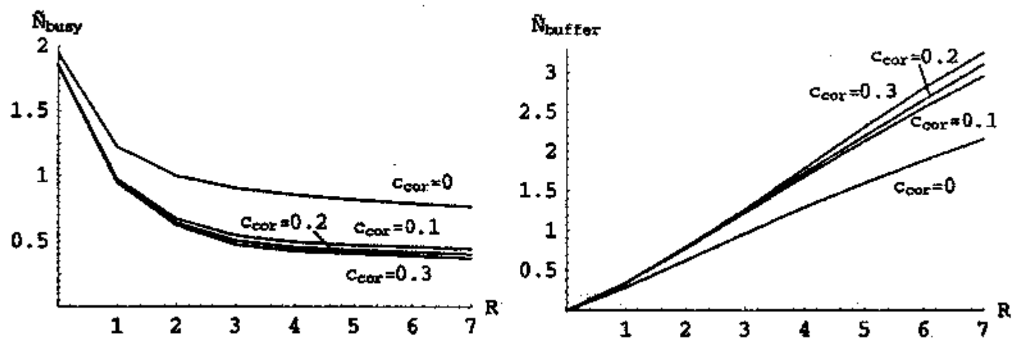
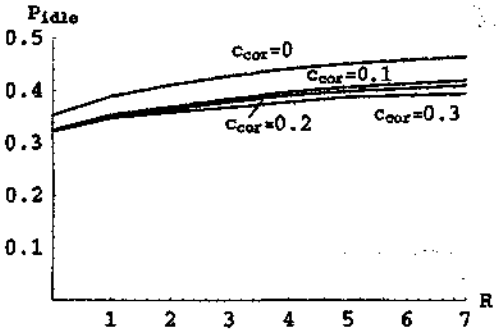


Fig. 2. The mean number of busy servers at the second phase and customers in the buffer as functions of the buffer size for the BMAPs with different correlation



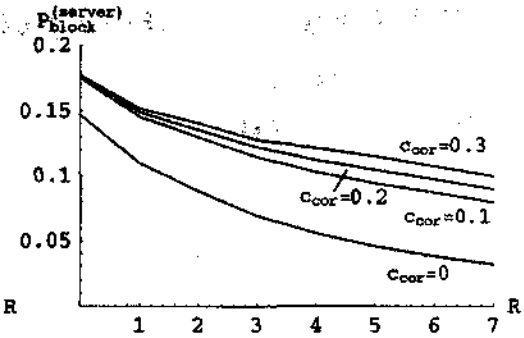


Fig. 3. The probability that the server of the first phase is idle and blocked as functions of the buffer size for the BMAPs with different correlation

Based on the figures one can conclude that the values of $\bar{L}_1$, $\tilde{N}_{busy}$, $P^{(server)}_{block}$ decrease and the values of $\bar{L}_2$, $\tilde{N}_{buffer}$, P_{idle} increase when the buffer size R increases for all $BMAP$s. It is clear that under the fixed value of the buffer size R the increase of the correlation in the input flow has negative impact on the key performance measures of both phases of the tandem and the quality of service in the system becomes worse when the correlation increases. So, it is evident that correlation of the arrival process must be taken into account because its ignoring can cause errors in prediction of system operation.

5. CONCLUSION

In this paper, the two-phase queue with intermediate buffer is studied. The process of the system states at an arbitrary time is investigated. The formulas for calculating the steady state probabilities are obtained. The numerical examples illustrated importance of the special treatment of the tandem queueing models with the *BMAP* are presented. The results can be used for capacity planning, performance evaluations, and optimization of real-world two-node networks in case of correlated bursty input and the discipline of admission to the second phase providing losses of customers and blocking the first phase server.

REFERENCES

1. *Taramin O.S.* Two-phase queue with Markovian arrival process // Vestnik BGU. 2008 (in Russian, to appear).
2. *Cinlar E.* Introduction to stochastic process // N.J.: Prentice-Hall. 1975.
3. *Klimenok V.I., Dudin A.N.* Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory // Queueing Systems. 2006. V. 54. P. 245–259.
4. *Graham A.* Kronecker products and matrix calculus with applications // Cichester: Ellis Horwood. 1981.