BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 58, No. 4, 2010
DOI: 10.2478/v10175-010-0072-9

Loss tandem networks with blocking – a semi-Markov approach

W. ONISZCZUK*

Faculty of Computer Science, Bialystok University of Technology, 45A Wiejska St., 15-351 Białystok, Poland

Abstract. Based on the semi-Markov process theory, this paper describes an analytical study of a loss multiple-server two-station network model with blocking. Tasks arrive to the tandem in a Poisson fashion at a rate λ, and the service times at the first and second stations are non-exponentially distributed with means s^A and s^B, respectively. Between these two stations there is a buffer with finite capacity. In this type of network, if the buffer is full, the accumulation of new tasks (jobs) by the second station is temporarily suspended (blocking factor) and tasks must wait on the first station until the transmission process is resumed. Any new task that finds all service lines at the first station occupied is turned away and is lost (loss factor). Initially, in this document, a Markov model of the loss tandem with blocking is investigated. Here, a two-dimensional state graph is constructed and a set of steady-state equations is created. These equations allow the calculation of state probabilities for each graph state. A special algorithm for transforming the Markov model into a semi-Markov process is presented. This approach allows calculating steady-state probabilities in the semi-Markov model. In the next part of the paper, the algorithms for calculation of the main measures of effectiveness in the semi-Markov model are presented. Finally, the numerical part of this paper contains an investigation of some special semi-Markov models, where the results are presented of the calculation of the quality of service (QoS) parameters and the main measures of effectiveness.

Key words: loss two-station network with blocking, exact Markovian algorithm, semi-Markov model.

1. Introduction

In the mathematical models of discrete flow systems, which are realistic and effective tools for performance analysis of a wide class of systems such as computer systems and networks, telecommunication networks, transportation networks, production lines, or flexible manufacturing systems, queuing network models (QNM) with finite capacity queues and blocking are often used [1–8]. Finite capacity queuing network models are of great value towards effective congestion control and quality of service (QoS) protection of modern discrete flow networks. Blocking in such networks arises because the traffic of jobs through one queue may be momentarily halted if the destination queue has reached its capacity. Over the years, many publications related to the analysis and application of QNMs with finite capacity queues and blocking in the field of computer science, operations research, traffic engineering or industrial engineering have been written [9–16].

Exact closed-form solutions for QNMs with blocking are not generally attainable except for some special cases. As a consequence, numerical techniques and analytic approximation have been proposed for the study of arbitrary QNMs with non-Markovian service times under various types of blocking mechanisms. Authoritative expositions of the subject appear in Perros [17] and Balsamo et al. [18]. However, there is still a great interest in the systems with buffer capacity limitations under different blocking mechanisms [19–21]. A blocking mechanism restricts the total intensity of input streams by enforcing certain limitations the blocking and synchronization procedures [22–25]. Such models are in constant demand for the performance evaluation and predication of more complex systems such as high-speed telecommunication networks or flexible manufacturing systems, etc.

Most research in the area of two-station (tandem) open networks with blocking (see for example [17]) assumes that each queue is served by a single server, where the first station has an infinite or a finite capacity and the second station has a finite capacity. The state of this queuing network can be described by the pair of variables indicating the number of tasks in the first station and the number of tasks in the second station. The various closed-form results related to the single server queuing network include the following two limiting cases: when a task at the first station receives an infinitesimal amount of service and when the first station is saturated. Another special tandem model with blocking assumes that multiple servers serve each queue. In this case, upon completion of service at the first station, a task will get blocked if at that moment the second station is full. We say that a station is saturated when there is always at least one task waiting for service, i.e. the station is never empty. Similarly, other authors studied the tandem configuration with exponential service times and no intermediate buffers, and no queue in front of the first station or where the first station was assumed to have an infinite (or a finite) capacity.

This paper extends the author's previous research on the open tandem model with blocking [6]. The former paper only considers Markov multiple server two-station queuing networks with blocked separated serving lines assuming that the first station is under heavy load. The current publication examines an open non-Markov loss tandem with blocked separated lines at the first station assuming that the first station has no queue. In both cases, when a departure occurs from the sec-

*e-mail: w.oniszczuk@pb.edu.pl

ond station, one of the blocked tasks will enter the second station and its associated serving line will become unblocked.

This paper provides the mathematical study of a special type of network configuration (tandem), as shown in Fig. 1. This kind of network has N parallel lines at the first station, and the other station with c parallel servicing lines. Between these stations is a common waiting buffer with finite capacity, for example equal to m. When the buffer is full, the accumulation of new tasks from the first station is temporarily suspended and a phenomenon called blocking occurs, until the queue empties and allows new inserts. This is the classical mechanism for controlling the intensity of the arriving task stream, which comes to the two-station network.

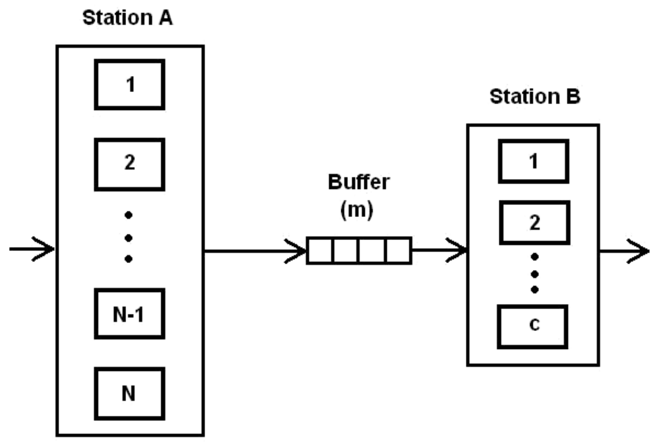

Fig. 1. Loss tandem network with blocking

In this kind of tandem configuration, no more than $N + m + c$ tasks can be processed simultaneously and then the tandem becomes idle, if there are no tasks in both stations. Assuming that the input stream to the tandem network represents a Poisson process and the service time in both stations corresponds to a random variable with non-exponential distribution, it is a non-Markov model of loss tandem with blocking. At the beginning of this paper all states of the tandem network are defined, then steady state probabilities and the main tandem measures of effectiveness are calculated. Additionally, algorithms for calculation of blocking and loss probabilities, delay time in the buffer, blocking time in the station A, the percentage of buffer filling, etc. is shown.

The structure of the paper is as follows. Section 2 specifies the tandem model and shows procedures for finding the state probabilities in a semi-Markov tandem model, in Sec. 3, the procedures for calculating the main measures of effectiveness are given. Model implementation and a numerical example are described in Sec. 4. Finally, conclusions are drawn in Sec. 5.

2. Semi-Markov analysis of a loss tandem with blocking

Queuing networks with finite capacity queues have been introduced to represent systems with finite capacity resources and population constraints. When a queue reaches its maximum capacity then the flow of jobs (tasks) into the service station is stopped, and the blocking phenomenon arises. Let us consider the two-station network with blocking as shown in Fig. 1. The input task stream comes to station A. This station has no buffer and it can accept only N incoming tasks. New tasks, which arrive at the full first station, are not accepted and are rejected. Each task at the first station is processed on the parallel service lines and upon service completion sent to station B. If there are free lines on this station, the service process starts immediately, if not, the tasks must wait in the buffer. If the buffer is full, any task upon service completion at the station A, is forced to wait and blocks this service line.

The general assumptions for this tandem model are:

- external task stream arriving at station A is assumed to be a Poisson stream, with rate $\lambda = 1/a$, where a is the mean inter-arrival time,
- station A has N parallel service lines,
- c service lines are available at station B,
- in both stations the service time for each task represents a non-exponentially distributed random variable, with mean $s^A = 1/\mu^A$ and $s^B = 1/\mu^B$, where μ is mean service rate,
- the buffer capacity is finite, for example equal to m.

Under these assumptions, if the buffer is full, any task upon completion of service at station A, is forced to wait in its service line, because the transfer process from the station A, depends only on the service process in station B. Physically, blocked tasks stay on station A, but the nature of the service process in station B, allows one to treat them as located in additional places in the buffer and they belong to station B. In this case, there can be a maximum of $c + m + N$ tasks assigned to the second station including all tasks in the first station that can be blocked (the maximum number of states in the two-dimensional tandem state space, that may belong to the second station is equal to $c + m + N$).

In turn, the maximal number of non-blocked tasks (the maximal number of unblocked, active servers) in the first station is equal to N. It means that the current number of tasks that belong to the second station depends on the number of non-blocked tasks in station A (let it be fixed as i). Therefore, the current number of states in station B (let us denote it as j) is equal to $j = c + m + N - i$. If the numbers of tasks located simultaneously at the tandem in the first and second stations are denoted by i and j, then a semi-Markov model with two-dimensional state space and with a unique path from the state $(0, 0)$ to any state (i, j) and back to the state $(0, 0)$ is defined in this paper (see Figs. 2 and 3).

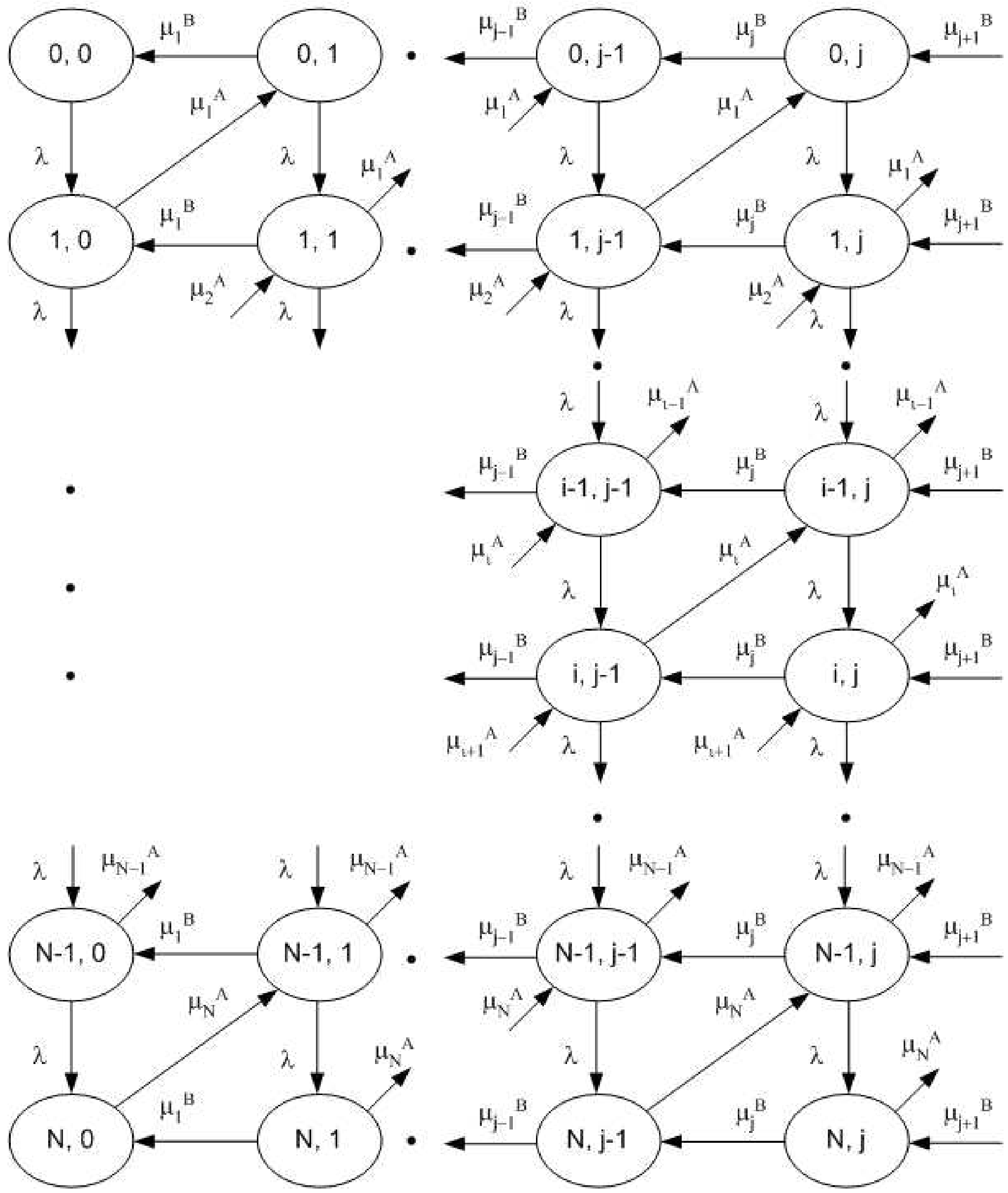

Fig. 2. Two-dimensional tandem state diagram (first part)

According to the general approach for analyzing semi-Markov models, the first step is to find solutions to the classical Markov model. Generally, queuing networks with blocking are difficult to solve, because their steady state probabilities cannot be shown to have a product-form solution. Hence, most of the techniques that are employed to analyse these networks are in the form of approximation or numerical techniques. Numerical methods are particularly useful in cases where it is not possible to obtain an analytic solution for the queuing system under study. The equivalent Markov queuing system under study (with the same service rates) is first formulated as a continuous-time Markov process with discrete states, and subsequently its steady-state probability vector is calculated using an equation solving technique [17, 18]. A queuing network with blocking, under appropriate assumptions, can be formulated as a Markov process and the stationary probability vector can be obtained using numerical methods for linear systems of equations.

Before describing the equations for calculation of steady-state probabilities, we need to define the service rates for station A:

$$\begin{gathered}\mu_1^A = \mu^A,\ \ \mu_2^A = 2\cdot\mu^A,\ \ldots,\\ \mu_i^A = i\cdot\mu^A,\ \ldots,\mu_N^A = N\cdot\mu^A\end{gathered} \tag{1}$$

and station B:

$$\begin{gathered}\mu_1^B = \mu^B,\ \ \mu_2^B = 2\cdot\mu^B,\ \ldots,\\ \mu_c^B = c\cdot\mu^B,\ \ldots,\mu_{c+m+N}^B = c\cdot\mu^B.\end{gathered} \tag{2}$$

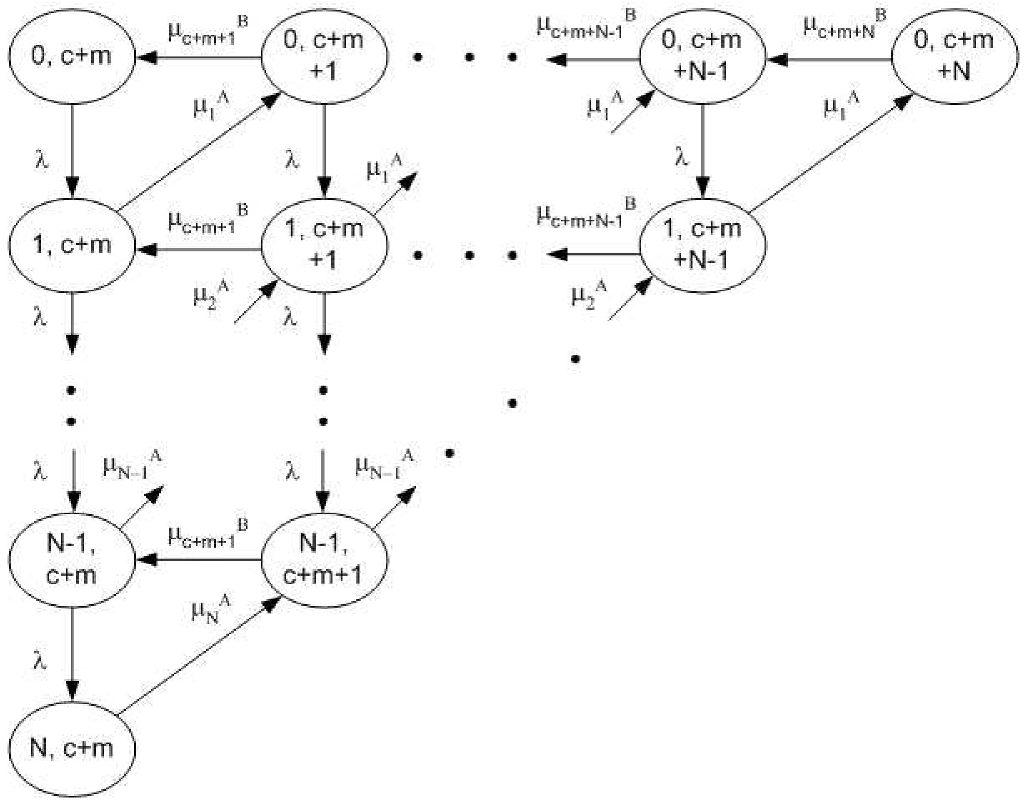

Fig. 3. Two-dimensional tandem state diagram (second part)

Based on an analysis of this state space diagram, the process of constructing the steady-state equations can be divided into several independent steps, which describe several similar, repeatable schemas (see Figs. 2 and 3). These steady-state equations are:

$$\begin{gathered}
\lambda \cdot p_{0,0} = \mu_1^B \cdot p_{0,1} \\
\text{for} \quad i=0, j=0, \\
(\lambda + \mu_j^B) \cdot p_{0,j} = \mu_1^A \cdot p_{1,j-1} + \mu_{j+1}^B \cdot p_{0,j+1} \\
\text{for} \quad i=0, j=1,\ldots,c+m, \\
(\lambda + \mu_i^A) \cdot p_{i,0} = \lambda \cdot p_{i-1,0} + \mu_1^B \cdot p_{i,1} \\
\text{for} \quad i=1,\ldots,N-1, j=0, \\
(\lambda + \mu_j^B + \mu_i^A) \cdot p_{i,j} = \lambda \cdot p_{i-1,j} + \\
+ \mu_{i+1}^A \cdot p_{i+1,j-1} + \mu_{j+1}^B \cdot p_{i,j+1} \\
\text{for} \quad i=1,\ldots,N-1, \;\; j=1,\ldots,c+m, \\
\mu_N^A \cdot p_{N,0} = \lambda \cdot p_{N-1,0} + \mu_1^B \cdot p_{N,1} \\
\text{for} \quad i=N, j=0, \\
(\mu_j^B + \mu_N^A) \cdot p_{N,j} = \lambda \cdot p_{N-1,j} + \mu_{j+1}^B \cdot p_{N,j+1} \\
\text{for} \quad i=N, j=1,\ldots,c+m-1, \\
(\mu_{c+m}^B + \mu_N^A) \cdot p_{N,c+m} = \lambda \cdot p_{N-1,c+m} \\
\text{for} \quad i=N, j=c+m.
\end{gathered} \tag{3}$$

And for states with blocking the equations are:

$$\begin{gathered}
(\lambda + \mu_j^B) \cdot p_{0,j} = \mu_1^A \cdot p_{1,j-1} + \mu_{j+1}^B \cdot p_{0,j+1} \\
\text{for} \quad i=0, j=c+m+1,\ldots,c+m+N-1, \\
\mu_{c+m+N}^B \cdot p_{0,c+m+N} = \mu_1^A \cdot p_{1,c+m+N-1} \\
\text{for} \quad i=0, j=c+m+N, \\
(\lambda + \mu_j^B + \mu_i^A) \cdot p_{i,j} = \\
= \lambda \cdot p_{i-1,j} + \mu_{i+1}^A \cdot p_{i+1,j-1} + \mu_{j+1}^B \cdot p_{i,j+1} \\
\text{for} \quad i=1,\ldots,N-2, \\
j=c+m+1,\ldots,c+m+N-1-i \\
(\mu_{c+m+N-i}^B + \mu_i^A) \cdot p_{i,j} = \\
= \lambda \cdot p_{i-1,j} + \mu_{i+1}^A \cdot p_{i+1,c+m+N-(i+1)} \\
\text{for} \quad i=1,\ldots,N-1, \qquad j=c+m+N-i.
\end{gathered} \tag{4}$$

This set of linear equations can be solved using classical numerical methods, based on algorithms typical for sparse and diagonal matrices.

Let us examine the semi-Markov model with a finite number of its states, which is for example equal to K. In the semi-Markov model, during the investigation procedure we try to find the steady-state probabilities $q_{i,j}$ that the model is in state (i,j). By using some convenient state ordering for tandem model, we may transform the state description (i,j) to the state number (k), where $k = 1, 2, \ldots, K$. Consider that we have a Markov tandem model with the identical state transition rates as the semi-Markov model. Traditionally, steady-state distributions of semi-Markov processes are found from the embedded Markov chain with a given rate transition matrix [26]. Clearly this still requires the calculation of the entire steady-state distribution of the embedded Markov process. What we achieve in this paper is the direct calculation method of individual steady-state probabilities of the semi-Markov process, which are functions solely of the sojourn time in a state [26].

Assume that for each model state is known the mean sojourn time $m_{i,j}$ (the expected time the process remains in the (i,j)th state during each visit) in a semi-Markov model [26], then if we choose the number of state changes L in the equivalent Markov model large enough, we can say that this model visited state (i,j) $L_{i,j} = p_{i,j} \cdot L$ times. By the way, the total sojourn time when the semi-Markov model was in any given state is equal to:

$$T_{i,j} = p_{i,j} \cdot L \cdot m_{i,j}. \tag{5}$$

Whereas in a semi-Markov model, the average time for L state changes, may be calculated from the following formula:

$$T = \sum_{k=1}^{K} T_k = L \cdot \sum_{k=1}^{K} (p_k \cdot m_k), \tag{6}$$

where $p_k = p_{i,j}$ and $m_k = m_{i,j}$.

Assuming that $q_{i,j}$ is the (i,j) state probability in the semi-Markov model then its entire mean sojourn time (for this state) during interval T is equal to:

$$T_{i,j} = q_{i,j} \cdot T = q_k \cdot L \cdot \sum_{k=1}^{K} (p_k \cdot m_k), \tag{7}$$

where $q_k = q_{i,j}$ and directly from relation (5) and (7) we have:

$$q_k = \frac{p_k \cdot m_k}{\sum\limits_{k=1}^{K} (p_k \cdot m_k)} \tag{8}$$

where $q_k = q_{i,j}$, $p_k = p_{i,j}$ and $m_k = m_{i,j}$ and

$$\sum_{k=1}^{K} q_k = 1. \tag{9}$$

Recapitulating, in the semi-Markov processes any steady-state probability can be calculated from the equivalent Markov models [26]. Here, the last problem, which must be solved, is how to calculate the $m_{i,j}$ parameter in a semi-Markov model, it means calculating the mean sojourn (service) time for state (i,j). At the beginning, we try to calculate another parameter $m^*_{i,j}$ – the mean time simultaneously spent by the tasks during processing by station A and station B, unless the new admission appears (the expected time until the next service completion in either node), where index i belongs to the station A and index j belongs to station B.

Let τ_A be the time until the next service completion in node A, and let τ_B be the time until the next service completion in node B (independent random variables). Let τ_{AB} be a random variable for duration of time until the next service completion in node A or B, whichever comes first (the simultaneous service time). This simultaneous service time (random variable) distribution for any tandem state can be given by the following expressions:

$$\begin{gathered} P(\tau_{AB} > x) = P(\tau_A > x, \tau_B > x) = \Phi_A(x) \cdot \Phi_B(x), \\ \Phi_A(x) = 1 - F_A(x), \\ \Phi_B(x) = 1 - F_B(x), \end{gathered} \tag{10}$$

where $F_A(x)$ and $F_B(x)$ are distribution functions of the random service times in the stations A and B. From relations (10) directly we have:

$$m^*_{AB} = \int_0^\infty \Phi_A(x) \cdot \Phi_B(x) dx. \tag{11}$$

This parameter is the mean sojourn time for the following tandem states: $(N,1)$, …, $(N,c+m)$, $(N-1,c+m+1)$, …, $(0,c+m+N)$ (see Figs. 2 and 3 – the bottom graph states, except state $(N,0)$). Calculation process of the mean sojourn time in the remaining tandem states (except state $(0,0)$) needs to include the task arrival factor, because any visit to state (i,j) would finish upon the next service completion or task arrival. According to the general assumptions for the tandem model, the external task stream is Poisson process, where the probability of k arrivals in an interval $(0-t)$ is given by:

$$p_k(t) = \frac{(\lambda \cdot t)^k}{k!} \cdot e^{-\lambda \cdot t} \quad \text{for} \quad k \geq 0, \ \ t \geq 0. \tag{12}$$

For the states (i,j) mentioned above, and during its simultaneity service time τ_{AB}, only one new task arrival may appear or not (if the state changes to $(i+1,j)$). It means that we have here only two events (a task occurrence or not):

$$p_0(t) + p_1(t) = 1 \quad \text{for} \quad t = \tau_{AB}, \tag{13}$$

where

$$p_0(t) = e^{-\lambda \cdot t}, \ p_1(t) = \lambda \cdot t \cdot e^{-\lambda \cdot t} = \lambda \cdot t \cdot p_o(t)$$

and

$$p_0(t) = \frac{1}{1 + \lambda \cdot t} \quad \text{for} \quad t = \tau_{AB}.$$

This is a task arrival factor. Remembering that a simultaneity service time τ_{AB} has the mean value equal to m^*_{AB} we may calculate a mean sojourn time m_{AB} directly from the following relation:

$$m_{AB} = m^*_{AB} \cdot p_0(t) = \frac{m^*_{AB}}{1 + \lambda \cdot m^*_{AB}} \quad \text{for} \quad t = m^*_{AB}. \tag{14}$$

3. Quality of service parameters and main measures of effectiveness for a loss tandem with blocking

The procedures for calculating quality of service (QoS) parameters and basic measures of effectiveness use the steady-state probabilities in the following manner:

1. Loss probability q_{loss}:

$$q_{loss} = \sum_{j=0}^{c+m} q_{N,j} + \sum_{j=c+m+1}^{c+m+N} q_{c+m+N-j,j} \tag{15}$$

2. Blocking probability q_{bl}:

$$q_{bl} = \sum_{i=0}^{N-1} \sum_{j=c+m+1}^{c+m+N-i} q_{i,j} \tag{16}$$

3. Idle probability q_{idle}:

$$q_{idle} = q_{0,0} \tag{17}$$

4. The average number of blocked lines (tasks) in station A:

$$n_{bl} = \sum_{i=0}^{N-1} \sum_{j=c+m+1}^{c+m+N-i} (j-c-m) \cdot q_{i,j} \tag{18}$$

5. The average number of active (non-blocked) tasks in station A:

$$l_A = \sum_{i=1}^{N} \sum_{j=0}^{c+m} i \cdot q_{i,j} + \sum_{i=1}^{N-1} \sum_{j=c+m+1}^{c+m+N-i} i \cdot q_{i,j} \tag{19}$$

6. The average number of tasks in station A:

$$n_A = \sum_{i=1}^{N} \sum_{j=0}^{c+m} i \cdot q_{i,j} + \sum_{i=1}^{N-1} \sum_{j=c+m+1}^{c+m+N-i} (i+j-c-m) \cdot q_{i,j} \tag{20}$$

7. The average number of tasks in the buffer v:

$$v = \sum_{i=0}^{N} \sum_{j=c+1}^{c+m} (j-c) \cdot q_{i,j} + m \cdot \sum_{i=0}^{N-1} \sum_{j=c+m+1}^{c+m+N-i} q_{i,j} \tag{21}$$

8. The average number of tasks in station B (buffer + server):

$$n_B = \sum_{i=0}^{N} \sum_{j=1}^{c+m} j \cdot q_{i,j} + (m+c) \cdot \sum_{i=0}^{N-1} \sum_{j=c+m+1}^{c+m+N-i} q_{i,j} \tag{22}$$

9. The average number of tasks on the service lines in station B:

$$l_B = \sum_{i=0}^{N} \sum_{j=1}^{c} j \cdot q_{i,j} + c \cdot \sum_{i=0}^{N} \sum_{j=c+1}^{c+m} q_{i,j} + c \cdot \sum_{i=0}^{N-1} \sum_{j=c+m+1}^{c+m+N-i} q_{i,j} \tag{23}$$

10. The mean blocking time in station A:

$$t_{bl} = \frac{n_{bl}}{c \cdot \mu^B} \tag{24}$$

11. The mean response time in station A:

$$q_A = \frac{1}{\mu^A} + t_{bl} \tag{25}$$

12. The mean waiting time in the buffer:

$$w = \frac{v}{c \cdot \mu^B} \tag{26}$$

13. The mean response time in station B:

$$q_B = w + \frac{1}{\mu^B} \tag{27}$$

14. The average tandem sojourn time:

$$t_{thr} = \frac{1}{\mu^A} + t_{bl} + q_B \tag{28}$$

15. The tandem throughput parameter:

$$thr = \frac{N}{t_{thr}} \tag{29}$$

4. Numerical examples: Erlang-*2* service time distribution in both station

According to the initial assumptions the service time for station A and station B has the Erlang-k distribution with the mean value equal to $m = k/\mu$ and the distribution function given by:

$$F(x) = 1 - \sum_{r=0}^{k-1} \frac{e^{-\mu \cdot x} (\mu \cdot x)^r}{r!} \tag{30}$$

therefore

$$\Phi(x) = 1 - F(x) = \sum_{r=0}^{k-1} \frac{e^{-\mu \cdot x} (\mu \cdot x)^r}{r!}. \tag{31}$$

For Erlang-2 distribution of the service time in stations A and B the function $\Phi(x)$ has the following form:

$$\Phi_A(x) = e^{-\mu^A \cdot x} \sum_{r=0}^{1} \frac{(\mu^A \cdot x)^r}{r!} = e^{-\mu^A \cdot x} \cdot (1 + \mu^A \cdot x) \tag{32}$$

$$\Phi_B(x) = e^{-\mu^B \cdot x} \sum_{r=0}^{1} \frac{(\mu^B \cdot x)^r}{r!} = e^{-\mu^B \cdot x} \cdot (1 + \mu^B \cdot x) \tag{33}$$

Then, based on expressions (10) and (11) we have:

$$\begin{aligned} m_{AB} &= \int_0^\infty \Phi_A(x) \cdot \Phi_B(x) dx = \\ &= \int_0^\infty e^{-\mu^A \cdot x} \cdot (1 + \mu^A \cdot x) \cdot e^{-\mu^B \cdot x} \cdot (1 + \mu^B \cdot x) dx = \\ &= \int_0^\infty e^{-(\mu^A + \mu^B) \cdot x} (1 + (\mu^A + \mu^B) \cdot x + \mu^A \cdot \mu^B \cdot x^2) dx = \\ &= \int_0^\infty e^{-(\mu^A + \mu^B) \cdot x} dx + (\mu^A + \mu^B) \int_0^\infty x \cdot e^{-(\mu^A + \mu^B) \cdot x} dx + \\ &+ \mu^A \cdot \mu^B \int_0^\infty x^2 \cdot e^{-(\mu^A + \mu^B) \cdot x} dx = \\ &= \frac{2}{\mu^A + \mu^B} + \frac{2 \cdot \mu^A \cdot \mu^B}{(\mu^A + \mu^B)^3} \end{aligned} \tag{34}$$

The general solution of the above integrals has the following form:

$$\int_0^\infty x^{n-1} e^{-q \cdot x} dx = \frac{1}{q^n} \cdot (n-1)! \tag{35}$$

According to the two-dimensional state diagram for a semi-Markov model, now we must calculate the adequate service rates in stations A and B, indicated here as $\mu^A_{\exp}$ and $\mu^B_{\exp}$, in the equivalent model with exponential distributed service times (see Fig. 2 and Fig. 3), thus:

$$\mu^A_{\exp} = \frac{\mu^A}{2} \qquad \mu^B_{\exp} = \frac{\mu^B}{2} \tag{36}$$

next from these formulas, we may calculate the service rates μ_j^B and μ_i^A according to algorithms given in (1) and (2) expressions:

$$\mu_j^B = \mu_{\exp}^B \cdot j, \qquad \text{for} \quad j = 0, 1, 2, \ldots, c,$$

$$\mu_j^B = \mu_{\exp}^B \cdot c, \qquad \text{for} \quad j = c+1, c+2, \ldots, c+m+N$$

and the service rate μ_i^A for all states without blocking is:

$$\mu_i^A = \mu_{\exp}^A \cdot i \qquad \text{for} \quad i = 0, 1, 2, \ldots, N \tag{37}$$

for all states with blocking:

$$\mu_i^A = \mu_{\exp}^A \cdot i + (j - c - m) \cdot \mu_{\exp}^B$$

for

$$j = c+m+1, \ldots, c+m+N, \quad i = 0, 1, \ldots, N+c+m-j$$

Based on the state diagrams from Fig. 2 and Fig. 3, the mean sojourn time m_{AB} and the mean simultaneity service time m_{AB}^* for each state without blocking can be calculated according to the following formulae:

$$\begin{aligned}
&m_{0,0} = \frac{1}{\lambda},\\
&m_{N,0} = \frac{1}{\mu_N^A},\\
&m_{i,j} = \frac{2}{\mu^A + \mu^B} + \frac{2 \cdot \mu_i^A \cdot \mu_j^B}{(\mu_i^A + \mu_j^B)^3}\\
&\text{for} \quad i = N, j = 1, 2, \ldots, c+m,\\
&m_{0,j}^* = \frac{1}{\mu_j^B}\\
&\text{for} \quad j = 1, 2, \ldots, c+m,\\
&m_{i,0}^* = \frac{1}{\mu_i^A} \quad \text{for} \quad i = 1, 2, \ldots, N-1,\\
&m_{i,j}^* = \frac{2}{\mu^A + \mu^B} + \frac{2 \cdot \mu_i^A \cdot \mu_j^B}{(\mu_i^A + \mu_j^B)^3}\\
&\text{for} \quad i = 1, 2, \ldots, N-1, \;\; j = 1, 2, \ldots, c+m
\end{aligned} \tag{38}$$

and for the states with blocking (here the service rates μ_i^A are calculated differently):

$$m_{0,c+m+N} = \frac{1}{\mu_{c+m+N}^B},$$

$$m_{i,j} = \frac{2}{\mu^A + \mu^B} + \frac{2 \cdot \mu_i^A \cdot \mu_j^B}{(\mu_i^A + \mu_j^B)^3}$$

$$\text{for} \quad j = c+m+1, \ldots, c+m+N-1, \quad i = N+c+m-j,$$

$$m_{0,j}^* = \frac{1}{\mu_j^B}$$

$$\text{for} \quad j = c+m+1, \ldots, c+m+N-1,$$

$$m_{i,j}^* = \frac{2}{\mu^A + \mu^B} + \frac{2 \cdot \mu_i^A \cdot \mu_j^B}{(\mu_i^A + \mu_j^B)^3}$$

$$\text{for} \quad j = c+m+1, \ldots, c+m+N-2, \quad i = 1, \ldots, N+c+m-j-1.$$

In this set of expressions, all $m_{i,j}^*$ can be transformed to $m_{i,j}$ (mean sojourn time) directly, by using relation (14).

In this section, to demonstrate the analysis of loss tandem with blocking, the following configuration is chosen: $N = 18$, $c = 6$, $m = 6$. The service rates in stations A and B are equal to: $\mu^A = 1.0$ (the Erlang-2 distributed service time) and $\mu^B = 1.6$ (the Erlang-2 distributed service time). The inter-arrival rate to the tandem changes within a range from 0.25 to 10.0 (for studying of a model with different utilizations). This model has 418 states, 247 states are without blocking and 171 states are with blocking.

For the model above, the following results were obtained, majority of which are presented in Fig. 4, Fig. 5 and Table 1.

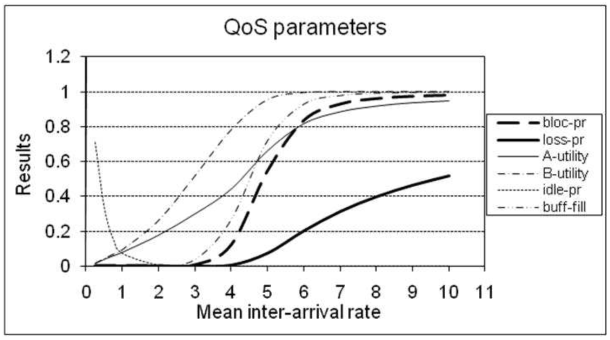

Fig. 4. Graphs of QoS parameters, where *bloc-pr* is the blocking probability, *loss-pr* is the loss probability, *A-utilization* is the station A utilization factor, *B-utilization* is the station B utilization factor, *idle-pr* is the idle probability, *buff-fill* is the buffer filling parameter

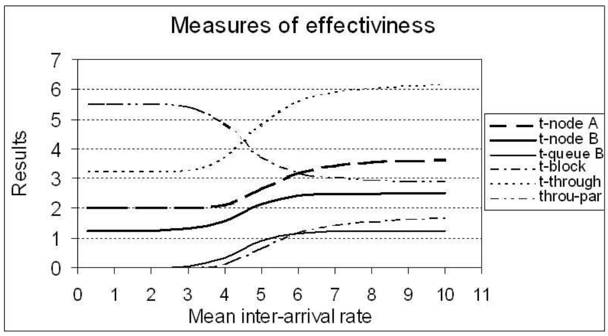

Fig. 5. The measures related to the mean time: *t-node A* is the mean response time in station A, *t-node B* is the mean response time in station B, *t-queue B* is the mean waiting time in the buffer, *t-block* is the mean blocking time, *t-through* is the average tandem throughput time, *throu-par* tandem throughput parameter

Table 1
The comparison of the average number of tasks in both stations

λ	The average number of tasks						Station A utilization	Station B utilization
	n_{bl}	v_B	l_A	l_B	n_A	n_B		
1.0	0.000	0.000	1.783	1.106	1.783	1.106	0.111	0.208
2.0	0.000	0.025	3.784	2.348	3.794	2.373	0.222	0.417
3.0	0.019	0.312	5.780	3.600	5.799	3.912	0.335	0.625
4.0	0.589	1.764	7.744	4.884	8.332	6.647	0.472	0.828
5.0	3.583	4.472	9.037	5.768	12.619	10.241	0.678	0.961
6.0	6.298	5.640	9.272	5.969	15.569	11.609	0.823	0.994
7.0	7.464	5.886	9.260	5.994	16.722	11.880	0.888	0.999
8.0	7.987	5.946	9.233	5.998	17.218	11.945	0.920	1.000
9.0	8.261	5.967	9.212	5.999	17.471	11.966	0.938	1.000
10.0	8.423	5.975	9.196	5.999	17.617	11.975	0.950	1.000

The results of this series of experiments show that the loss probability, blocking probability (quality of service parameters), filling buffer parameters or mean blocking times rapidly grow when the stream intensity to the tandem increases. In this case the number of blocked lines in the first station quickly grows simultaneously when the buffer is filled quickly and the number of rejected tasks quickly increase. Most of the average time measures behave similarly, if stream intensity to the tandem grows. In this situation the tandem throughput parameter falls down, because the blocking and response times are increased. All these negative factors depend on the second station utilization parameter. The tandem works properly when the second station utilization parameter is less than 0.75. In the moderate utilization interval, the tandem works properly and most of quality of service (QoS) parameter is easy to keep at the appropriate level.

5. Conclusions

In this paper, the mathematical model of a loss two-station stochastic transition system with blocking and rejection, treated as a semi-Markov process, is presented. In this tandem model a phenomenon of blocking and rejection appears simultaneously and the mathematical procedures allow for calculating the main measures of effectiveness including the loss and blocking probabilities. These measures may be calculated for any tandem configuration if we have given service rates in both stations and when the inter-arrival rate to the tandem is given.

The results of experiments presented in Sec. 4 show that depending on the model configuration and its characteristics the mathematical modelling allows finding the proper rate range for input stream that guarantees congestion avoidance in the tandem. In the opposite case if one is given an input stream rate, the analysis allows taking another model characteristic, which guarantees that blocking and loss probabilities should be in the proper range.

Acknowledgements. This work is supported by the Bialystok University of Technology under the W/WI/7/03 grant.

REFERENCES

[1] S. Balsamo, V. de Nito Persone, and P. Inverardi, "A review on queueing network models with finite capacity queues for software architectures performance predication", *Performance Evaluation* 51 (2–4), 269–288 (2003).

[2] A. Badrah, T. Czachórski, J. Domańska, J.-M. Fourneau, and F. Quessette, "Performance evaluation of multistage interconnection networks with blocking – discrete and continuous time Markov models", *Archives of Theoretical and Applied Informatics* 14 (2), 145–162 (2002), (in Polish).

[3] A. Economou and D. Fakinos, "Product form stationary distributions for queueing networks with blocking and rerouting", *Queueing Systems* 30 (3/4), 251–260 (1998).

[4] J.S. Kaufman and K.M. Rege, "Blocking in a shared resource environment with batched arrival processes", *Performance Evaluation* 24, 249–263 (1996).

[5] J.B. Martin, "Large tandem queueing networks with blocking", *Queueing Systems* 41 (1/2), 45–72 (2002).

[6] W. Oniszczuk, "Tandem models with blocking in the computer subnetworks performance analysis", in *Biometrics, Computer Security Systems and Artificial Intelligence Applications*, pp. 259–267, Springer, New York, 2006.

[7] M. Sereno, "Mean value analysis of product form solution queueing networks with repetitive service blocking", *Performance Evaluation* 36–37, 19–33 (1999).

[8] L. Zhuang, "Acceptance instant distributions in product-form closed queueing networks with blocking", *Performance Evaluation* 26, 133–144 (1996).

[9] I.F. Akyildiz, "Mean value analysis for blocking queuing Networks", *IEEE Transaction on Software Engineering* 14 (4), 418-428 (1988).

[10] S. Balsamo and V. de Nitto Persone, "A survey of product form queueing networks with blocking and their equivalences", *Annals of Operations research* 48 (1/4), 31–61 (1994).

[11] R.J. Boucherie and N.M. van Dijk, "On the arrival theorem for product form queueing networks with blocking", *Performance Evaluation* 29 (3), 155–176 (1997).

[12] M.C. Clo, "MVA for product-form cyclic queueing networks with blocking", *Annals of Operations Research* 79, 83–96 (1998).

[13] D. Kouvatsos and J. Almond, "Maximum entropy two-station cyclic queues with multiple general servers", *Acta Informatica* 26, 241–267 (1988).

[14] J.A. Morrison, "Blocking probabilities for multiple class batched arrivals to a shared resource", *Performance Evaluation* 25, 131–150 (1996).

[15] R. Onvural, "Survey of closed queuing networks with blocking", *Computer Survey* 22 (2), 83–121 (1990).

[16] V. Sharma and J.T. Virtamo, "A finite buffer queue with priorities", *Performance Evaluation* 47, 1–22 (2002).

[17] H.G. Perros, *Queuing Networks with Blocking. Exact and Approximate Solution*, Oxford University Press, Oxford, 1994.

[18] S. Balsamo, V. de Nito Persone, and R. Onvural, *Analysis of Queueing Networks with Blocking*, Kluwer Academic Publishers, Boston, 2001.

[19] A. Bouhchouch, Y. Frein, and Y. Dallery, "Performance evaluation of closed tandem queueing networks", *Performance Evaluation* 26, 115–132 (1996).

[20] A. Gomez-Corral, "A tandem queue with blocking and Markovian arrival process", *Queueing Systems* 41 (4), 343–370 (2002).

[21] J.Ch. Strelen, B. Bärk, J. Becker, and V. Jonas, "Analysis of queueing networks with blocking using a new aggregation technique", *Annals of Operations Research* 79, 121-142 (1998).

[22] D.D. Kouvatsos, I.U. Awan, R.J. Fretwell, and G. Dimakopoulos, "A cost-effective approximation for SRD traffic in arbitrary multi-buffered networks", *Computer Networks* 34, 97–113 (2000).

[23] S. Ramesh and H.G. Perros, "A two-level queueing network model with blocking and non-blocking messages", *Annals of Operations Research* 93 (1/4), 357–372 (2000).

[24] T. Tolio and S.B. Gershwin, "Throughput estimation in cyclic queueing networks with blocking", *Annals of Operations Research* 79, 207–229 (1998).

[25] L. Zhuang, J.A. Buzacott, and X-G. Liu, "Approximate mean value performance analysis of cyclic queueing networks with production blocking", *Queueing Systems* 16, 139–165 (1994).

[26] V.S. Korolyuk and V.V. Korolyuk, *Stochastic Models of Systems*, Kluwer Academic Publishers, Dordrecht, 1999.