Service Strategies in Two-Server Tandem Configurations – Modeling and Investigation

Walenty Oniszczuk

Service strategies in two-server tandem configurations – modeling and investigation

Walenty Oniszczuk

Faculty of Computer Science, Bialystok University of Technology, Bialystok, Poland

w.oniszczuk@pb.edu.pl

Abstract. In this document, two different service strategies in an Internet service station, which consists of with two separated servers (tandem), are investigated. At beginning the Internet tandem under study is first formulated as an open three-station queuing network with blocking assuming that there are finite capacity buffers at front of each server. For each Internet station configurations, the two-dimensional state graphs ware constructed and set of steady-state equations are created. These equations allow for calculating the state probabilities vectors for each tandem configurations, using an equation solving techniques. Many useful performance measures regarding the tandem configurations under study can be extracted from the steady-state probability vectors. The numerical part of this paper contains example for the investigation of different tandem configurations, where are presented the results of the calculation of the main measures of effectiveness and quality of service (QoS) parameters.

Keywords: Internet tandem configurations, open queuing network, blocking probabilities, performance measures

1 Introduction

A traditional Internet server processes requests in FIFO manner. During a high load period, each task has to wait in a queue for a long time before getting serviced. Overhead from tasks competing for limited resources, such as open connections and network bandwidth, is increased. Retry from impatient clients worsens the load situation and causes the snowball effect. More elaborate resource allocation schemas, rather than best-effort service model, need to be adopted to provide predictable services during high load periods. The existing best-effort service with the FIFO scheduling mode and dropping tasks (requests) if the queue is full of the Internet servers leads to misallocation of scarce and expensive network and CPU resources during heavy load periods, causing unpredictable response delay.

The next generation Internet will demand differentiated services from Internet servers. The most important function of the future Internet is to support the provisioning of reliable real-time services on a wide scale. In order to achieve this goal, per-aggregate resource management is, nowadays, regarded as a mandatory choice. The

classical example of the modern approach to service users tasks is divided they service processes onto several separated servers (for example, as mechanisms for resolving some security problems), which processes some independent parts of incoming tasks. In this paper, we analyze two different tandem configurations, which consists of with two separated servers, in specific way connected each other. At each server, FIFO multiplexing is in place, meaning that all tasks traversing the server are buffered in a single queue First-Come-First-Served. Nowadays the tandem networks have been studied extensively and applied in the evaluation of various systems as in design, capacity planning and performance evaluation of computer and communication systems, call centers, flexible manufacturing systems, etc. Some examples of their application in real systems (two transmitter communication networks with Dynamic Bandwidth Allocation, service facility with front and back room operations) can be found in [28], [1] respectively. The behavior of various systems, including communication and computer systems, as well as production and manufacturing procedures, can be represented and analyzed through queuing network models to evaluate their performance [11,12,13], [29,30]. System performance analysis usually includes the queue length distribution and various performance indicators such as response time, throughput and utilization [2], [6, 7], [10].

The theory behind tandem queues is well developed, see, e.g. [3, 4, 5], [14], [26, 27]. However, there is still a great interest around more complicated setups involving blocking phenomena as well as different mechanisms for offering services. An excellent survey may be found in the well known of Perros [26] and Balsamo [3] books. Over the years high quality research has appeared in diverse journals and conference proceeding in the field of computer science, traffic engineering and communication engineering [4], [14], and [26]. In particular, the two-node tandem queuing model with the Batch Markovian Arrival Process input flow and non-exponential service time distribution described in the paper [8]. Additionally, systems with finite capacity queues under various blocking mechanisms and scheduling constraints are analyzed by the author in [15-25]. In [15, 16], the closed type, multi-center computer networks with different blocking strategies are investigated and measures of effectiveness based on Quality of Service (QoS) are studied. Markovian and semi-Markovian approaches for analysis of open tandem networks with blocking are presented in [17], [19], [21], [23] [24] and [25]. Some two-stage tandem queues with blocking and an optional feedback are presented in [20] and [22]. In such systems, feedback is the likelihood of a task return, with fixed probability to the first server of the tandem immediately after the service at the second one [8]. Tandems with feedback are usually more complex than the ones without and they are mostly investigated given stationary Poisson arrival process and exponential service time distribution [1], [28]. Blocking and dead-locking phenomena in an open series, linked network model with HOL (head-of-line) priority feedback service was investigated and presented by the author in [18].

The remainder of this paper is organized as follows: Section 2 presents and explains the models specification and description. Section 3, analyzes a tandem as an open linked series three-station network with blocking. Section 4 explains a tandem as a linked rerouting two-server network the same with blocking. Section 5 describes the numerical results obtained using our solution technique, followed by the concluding remarks in Section 6.

2   Models specification and description

In this research, two different configuration of Internet tandem is presented. Each of these kinds of tandem networks has a single service line at the main server, and the other (additional) server with a single service line. Between these servers is a common waiting buffer with finite capacity, for example equal to *m2*. When this buffer is full, the accumulation of new tasks from the main server is temporarily suspended and a phenomenon called blocking occurs until the queue empties and allows new inserts. Similarly, if the first buffer (with capacity *m1*) ahead of the main server is full, then the Internet source node (station) is blocked. This is the classical mechanism for controlling the intensity of an arriving task stream, which comes from the Internet users to the servicing tandem. In this kind of network configuration, no more than *m1+m2+2* tasks can be processed simultaneously and the Internet tandem network becomes idle, if there are no tasks in both servers.

Let us consider the Internet two-server tandems with blocking, both configurations, as shown in Figure 1 and Figure 2. On these figures we present models of generalized Internet tandem networks and analyze the feasibility of service with blocking and rerouting. These networks contain of four major logical components:

1. Main server, for the first or the first and the second stage tasks processing with a task initiator, a task controller (realization ON-OFF strategy for incoming requests) and a task dispatcher.
2. Second sever – for the second stage tasks processing.
3. Communication channels.
4. Internet source and λ represents the request arrival process from the clients.

After processing the responses are sent back to the clients through the communication channels. To simplify the model, we assume that the tandem of servers connect to the client through a high speed network. We have ignored the channel delay and flow control on a client side as they are beyond the scope of the proposed study.

The input task stream comes from the Internet source to the main server. This server has a finite capacity buffer and it can accept only *m1+1* tasks. A new task, which arrives at the full main server buffer, is forced to wait in the Internet source station and blocks it. For linked in series servers configuration of tandem (see Fig. 1), each task at the main server is processed on the service line and upon service completion sent to the second server. If there is a free service line on this server, the service process starts immediately, if not, the task must wait in the buffer. If the buffer is full, any task upon service completion at the main server is forced to wait and blocks this server. In case of tandem with rerouting configuration (see Fig. 2), after service completion at the main server, the task proceeds to the second server with probability *1- σ*, and with probability σ the task departs from tandem. Tasks leaving the second server are always feed back to the main server. If the main server buffer is full, similarly as in the previous configuration, the tasks block the second server.

The general assumptions for these two-server tandem models are:

1. Internet tasks stream arriving to the main server is assumed to be a Poisson stream, with rate $\lambda = 1/a$, where *a* is the mean inter-arrival time,
2. a single service line is on the main server,
3. a single service line is available on the second server,

4. in both servers the service time represents exponentially distributed random variables, with mean $s^A = 1/\mu^A$ and $s^B = 1/\mu^B$, where μ is the mean service rate,
5. the buffers capacities are finite, for example equal to *m1* and *m2*, respectively to the main and the second servers,
6. service strategies restrict and forbid tasks truncations, if the buffers are full in any case (**strategy network drop-tail is forbidden**) and tasks lose

In this special type of multi-stage network with blocking a deadlock may occur. We assume that a deadlock is detected instantaneously and resolved without any delay time by simultaneously exchanging both blocked tasks [25].
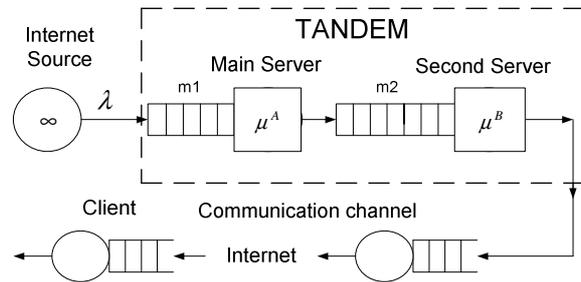

Fig. 1. Tandem configuration as open linked series servers.

Generally, blocking phenomena is a very important mechanism for controlling and regulating intensity of arriving tasks from the Internet source (users) to the Internet tandem servers. A blocking strategies in the main server are realized by a Controller mechanism, it means that controller temporarily suspend and resume (ON-OFF strategy) transfer process from users, because "**network drop-tail strategy is forbidden**" must be realized. The arrival rate to the main server depends on the state of the tandem and blocking factor that reduces the rate at which users are sending tasks to the tandems.
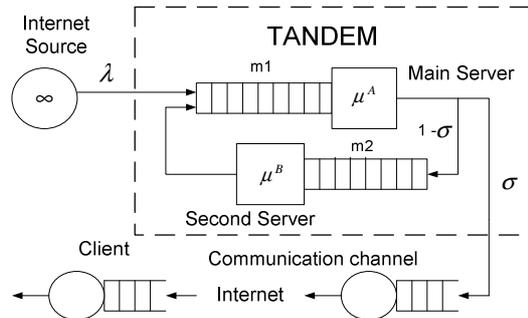

Fig. 2. Tandem configuration as rerouting networks.

3　　Tandem as an open linked series servers

Let us consider the Internet two-server tandem shown in figure 1 as the three-station queuing network with blocking. Each queuing system can, in principle, be mapped onto an instance of a Markov process and then mathematically evaluated in terms of

this process. According to general assumptions, a continuous-time homogeneous Markov chain can represent a tandem network. The queuing network model reaches a steady-state condition and the underlying Markov chain has a stationary state distribution. Also, such queuing network with finite capacity queues has finite state space. The solution of the Markov chain representation may then be computed and the desired performance characteristics, as queue length distribution, utilizations, and throughputs, obtained directly from the stationary probability vector.

In theory, any Markov model can be solved numerically. In particular, solution algorithm for Markov queuing networks with blocking is a five-step procedure:

1. Definition of the series network state space (choosing a state space representation).
2. Enumerating all the transitions that can possible occur among the states.
3. Definition of the transition rate matrix Q that describes the network evaluation (generating the transition rate).
4. Solution of linear system of the global balance equations to derive the stationary state distribution vector (computing appropriate probability vector).
5. Computation from the probability vector of the average performance indices.

In this type of an open series network we may denote its state by the pair (i,j), where i represents the number of tasks in main server and j denotes the number in second server, including all tasks in service and in blocking. Here some blocked tasks, physically, are located on the Internet source station or on the main server, but the nature of the service process in both servers, allows one to treat them as located in additional places in the buffers and they belong to the main server or the second server. In this case, there can be a maximum of $m1+2$ tasks assigned to the main server including the task in the Internet source that can be blocked. Similarly, there can be a maximum of $m2+2$ tasks assigned to second server with a task blocked in the main server. For any non-negative integer values of i and j, (i,j) represent a feasible state of this queuing network, and $p_{i,j}$ denotes the probability for that state in equilibrium. These states and the possible transitions among them are shown in Figure 3. This state diagram of the series server network contains all possible non-blocked states (marked by ovals) as well as the blocking states (marked by rectangles).

Based on an analysis the state space diagrams, the process of constructing the corresponding differential-difference equations, differential equations in t, and difference in (i,j) state, can be divided into several independent steps, which describe some similar, repeatable schemas (see Fig. 3).

These equations are:

a) for states without blocking factor:

$$p'_{0,0}(t) = - \lambda \cdot p_{0,0}(t) + \mu^B \cdot p_{0,1}(t)$$

$$p'_{0,j}(t) = - (\lambda + \mu^B) \cdot p_{0,j}(t) + \mu^A \cdot p_{1,j-1}(t) + \mu^B \cdot p_{0,j+1}(t) \qquad \text{for } j = 1, \dots , m2+1$$

$$p'_{i,0}(t) = - (\lambda + \mu^A) \cdot p_{i,0}(t) + \lambda \cdot p_{i-1,0}(t) + \mu^B \cdot p_{i,1}(t) \qquad \text{for } i = 1, \dots , m1+1$$

$$p'_{i,j}(t) = - (\lambda + \mu^B + \mu^A) \cdot p_{i,j}(t) + \lambda \cdot p_{i-1,j}(t) + \mu^A \cdot p_{i+1,j-1}(t) + \mu^B \cdot p_{i,j+1}(t) \qquad (1)$$
$$\text{for } i = 1, \dots , m1+1, \; j = 1, \dots , m2+1$$

b) for states with Main Server blocking:

$$p'_{0,m2+2}(t) = - (\lambda + \mu^B) \cdot p_{0,m2+2}(t) + \mu^A \cdot p_{1,m2+1}(t)$$

$$p'_{i,m2+2}(t) = - (\lambda + \mu^B) \cdot p_{i,m2+2}(t) + \lambda \cdot p_{i-1,m2+2}(t) + \mu^A \cdot p_{i+1,m2+1}(t) \qquad (2)$$
$$\text{for } i = 1, \dots , m1$$

c) for states with both Internet source and Main Server simultaneous blocking:

$$p'_{m1+1,m2+2}(t) = - \mu^B \cdot p_{m1+1,m2+2}(t) + \lambda \cdot p_{m1,m2+2}(t) + \mu^A \cdot p_{m1+2,m2+1}(t) \qquad (3)$$

d) for states with Internet source blocking:

$$p^{'}_{m1+2,0}(t) = - \mu^A \cdot p_{m1+2,0}(t) + \lambda \cdot p_{m1+1,0}(t) + \mu^B \cdot p_{m1+2,1}(t)$$

$$p^{'}_{m1+2,j}(t) = - (\mu^A + \mu^B) \cdot p_{m1+2,j}(t) + \lambda \cdot p_{m1+1,j}(t) + \mu^B \cdot p_{m1+2,j+1}(t) \qquad (4)$$

$$\text{for } j = 1, \dots , m2$$

$$p^{'}_{m1+2,m2+1}(t) = - (\mu^A + \mu^B) \cdot p_{m1+2,m2+1}(t) + \lambda \cdot p_{m1+1,m2+1}(t)$$

The solution for the equilibrium states (or the stationary states), if it exists, must satisfy:

$$\lim_{t \to \infty} p^{'}_{i,j}(t) = 0 \qquad (5)$$

and if we let

$$p_{i,j} = \lim_{t \to \infty} p^{'}_{i,j}(t) \qquad (6)$$

this leads to the set of equilibrium equations.

Here, a queuing network with blocking linked in series, under appropriate assumptions, is formulated as a Markov process and the stationary probability vector can be obtained using numerical methods for linear systems of equations. If there is a model network with finite number of states, its steady-state probabilities can be found directly from equations (1) - (6) by using some iteration method and the normalizing condition for the sum of state probabilities.
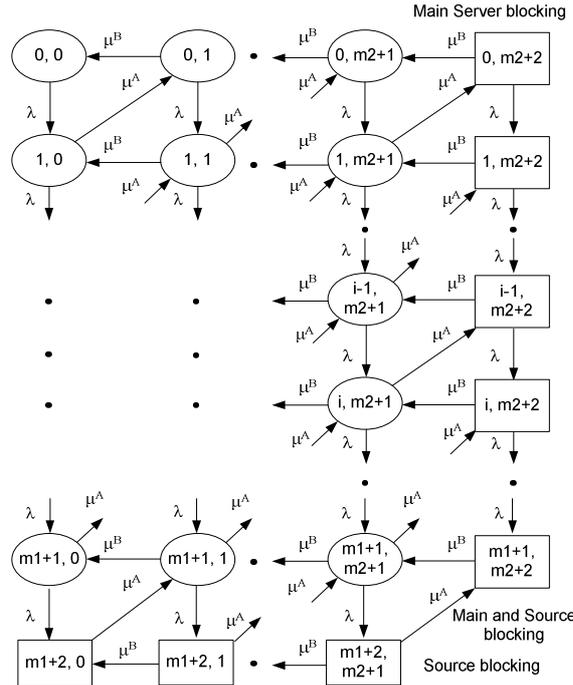

Fig. 3. State transmission diagram for linked series servers

Some specialized software for the solution of nonsymmetrical linear systems of equations by using iterative methods was created by the author. The package is written entirely in the C programming language, and data structures are managed dynami-

cally. This package allows efficiently calculate the steady-state probability vectors in Markovian models and automatically generates more than twenty different performance measures for analyzed linked series queuing network.

4 Tandem as rerouting networks

The same, as in the Section 3, let us consider the Internet two-server tandem shown in figure 2 as the three-station queuing network with blocking. The state diagram of this rerouting network is presented on the Figure 4. Here are shown all possible non-blocked states (marked by ovals) as well as the blocking states (marked by rectangles). Based on an analysis this state space diagrams, the process of constructing the corresponding differential-difference equations, differential equations in t, and difference in (i,j) state, can be divided into several independent steps, which describe some similar, repeatable schemas.

These equations are:

a) for states without blocking factor:

$$p'_{0,0}(t) = - \lambda \cdot p_{0,0}(t) + \mu^A\sigma \cdot p_{1,0}(t)$$

$$p'_{0,j}(t) = - (\lambda + \mu^B) \cdot p_{0,j}(t) + \mu^A(1-\sigma) \cdot p_{1,j-1}(t) + \mu^A\sigma \cdot p_{1,j}(t) \qquad \text{for } j = 1, \dots , m2+1$$

$$p'_{i,0}(t) = - (\lambda + \mu^A\sigma + \mu^A(1-\sigma)) \cdot p_{i,0}(t) + \lambda \cdot p_{i-1,0}(t) + \mu^B \cdot p_{i-1,1}(t) + \mu^A\sigma \cdot p_{i+1,0}(t)$$
$$\text{for } i = 1, \dots , m1$$

$$p'_{i,j}(t) = - (\lambda + \mu^A\sigma + \mu^A(1-\sigma) + \mu^B) \cdot p_{i,j}(t) + \lambda \cdot p_{i-1,j}(t) + \mu^B \cdot p_{i-1,j+1}(t) +$$
$$+ \mu^A\sigma \cdot p_{i+1,j}(t) + \mu^A(1-\sigma) \cdot p_{i+1,j-1}(t) \qquad \text{for } i = 1, \dots , m1, \; j = 1, \dots , m2+1$$

$$p'_{m1+1,0}(t) = - (\lambda + \mu^A\sigma + \mu^A(1-\sigma)) \cdot p_{m1+1,0}(t) + \lambda \cdot p_{m1,0}(t) + \mu^B \cdot p_{m1,1}(t) +$$
$$+ \mu^A\sigma \cdot p_{m1+2,0}(t) + \mu^A\sigma \cdot p_{m1+3,0}(t) \qquad\qquad (7)$$

$$p'_{m1+1,j}(t) = - (\lambda + \mu^A\sigma + \mu^A(1-\sigma) + \mu^B) \cdot p_{m1+1,j}(t) + \lambda \cdot p_{m1,j}(t) + \mu^B \cdot p_{m1,j+1}(t) +$$
$$+ \mu^A\sigma \cdot p_{m1+2,j}(t) + \mu^A\sigma \cdot p_{m1+3,j}(t) + \mu^A(1-\sigma) \cdot p_{m1+3,j-1}(t) \qquad \text{for } j = 1, \dots , m2$$

$$p'_{m1+1,m2+1}(t) = - (\lambda + \mu^A\sigma + \mu^A(1-\sigma) + \mu^B) \cdot p_{m1+1,m2+1}(t) + \lambda \cdot p_{m1,m2+1}(t) +$$
$$+ \mu^B \cdot p_{m1,m2+2}(t) + \mu^A\sigma \cdot p_{m1+2,m2+1}(t) + \mu^A(1-\sigma) \cdot p_{m1+3,m2}(t)$$

b) for states with Main Server blocking:

$$p'_{0,m2+2}(t) = - (\lambda + \mu^B) \cdot p_{0,m2+2}(t) + \mu^A(1-\sigma) \cdot p_{1,m2+1}(t) \qquad\qquad (8)$$

$$p'_{i,m2+2}(t) = - (\lambda + \mu^B) \cdot p_{i,m2+2}(t) + \lambda \cdot p_{i-1,m2+2}(t) + \mu^A(1-\sigma) \cdot p_{i+1,m2+1}(t)$$
$$\text{for } i = 1, \dots , m1$$

c) for states with both Internet source and Main Server simultaneous blocking:

$$p'_{m1+1,m2+2}(t) = - \mu^B \cdot p_{m1+1,m2+2}(t) + \lambda \cdot p_{m1,m2+2}(t) + \mu^A(1-\sigma) \cdot p_{m1+2,m2+1}(t) \qquad (9)$$

d) for states with Source blocking:

$$p'_{m1+2,j}(t) = - (\mu^A\sigma + \mu^A(1-\sigma)) \cdot p_{m1+2,j}(t) + \lambda \cdot p_{m1+1,j}(t) \qquad \text{for } j = 0, \dots , m2$$

$$p'_{m1+2,m2+1}(t) = - (\mu^A\sigma + \mu^A(1-\sigma)) \cdot p_{m1+2,m2+1}(t) + \lambda \cdot p_{m1+1,m2+1}(t) + \qquad (10)$$
$$+ \mu^B \cdot p_{m1+1,m2+2}(t)$$

e) for states with Second Server blocking:

$$p'_{m1+3,j}(t) = - (\mu^A\sigma + \mu^A(1-\sigma)) \cdot p_{m1+3,j}(t) + \mu^B \cdot p_{m1+1,j+1}(t) \text{ for } j = 0, \dots , m2 \qquad (11)$$

The solution of these equations for the equilibrium states, if it exists, must satisfy formulas (5) and (6) and this leads to the set of stationary equations. In this case a queuing rerouting network with blocking is formulated as a Markov process and the stationary probability vector can be obtained using numerical methods for linear systems of equations using equations (7) - (11) with (5) and (6) and dedicated software for the solution of linear systems of equations was created by the author. This soft-

ware allows efficiently calculate the steady-state probability vectors and automatically generates more than twenty different performance measures for analyzed queuing network.
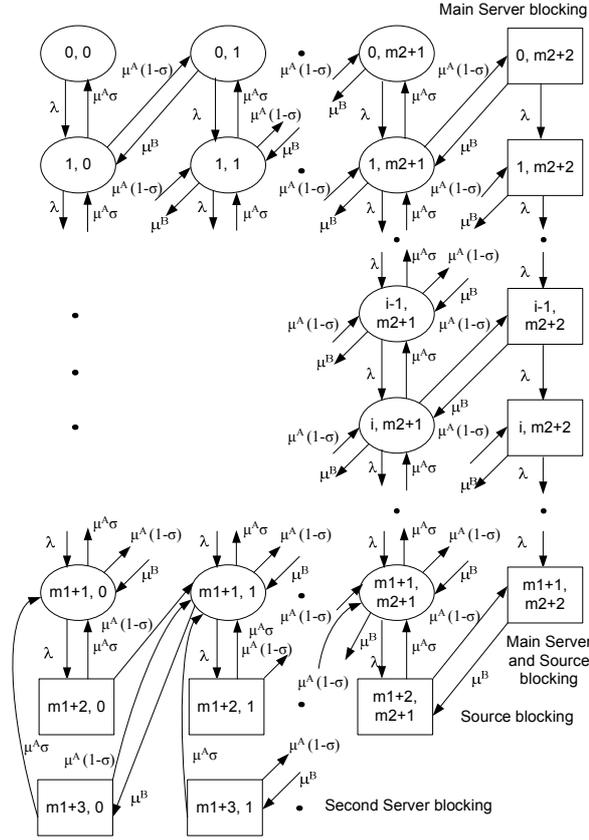

Fig. 4. State transmission diagram for rerouting network

5   Numerical Results

In this section, to demonstrate our analysis of two service strategies in the different Internet two-serves tandem configurations presented in Sections 3 and Section 4, we have performed numerous calculations. We will concentrate on the several important performance descriptors, such as the probabilities that the different tandem configurations are blocked, it means how are blocked separated servers or blocked Internet sources or how are changed various time measures and how are changed buffers filling parameters if the inter-arrival rate λ from the Internet source (users) to tandems are changed within a range from *1.0* to *10.0* per time unit. To demonstrate this, the following tandem parameters were chosen: the service rates in main server and second server are equal to: $\mu^A = 8.0$, $\mu^B = 5.0$ respectively. The buffer capacities are chosen as equal to: *m1 = 40, m2 = 30*. The depart probability σ for rerouting tandem

configuration is chosen as *0.5* and in this configuration the service rates for main server is chosen as equal to: $\mu^A = 16.0$ per time unit. This is necessary for guarantee the same main server utilization parameter, as in the series tandem model. In this case the mean service rate for tandem rerouting model, must be equal to previous service rate divided by depart probability, e.g. μ^A/σ.

Based on such chosen parameters the following results were obtained and the majority of them are presented in Tables 1, 2 and 3. In all of these tables the columns marked as "*italic*" contain the results for the first tandem configuration, e.g. the open linked series servers (first service strategy). The columns marked as "bold" contains the results for the second tandem configuration, e.g. the rerouting tandem (second service strategy).

In the first table, the λ is the inter-arrival rate from the Internet source to the both tandems, *Idle-s* and *Idle-r* are the idle tandem probabilities for series and rerouting tandems, *Bl-M-s* and *Bl-M-r* are the main server blocking probabilities for series and rerouting tandems, *Bl-S-s* and *Bl-S-r* are the source blocking probabilities for series and rerouting tandems.

Table 1. Measures of effectiveness – the probabilities

λ	*Idle-s*	**Idle-r**	*Bl-M-s*	**Bl-M-r**	*Bl-S-s*	**Bl-S-r**
1.0	*0.700*	**0.700**	*0.000*	**0.000**	*0.000*	**0.000**
2.0	*0.450*	**0.450**	*0.000*	**0.000**	*0.000*	**0.000**
3.0	*0.250*	**0.250**	*0.000*	**0.000**	*0.000*	**0.000**
4.0	*0.100*	**0.100**	*0.000*	**0.000**	*0.000*	**0.000**
5.0	*0.005*	**0.011**	*0.219*	**0.095**	*0.014*	**0.005**
6.0	*0.000*	**0.000**	*0.375*	**0.320**	*0.167*	**0.075**
7.0	*0.000*	**0.000**	*0.375*	**0.414**	*0.286*	**0.211**
8.0	*0.000*	**0.000**	*0.375*	**0.464**	*0.375*	**0.362**
9.0	*0.000*	**0.000**	*0.375*	**0.499**	*0.444*	**0.487**
10.0	*0.000*	**0.000**	*0.375*	**0.520**	*0.500*	**0.570**

Table 2. Measures of effectiveness – the responses (the time parameters)

λ-*decl*	λ-*s-eff*	**λ-r-eff**	*w-s*	**w-r**	*ro-s*	**ro-r**
1.0	*1.000*	**1.000**	*0.010*	**0.010**	*0.200*	**0.200**
2.0	*2.000*	**2.000**	*0.053*	**0.053**	*0.400*	**0.400**
3.0	*3.000*	**3.000**	*0.180*	**0.180**	*0.600*	**0.600**
4.0	*4.000*	**4.000**	*0.640*	**0.640**	*0.800*	**0.800**
5.0	*4.992*	**4.942**	*3.331*	**4.512**	*0.972*	**0.986**
6.0	*5.836*	**5.000**	*5.469*	**5.791**	*0.999*	**1.000**
7.0	*6.408*	**5.000**	*5.829*	**5.792**	*1.000*	**1.000**
8.0	*6.774*	**5.000**	*5.928*	**5.792**	*1.000*	**1.000**
9.0	*7.064*	**5.000**	*5.978*	**5.792**	*1.000*	**1.000**
10.0	*7.372*	**5.000**	*6.004*	**5.792**	*1.000*	**1.000**

In the second table, the λ-*decl* is the declared inter-arrival input stream intensities from the Internet source to the tandems, λ-*s-eff* and λ-*r-eff* are effective input stream intensities for series and rerouting tandems, *w-s* and *w-r* are the mean waiting times

for series and rerouting tandems, *ro-s* and *ro-r* are the second server utilization factors for series and rerouting tandems.

In the next table, the λ is the inter-arrival rate from the Internet source to the tandems, *v-M-s* and *v-M-r* are the number of tasks in the Main server buffer for series and rerouting tandems, *v-S-s* and *v-S-r* are the number of tasks in the Second server buffer for series and rerouting tandems, *ro-s* and *ro-r* are the main server utilization factor for series and rerouting tandems.

Table 3. Measures of effectiveness – the occupation parameters

λ	*v-M-s*	*v-M-r*	*v-S-s*	*v-S-r*	*ro-s*	*ro-r*
1.0	*0.018*	**0.018**	*0.050*	**0.050**	*0.125*	**0.125**
2.0	*0.083*	**0.083**	*0.267*	**0.267**	*0.250*	**0.250**
3.0	*0.225*	**0.225**	*0.900*	**0.900**	*0.375*	**0.375**
4.0	*0.504*	**0.508**	*3.198*	**3.198**	*0.500*	**0.500**
5.0	*12.531*	**4.802**	*22.562*	**16.653**	*0.836*	**0.704**
6.0	*35.948*	**20.372**	*28.955*	**27.289**	*1.000*	**0.932**
7.0	*38.302*	**30.732**	*28.958*	**29.000**	*1.000*	**0.988**
8.0	*39.028*	**35.958**	*28.958*	**29.412**	*1.000*	**0.998**
9.0	*39.363*	**38.353**	*28.958*	**29.611**	*1.000*	**1.000**
10.0	*39.548*	**39.242**	*28.958*	**29.719**	*1.000*	**1.000**

The results of the experiments clearly show that the effect of the properly chosen service strategy in Internet tandem network must be taken into account when analyzing performance such computer networks. Also, results of these calculations evidently show that the blocking phenomena must be taken into account because variation of inter-arrival rate drastically changes the main performace parameters. Which tandem configurations and service strategies are better? It all depends, which measures or time parameters are more preferable during our analysis and obtained results applications.

Conclusions

An approach to compare the effectiveness of two service strategies in Internet servers linked to tandem with blocking has been presented. Tasks blocking probabilities and some other fundamental performance characteristics of such networks are derived, followed by numerical examples. The results confirm importance of a special treatment for the models with blocking, which justifies this research. Moreover, our proposal is useful in designing buffer sizes or channel capacities for a given blocking probability requirement constraint. The results can be used for capacity planning and performance evaluation of real-time computer networks where blocking are present.

References

1. Arivudainambi, D., Poongothai, V.: Analysis of a service facility with cross trained servers and optional feedback. Int. J. Oper. Res. **18**(2), 218-237 (2013)

2. Atencia, I.: A discrete-time system with service control and repairs. Int. J. Appl. Math. Comput. Sci. **24**(3), 471-484 (2014)
3. Balsamo, S., De Nito Persone, V., Onvural, R.: Analysis of Queueing Networks with Blocking. Kluwer Academic Publishers, Boston (2001)
4. Clo, M.C.: MVA for product-form cyclic queueing networks with blocking. Ann. Oper. Res. **79**, 83-96 (1998)
5. Economou, A., Fakinos, D.: Product form stationary distributions for queueing networks with blocking and rerouting. Queueing Syst. **30**(3/4), 251-260 (1998)
6. Gemikonakli, E., Mapp, G., Gemikonakli, O., Ever, E.: Exploring service and buffer management issues to provide integrated voice and data services in single and multi-channel wireless networks. In: 2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA), pp.1056-1063. IEEE Conference Publications (2013). doi:10.1109/AINA.2013.57
7. Itoh, H., Fukumoto, H., Wakuya, H., Furukawa T.: Bottom-up learning of hierarchical models in a class of deterministic POMDP environments. Int. J. Appl. Math. Comput. Sci. **25**(3), 597-615 (2015)
8. Kim, C.S., Klimenok, V., Tsarenkov, G., Breuer, L., Dudin, A.: The BMAP/G/1-> ·/PH/1/M tandem queue with feedback and losses. Perform. Eval. **64**, 802-818 (2007)
9. Korolyuk, V.S., Korolyuk V.V.: Stochastic Models of Systems. Kluwer Academic Publishers, Dordrecht/Boston/London (1999)
10. Kwiecień, J., Filipowicz, B.: Firefly algorithm in optimization of queueing systems. Bull. Pol. Ac.: Tech. **60**(2), 363-368 (2012)
11. Legros, B., Jouini, O., Koole, G.: Optimal scheduling in call centers with a callback option. Perform. Eval. **95**, 1-40 (2016)
12. Malekian, R., Abdullah, A.H., Ye, N.: Novel packet queuing algorithm on packet delivery in mobile internet protocol version 6 networks. Appl. Math. Inf. Sci. **7**(3), 881-887 (2013)
13. Mangili, M., Martignon F., Capone, A.: Performance analysis of Content-Centric and Content-Delivery networks with evolving object popularity. Comp. Netw. **94**, 80-98 (2016)
14. Martin, J.B.: Large tandem queueing networks with blocking. Queueing Syst. **41**(1/2), 45-72 (2002)
15. Oniszczuk W.: Quality of service requirements in computer networks with blocking. In: Saeed, K., Pejas, J. (eds.) Information Processing and Security Systems, pp. 245-254. Springer Science+Business Media, New York (2005)
16. Oniszczuk, W.: Modeling of dynamical flow control procedures in closed type queuing models of a computer network with blocking. Automat. Contr. Comput. Sci. **39**(4), 60-69 (2005)
17. Oniszczuk W.: Tandem models with blocking in the computer subnetworks performance analysis. In: Saeed, K., Pejas, J., Mosdorf, R. (eds.) Biometrics, Computer Security Systems and Artificial Intelligence Applications, pp. 259-267. Springer Science+Business Media, New York (2006)
18. Oniszczuk, W.: Blocking and deadlock factors in series linked servers with HOL priority feedback service. Pol. J. Environ. Stud. **16**(5B), 145-151 (2007)
19. Oniszczuk W.: Analysis of an open linked series three-station network with blocking. In: Pejas, J., Saeed, K. (eds.) Advances in Information Processing and Protection, pp. 419-429. Springer Science+Business Media, New York (2007)
20. Oniszczuk, W.: An intelligent service strategy in linked networks with blocking and feedback. In: Nguyen, N.T., Katarzyniak, R. (eds.) Studies in Computational Intelligence

New Challenges in Applied Intelligence Technologies, SCI, vol. 134, pp. 351-361, Springer-Verlag, Berlin, Heidelberg (2008)

21. Oniszczuk, W.: Semi-Markov-based approach for analysis of open tandem networks with blocking and truncation. Int. J. Appl. Math. Comput. Sci. **19**(1), 151-163 (2009)

22. Oniszczuk, W.: Analysis of linked in series servers with blocking, priority feedback service and threshold policy. Int. J. Comput. Syst. Sci. Eng. **5**(1), 1-8 (2009)

23. Oniszczuk, W.: Loss tandem networks with blocking analysis – a semi-Markov approach. Bull. Pol. Ac.: Tech. **58**(4), 673-681 (2010)

24. Oniszczuk, W.: Open tandem networks with blocking analysis – two approaches. Control Cybern. **43**(1), 111- 132 (2014)

25. Oniszczuk W.: Blocking and deadlocking phenomena in two-server tandem configuration with optional feedback – modeling and parameter sensitivity investigation. In: Saeed K., Homenda W. (eds.) Computer Information Systems and Industrial Management. Lecture Notes in Computer Science, vol. 9842, pp. 441-452. Springer (2016). doi:10.1007/978-3-319-45378-1_40

26. Onvural, R.: Survey of closed queuing networks with blocking. Comput. Surv. **22**(2), 83-121 (1990)

27. Perros, H.G.: Queuing Networks with Blocking. Exact and Approximate Solution. Oxford University Press, New York (1994)

28. Raghavendran, CH.V., Naga Satish, G., Rama Sundari, M.V., Suresh Varma, P.: Tandem communication network model with DBA having Non Homogeneous Poisson arrivals and feedback for first node. International Journal of Computers and Technology **13**(9), 4922-4932 (2014)

29. Sunitha, G.P., Kumar, S.M.D., Kumar B.P.V.: A pre-emptive multiple queue congestion control for different traffic classes in WSN. In: 2014 International Conference on Circuits, Communication, Control and Computing, pp. 212-218. IEEE Conference Publications (2014). doi:10.1109/CIMCA.2014.7057793

30. Tikhonenko, O., Kempa, W.M.: On the queue-size distribution in the multi-server system with bounded capacity and packet dropping. Kybernetika **49**(6), 855-867 (2013)