Reliability Analysis of a Two-Server Heterogeneous Unreliable Queueing System with a Threshold Control Policy

Dmitry Efrosinin[1,2(✉)], Janos Sztrik[3], Mais Farkhadov[4], and Natalia Stepanova[4]

[1] Peoples' Friendship University of Russia (RUDN University), Miklukho-Maklaya Street 6, 117198 Moscow, Russia
[2] Johannes Kepler University Linz, Altenbergerstrasse 69, 4040 Linz, Austria
dmitry.efrosinin@jku.at
[3] University of Debrecen, Egyetem tér 1, Debrecen 4032, Hungary
sztrik.janos@inf.unideb.hu
[4] Institute of Control Sciences, RAS, Profsoyuznaya Street 65, 117997 Moscow, Russia
http://www.rudn.ru, http://www.jku.at, http://www.unideb.hu, http://www.ipu.ru

Abstract. Heterogeneous servers which can differ in service speed and reliability are getting more popular in modeling of modern communication systems. For a two-server queueing system with unreliable servers the allocation of customers between the servers is performed via a threshold control policy which prescribes to use the fastest server whenever it is free and the slower one only if the number of waiting customers exceeds some threshold level depending on the state of faster server. The main task of the paper consists in reliability analysis of the proposed system including evaluation of the stationary availability and reliability function. The effects of different parameters on introduced reliability characteristics are analyzed numerically.

Keywords: Reliability analysis · Quasi-birth-and-death process · Heterogeneous servers · Threshold policy · Matrix-geometric solution method

1 Introduction

To make modern communication systems superior in performance and reliability to the previous generation systems they can be supplied with heterogeneous communication links. Such links can differ in availability, link data throughputs,

D. Efrosinin—The publication was financially supported by the Ministry of Education and Science of the Russian Federation (the Agreement number 02.a03.21.0008), by the Russian Foundation for Basic Research, Project No. 16-37-60072 mol_a_dk and No. 15-08-08677 A, by the Austro-Hungarian Scientific Cooperation OMAA 96öu8.

A. Dudin et al. (Eds.): ITMM 2017, CCIS 800, pp. 13–27, 2017.
DOI: 10.1007/978-3-319-68069-9_2

power consumption and reliability characteristics. To model the dynamic behaviour of the data transmission links subject to breakdowns a queueing system with non-reliable servers can be used. The analysis of multi-server queueing systems generally assume the servers to be homogeneous. Mitrany and Avi-Izhak [11] and Neuts and Lucantoni [13] have studied the $M/M/s$ queueing system with server breakdowns and repairs. In paper of Levi and Yechiali [9] the queue $M/M/s$ with servers' vacations was analyzed. A recent paper of Efrosinin et al. [3] deals with an stationary analysis performed on the busy period for the multi-server Markovian queueing system with simultaneous failures of servers. The queues with heterogeneous non-reliable servers occur quite rarely as a research subject. A queueing system with two heterogeneous servers and multiple vacations was studied by Kumar and Madheswari [6], who obtained the stationary queue length distribution by using matrix geometric method and provided analysis of busy period and waiting time. In Kumar et al. [7] the same authors have introduced the $M/M/2$ queueing system with heterogeneous servers subject to catastrophes and provided a transient solution for the system under study. A heterogeneous two-server queueing system with balking and server breakdowns has been studied by Yue et al. [16]. In their study, some stationary mean performance measures are obtained using the matrix-geometric solution method.

In heterogeneous queueing system with one common queue, especially in case of the service without preemption, when the customer can not change the server during a service time, the customer allocation mechanism between the servers must be specified. The majority of heterogeneous systems investigated use heuristic service policies (e.g. the Fastest Free Server (FFS) or Random Service Selection (RSS) policies). In fact these policies are not optimal, if e.g. the mean response time must be minimized. As it is already known, see. e.g. the results of Efrosinin [1], Koole [5], Legros and Jouini [8], Lin and Kumar [10], Rykov and Efrosinin [15], for the heterogeneous queueing systems the optimal allocation policy belongs to a class of threshold policies, where the less effective server must be used only if the number of customers in the queue has reached some prespecified threshold level. The same result was confirmed for the queueing system with faster non-reliable server and absolutely reliable slower server in Efrosinin [2], Özkan and Kharoufeh [14] and for two non-reliable heterogeneous servers in system with a constant retrial discipline in Efrosinin and Sztrik [4]. In the latter paper it was shown that for the fixed threshold policy the corresponding Markov process is of the QBD (Quasi-birth-and-death) type with a tri-diagonal block infinitesimal matrix with a large number of bounding states.

While the first steps in performance analysis of controllable heterogeneous queueing systems have already been performed for completely reliable servers, a missing link to an applicability of heterogeneous models is a reliability analysis of such queues with servers subject to failures. In this paper we use a forward-elimination-backward-substitution method expressed in matrix form in terms of the Laplace-Stiltjes transforms (LST) combined with probability generating function (PGF) approach to evaluate reliability measures such as reliability function, which represents the complementary cumulative distribution function of the life time, and mean time to the first failure for each server separately and for the group of servers under the fixed threshold allocation control policy. The reliability functions are obtained in terms of the Laplace transform (LT) and numerical inversion algorithm is used to get the time dependent functions. Additionally a new discrete reliability metric in form of the distribution of the number of failures during a certain life time is introduced. We expect that the proposed results can be generalized to the case of an arbitrary controllable non-reliable queueing model with a QBD structure.

The rest of the paper is organized as follows. In Sect. 2, we describe the mathematical model and give a presentation of the stationary distribution of the system state using a matrix-geometric solution method. In Sect. 3, we develop computational analysis for the stationary reliability characteristics, for the reliability function and mean time to failure. The number of failures during a certain life time is investigated in Sect. 4. In Sect. 5, numerical illustrations are provided to highlight the effect of some parameters on the derived reliability characteristics.

Hereafter, the notations $\mathbf{e}(n)$, $\mathbf{e}_j(n)$, and I_n are used respectively for the column-vector consisting of 1's, the column vector with 1 in the j-th (beginning from 0-th) position and 0 elsewhere, and an identity matrix of the dimension n. When there is no need to emphasize the dimension of these vectors the suffix will be suppressed and dimension is determined by the context. The expressions $diag(a_1,\ldots,a_n)$, $diag^+(a_1,\ldots,a_n)$, and $diag^-(a_1,\ldots,a_n)$ denote respectively the diagonal matrix, the upper diagonal matrix, and the lower diagonal matrix with entries $a_1,\ldots,a_n$ that could be scalars or matrices.

2 Mathematical Model and Stationary Distribution

In the present paper we deal with a two-server heterogeneous non-reliable queueing model of the type $M/M/2$. The customers arrive according to a Poisson process with arrival rate λ. The service times are exponentially distributed with rates μ_1 and μ_2, where $\mu_1 \geq \mu_2$. We assume that the server fails respectively at an exponential rate α_1 and α_2. The servers can fail only if they are busy. The failed server is repaired immediately and the time required to repair it is exponentially distributed respectively with rate β_1 and β_2. The customer being served at the failure moment is left at this server during the repair time and can be served when the server becomes operational again. The allocation mechanism between two servers is based on a threshold policy: depending on the state of faster server the slower one is used whenever the number of customers in the queue exceeds a certain threshold level.

Let $Q(t)$ and $D(t) = \{D_1(t), D_2(t)\}$ denote, respectively, the number of customers in the queue and the vector state of servers at time t, where

$$D_j(t) = \begin{cases} 0, & \text{the server } j \text{ is idle,} \\ 1, & \text{the server } j \text{ is busy and operational,} \\ 2, & \text{the server } j \text{ is failed.} \end{cases}$$

The threshold policy $f = (q_1, q_2)$ is defined by two threshold levels $1 \leq q_2 \leq q_1 < \infty$. According to this policy server 1 must be activated whenever it is free and there are customers in the queue, whereas server 2 is used only if server 1 is in state 1 or 2 and the number of customers in the queue has reached the value q_1 or q_2. The process

$$\{X(t)\}_{t\geq 0} = \{Q(t), D(t)\}_{t\geq 0} \tag{1}$$

is a continuous-time Markov chain with a state space given by

$$E = \{x = (q, d_1, d_2); q \in \mathbb{N}_0,\, (d_1, d_2) \in E_D\}, \tag{2}$$

where E_D is a set of states of servers which is defined as

$$E_D = \left\{ (d_1, d_2); \begin{array}{l} d_j \in \{0,1,2\},\, j \in \{1,2\},\, q = 0 \\ d_1 \in \{1,2\}, d_2 \in \{0,1,2\},\, 1 \leq q \leq q_2 - 1, \\ d_1 \in \{1,2\}, d_2 \in \{0,1,2\}, (d_1, d_2) \neq (2,0),\, q_2 \leq q \leq q_1 - 1, \\ d_j \in \{1,2\},\, j \in \{1,2\},\, q \geq q_1, \end{array} \right\}.$$

Next we partition E in blocks as follows,

$$(\mathbf{0}, \mathbf{0}) = \{(0,0,d_2); d_2 \in \{0,1,2\}\},$$

$$(\mathbf{q}, \mathbf{1}) = \begin{cases} \{(q,1,0),(q,2,0),(q,1,1),(q,2,1),(q,1,2),(q,2,2)\}, & 0 \leq q \leq q_2 - 1, \\ \{(q,1,0),(q,1,1),(q,2,1),(q,1,2),(q,2,2)\}, & q_2 \leq q \leq q_1 - 1, \\ \{(q,1,1),(q,2,1),(q,1,2),(q,2,2)\}, & q \geq q_1. \end{cases}$$

Due to above notation, the infinitesimal generator olude the rates of transition fromf the Markov chain $\{X(t)\}_{t\geq 0}$ has the block-tridiagonal structure,

$$\begin{aligned} \Lambda = [\lambda_{xy}]_{x,y\in E} = diag(Q_{1,0}, \underbrace{Q_{1,1}, \ldots, Q_{1,1}}_{q_2-1}, Q_{1,2}, \underbrace{Q_{1,3}, \ldots, Q_{1,3}}_{q_1-q_2-1}, Q_{1,4}, Q_{1,5}, \ldots) \\ + \, diag^{+}(Q_{0,1}, \underbrace{Q_{0,2}, \ldots, Q_{0,2}}_{q_2-1}, Q_{0,3}, \underbrace{Q_{0,4}, \ldots, Q_{0,4}}_{q_1-q_2-1}, Q_{0,5}, Q_{0,6}, \ldots) \\ + \, diag^{-}(Q_{2,1}, \underbrace{Q_{2,2}, \ldots, Q_{2,2}}_{q_2-1}, Q_{2,3}, \underbrace{Q_{2,4}, \ldots, Q_{2,4}}_{q_1-q_2-1}, Q_{2,5}, Q_{2,6}, \ldots). \end{aligned}$$

The square matrices $Q_{1,n}, 0 \leq n \leq 5$, include the rates of the output from the current block of states,

$$Q_{1,0} = \begin{pmatrix} -\lambda & 0 & 0 \\ \mu_2 & -(\lambda + \alpha_2 + \mu_2) & \alpha_2 \\ 0 & \beta_2 & -(\lambda + \beta_2) \end{pmatrix},$$

$$Q_{1,1} = \begin{pmatrix} -(\lambda+\mu_1+\alpha_1) & \alpha_1 & 0 & 0 & 0 & 0 \\ \beta_1 & -(\lambda+\beta_1) & 0 & 0 & 0 & 0 \\ \mu_2 & 0 & -(\lambda+\mu+\alpha) & \alpha_1 & \alpha_2 & 0 \\ 0 & \mu_2 & \beta_1 & -(\lambda+\alpha_2+\beta_1+\mu_2) & 0 & \alpha_2 \\ 0 & 0 & \beta_2 & 0 & -(\lambda+\alpha_1+\beta_2+\mu_1) & \alpha_1 \\ 0 & 0 & 0 & \beta_2 & \beta_1 & -(\lambda+\beta) \end{pmatrix},$$

$Q_{1,2} = Q_{1,1} + \lambda \mathbf{e}_1(6) \otimes \mathbf{e}'_3(6),$

$$Q_{1,3} = \begin{pmatrix} -(\lambda+\mu_1+\alpha_1) & 0 & 0 & 0 & 0 \\ \mu_2 & -(\lambda+\mu+\alpha) & \alpha_1 & \alpha_2 & 0 \\ 0 & \beta_1 & -(\lambda+\alpha_2+\beta_1+\mu_2) & 0 & \alpha_2 \\ 0 & \beta_2 & 0 & -(\lambda+\alpha_1+\beta_2+\mu_1) & \alpha_1 \\ 0 & 0 & \beta_2 & \beta_1 & -(\lambda+\beta) \end{pmatrix},$$

$Q_{1,4} = Q_{1,3} + \lambda \mathbf{e}_0(5) \otimes \mathbf{e}'_1(5),$

$$Q_{1,5} = \begin{pmatrix} -(\lambda+\mu+\alpha) & \alpha_1 & \alpha_2 & 0 \\ \beta_1 & -(\lambda+\alpha_2+\beta_1+\mu_2) & 0 & \alpha_2 \\ \beta_2 & 0 & -(\lambda+\alpha_1+\beta_2+\mu_1) & \alpha_1 \\ 0 & \beta_2 & \beta_1 & -(\lambda+\beta) \end{pmatrix}.$$

The rectangular matrices $Q_{0,n}, 1 \le n \le 6$, include the rates of transitions from subsequent block to the current one,

$$Q_{0,1} = \lambda \begin{pmatrix} 1&0&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&0&1&0 \end{pmatrix}, \quad Q_{0,3} = \lambda \begin{pmatrix} 1&0&0&0&0 \\ 0&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&1 \end{pmatrix}, \quad Q_{0,5} = \lambda \begin{pmatrix} 0&0&0&0 \\ 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix},$$

$Q_{0,2} = \lambda I_6, \quad Q_{0,4} = \lambda I_5, \quad Q_{0,6} = \lambda I_4$, $\mu = \mu_1 + \mu_2$, $\alpha = \alpha_1 + \alpha_2$, $\beta = \beta_1 + \beta_2$.

The rectangular matrices $Q_{2,n}, 1 \le n \le 6$, include the rates of transition from the previous block to the current one,

$$Q_{2,1} = \begin{pmatrix} \mu_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \mu_1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mu_1 \\ 0 & 0 & 0 \end{pmatrix}, \quad Q_{2,2} = \begin{pmatrix} \mu_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mu_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad Q_{2,3} = \begin{pmatrix} \mu_1 & 0 & 0 & \alpha_1 & 0 & 0 \\ 0 & 0 & \mu_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$Q_{2,4} = \begin{pmatrix} \mu_1 & 0 & \alpha_1 & 0 & 0 \\ 0 & \mu_1 & 0 & 0 & 0 \\ 0 & 0 & \mu_2 & 0 & 0 \\ 0 & 0 & 0 & \mu_1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad Q_{2,5} = \begin{pmatrix} 0 & \mu & 0 & 0 & 0 \\ 0 & 0 & \mu_2 & 0 & 0 \\ 0 & 0 & 0 & \mu_1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad Q_{2,6} = \begin{pmatrix} \mu & 0 & 0 & 0 \\ 0 & \mu_2 & 0 & 0 \\ 0 & 0 & \mu_1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Denote by $\boldsymbol{\pi} = (\boldsymbol{\pi}_{0,0}, \boldsymbol{\pi}_{0,1}, \boldsymbol{\pi}_{1,1}, \boldsymbol{\pi}_{2,1}, \dots)$ the stationary probability vector of Λ which satisfies

$$\boldsymbol{\pi}\Lambda = \mathbf{0}, \ \boldsymbol{\pi}\mathbf{e} = 1. \tag{3}$$

The computation of the stationary distribution is reduced to solving a block-tridiagonal system. The process $\{X(t)\}_{t \ge 0}$ is in the format of a quasi-birth-and-death (QBD) process which allows to apply the matrix-analytic approach.

By [12, Theorem 3.1.1] it is well known that the stationary probability vector $\boldsymbol{\pi}$ of the QBD process exists if and only if

$$\mathbf{p}Q_{0,6}\mathbf{e}(4) < \mathbf{p}Q_{2,6}\mathbf{e}(4),$$

where $\mathbf{p} = (p_1, p_2, p_3, p_4)$ is the invariant probability of the matrix $Q_{0,6} + Q_{1,5} + Q_{2,6}$. This vector can be obtained by solving the system $\mathbf{p}(Q_{0,6} + Q_{1,5} + Q_{2,6}) = \mathbf{0}$ and $\mathbf{pe}(4) = 1$. After some routine manipulation we can obtain the condition

$$\rho = \frac{\lambda}{\sum_{j=1}^{2} \frac{\beta_j \mu_j}{\alpha_j + \beta_j}} < 1. \tag{4}$$

Theorem 1. *The vectors of stationary probabilities $\boldsymbol{\pi}_{q,i}$, $q \geq 0$, can be computed as follows,*

$$\begin{aligned}
\boldsymbol{\pi}_{0,0} &= \boldsymbol{\pi}_{q_1,1} \prod_{j=0}^{q_1} M_{q_1-j}, \qquad (5)\\
\boldsymbol{\pi}_{q,1} &= \boldsymbol{\pi}_{q_1,1} \prod_{j=0}^{q_1-q-1} M_{q_1-j},\ 0 \leq q \leq q_1 - 1,\\
\boldsymbol{\pi}_{q,1} &= \boldsymbol{\pi}_{q_1,1} R^{q-q_1},\ q \geq q_1,
\end{aligned}$$

where the matrices $M_i, 0 \leq i \leq q_1$, are recursively defined

$$\begin{aligned}
M_0 &= -Q_{2,1}Q_{1,0}^{-1},\ M_1 = -Q_{2,2}(M_0 Q_{0,1} + Q_{1,1})^{-1}, \qquad (6)\\
M_q &= -Q_{2,2}(M_{q-1}Q_{0,2} + Q_{1,1})^{-1},\ 2 \leq q \leq q_2 - 1,\\
M_{q_2} &= -Q_{2,3}(M_{q_2-1}Q_{0,2} + Q_{1,2})^{-1},\ M_{q_2+1} = -Q_{2,4}(M_{q_2}Q_{0,3} + Q_{1,3})^{-1},\\
M_q &= -Q_{2,4}(M_{q-1}Q_{0,4} + Q_{1,3})^{-1},\ q_2 + 2 \leq q \leq q_1 - 1,\\
M_{q_1} &= -Q_{2,5}(M_{q_1-1}Q_{0,4} + Q_{1,4})^{-1}.
\end{aligned}$$

The vector $\boldsymbol{\pi}_{q_1,1}$ is a unique solution of the system of equations

$$\begin{aligned}
&\boldsymbol{\pi}_{q_1,1}\Big[\sum_{q=-1}^{q_1-1} \prod_{j=0}^{q_1-q-1} M_{q_1-j} + (I-R)^{-1}\Big]\mathbf{e}(4) = 1, \qquad (7)\\
&\boldsymbol{\pi}_{q_1,1}(M_{q_1}Q_{0,5} + Q_{1,5} + RQ_{2,6}) = \mathbf{0}.
\end{aligned}$$

The matrix R is a minimal solution of the matrix quadratic equation,

$$R^2 Q_{2,6} + RQ_{1,5} + Q_{0,6} = 0. \tag{8}$$

Proof. The last row of (5) and equation $R^2 Q_{2,6} + RQ_{1,5} + Q_{0,6} = 0$ follow from the properties of the QBD process [12]. If the stability condition holds, then (3) yields the system,

$$\begin{aligned}
&\boldsymbol{\pi}_{0,0}Q_{1,0} + \boldsymbol{\pi}_{0,1}Q_{2,1} = \mathbf{0},\\
&\boldsymbol{\pi}_{q-1,1}Q_{0,1} + \boldsymbol{\pi}_{q,1}Q_{1,1} + \boldsymbol{\pi}_{q+1,1}Q_{2,2} = \mathbf{0},\ 2 \le q \le q_2 - 1,\\
&\boldsymbol{\pi}_{q_2-1,1}Q_{0,2} + \boldsymbol{\pi}_{q_2,1}Q_{1,2} + \boldsymbol{\pi}_{q_2+1,1}Q_{2,3} = \mathbf{0},\\
&\boldsymbol{\pi}_{q_2,1}Q_{0,3} + \boldsymbol{\pi}_{q_2+1,1}Q_{1,3} + \boldsymbol{\pi}_{q_2+2,1}Q_{2,4} = \mathbf{0},\\
&\boldsymbol{\pi}_{q-1,1}Q_{0,4} + \boldsymbol{\pi}_{q,1}Q_{1,3} + \boldsymbol{\pi}_{q+1,1}Q_{2,4} = \mathbf{0},\ q_2 + 2 \le q \le q_1 - 1,\\
&\boldsymbol{\pi}_{q_1-1,1}Q_{0,4} + \boldsymbol{\pi}_{q_1}Q_{1,4} + \boldsymbol{\pi}_{q_1+1}Q_{2,5} = \mathbf{0},\\
&\boldsymbol{\pi}_{q_1,1}R^{q-q_1-1}Q_{0,5} + \boldsymbol{\pi}_{q_1,1}R^{q-q_1}Q_{1,5} + \boldsymbol{\pi}_{q_1,1}R^{q-q_1+1}Q_{2,6} = \mathbf{0},\ q \ge q_1 + 1.
\end{aligned}$$

The routine of substitution applied to the previous system leads to recursive relations,

$$\begin{aligned}
&\boldsymbol{\pi}_{0,0} = \boldsymbol{\pi}_{0,1}M_0,\\
&\boldsymbol{\pi}_{q,1} = \boldsymbol{\pi}_{q+1,1}M_{q+1},\ 1 \le q \le q_1 - 1,
\end{aligned} \tag{9}$$

where M_q is defined by (6). Hence it implies the first two rows of (5). Finally the vector $\boldsymbol{\pi}_{q_1,1}$ is obviously a unique solution of the system of equations (7) which consists of the normalizing condition and the balance equation for the probability vector $\boldsymbol{\pi}_{q_1,1}$ of the boundary states.

3 Reliability Characteristics of the System and Servers

In this section we consider some reliability quantities of the system and servers. Denote by

$$\begin{aligned}
A_1(t) &= \mathbb{P}[X(t) = (q, d_1, d_2); d_1 \ne 2 \vee d_2 \ne 2],\\
A_2(t) &= \mathbb{P}[X(t) = (q, d_1, d_2); d_1 \ne 2 \wedge d_2 \ne 2],\\
A_3(t) &= \mathbb{P}[X(t) = (q, d_1, d_2); d_1 \ne 2],\\
A_4(t) &= \mathbb{P}[X(t) = (q, d_1, d_2); d_2 \ne 2],
\end{aligned}$$

the pointwise availability of the system and servers. The stationary availability in case $n, 1 \le n \le 4$, is defined as $A_n = \lim_{t\to\infty} A_n(t)$.

Corollary 1. *The stationary availability can be computed by*

$$A_n = \boldsymbol{\pi}_{0,0}\mathbf{x}_{n,1} + \sum_{q=0}^{q_2-1} \boldsymbol{\pi}_{q,1}\mathbf{x}_{n,2} + \sum_{q=q_2}^{q_1-1} \boldsymbol{\pi}_{q,1}\mathbf{x}_{n,3} + \boldsymbol{\pi}_{q_1,1}(I-R)^{-1}\mathbf{x}_{n,4},\ 1 \le n \le 4,$$

where $A_2 = A_3 + A_4 - A_1$ *and*

$$\mathbf{x}_{1,1} = \mathbf{e}(3),\ \mathbf{x}_{1,2} = \sum_{k=0}^{4} \mathbf{e}_k(6),\ \mathbf{x}_{1,3} = \sum_{k=0}^{3} \mathbf{e}_k(5),\ \mathbf{x}_{1,4} = \sum_{k=0}^{2} \mathbf{e}_k(4),$$

$$\mathbf{x}_{2,1} = \sum_{k=0}^{1} \mathbf{e}_k(3),\ \mathbf{x}_{2,2} = \sum_{k=0}^{1} \mathbf{e}_{2k}(6),\ \mathbf{x}_{2,3} = \sum_{k=0}^{1} \mathbf{e}_k(5),\ \mathbf{x}_{2,4} = \mathbf{e}_0(4),$$

$$\mathbf{x}_{3,1} = \mathbf{e}(3),\ \mathbf{x}_{3,2} = \sum_{k=0}^{2} \mathbf{e}_{2k}(6),\ \mathbf{x}_{3,3} = \mathbf{e}_0 + \sum_{k=0}^{1} \mathbf{e}_{2k+1}(5),\ \mathbf{x}_{3,4} = \sum_{k=0}^{1} \mathbf{e}_{2k}(4),$$

$$\mathbf{x}_{4,1} = \sum_{k=0}^{1} \mathbf{e}_k(3),\ \mathbf{x}_{4,2} = \sum_{k=0}^{3} \mathbf{e}_k(6),\ \mathbf{x}_{4,3} = \sum_{k=0}^{2} \mathbf{e}_k(5),\ \mathbf{x}_{4,4} = \sum_{k=0}^{1} \mathbf{e}_k(4).$$

Corollary 2. *The stationary failure frequency of the server* $l \in \{1, 2\}$ *can be computed by*

$$B_l = \alpha_l \boldsymbol{\pi}_{0,0}\mathbf{y}_{l,1} + \sum_{q=0}^{q_2-1} \boldsymbol{\pi}_{q,1}\mathbf{y}_{l,2} + \sum_{q=q_2}^{q_1-1} \boldsymbol{\pi}_{q,1}\mathbf{y}_{l,3} + \boldsymbol{\pi}_{q_1,1}(I - R)^{-1}\mathbf{y}_{l,4},\ 1 \le l \le 2,$$

where

$$\mathbf{y}_{1,1} = \mathbf{0},\ \mathbf{y}_{1,2} = \sum_{k=0}^{2} \mathbf{e}_{2k}(6),\ \mathbf{y}_{1,3} = \mathbf{e}_0(5) + \sum_{k=0}^{1} \mathbf{e}_{2k+1}(5),\ \mathbf{y}_{1,4} = \sum_{k=0}^{1} \mathbf{e}_{2k}(4),$$

$$\mathbf{y}_{2,1} = \mathbf{e}_1(3),\ \mathbf{y}_{2,2} = \sum_{k=2}^{3} \mathbf{e}_k(6),\ \mathbf{y}_{2,3} = \sum_{k=1}^{2} \mathbf{e}_k(5),\ \mathbf{y}_{2,4} = \sum_{k=0}^{1} \mathbf{e}_k(4).$$

Denote by T the random time to the first failure of one of server. The corresponding reliability function, which is the same as the complementary cumulative distribution function of the life time T, is then defined as

$$R(t) = \mathbb{P}[T > t].$$

In this section we intend to obtain this function in terms of the Laplace transform $\tilde{R}(s) = \int_0^\infty R(s)e^{-st}dt, Re[s] > 0$. In order to realize it we let the corresponding failure states be absorbing states. In this case we obtain new process which can be modelled by the auxiliary continuous-time absorbing Markov chains $\{\hat{X}(t)\}_{t\ge 0}$ with state space $\hat{E} = E \setminus \{x = (q, d_1, d_2); q \in \mathbb{N}_0,\ d_1 = 2 \vee d_2 = 2\}$. We describe two main approaches to get the function $\tilde{R}(s)$: By means of the transient solution of the absorbing Markov chain and using the remaining life time.

Theorem 2. *The Laplace transform of* $R(t)$ *is given by*

$$\tilde{R}(s) = \tilde{P}_{1,0}(s, 1) + \tilde{P}_{1,1}(s, 1) + \tilde{P}_{1,2}(s, 1), \tag{10}$$

where

$$\tilde{P}_{1,0}(s,1) = \frac{1+\alpha_1\tilde{\pi}_{(0,0,0)}(s) - \lambda\tilde{\pi}_{(q_1-1,1,0)}(s) + \mu_2\tilde{P}_{1,1}(s,1)}{s+\alpha_1}, \tag{11}$$

$$\tilde{P}_{1,1}(s,1) = \frac{\alpha_1\tilde{\pi}_{(0,0,1)}(s) + \lambda(\tilde{\pi}_{(q_1-1,1,0)}(s) - \tilde{\pi}_{(q_1-1,1,1)}(s)) + \mu\tilde{\pi}_{(q_1,1,1)}(s)}{s+\alpha+\mu_2},$$

$$\tilde{P}_{1,2}(s,1) = \frac{\lambda\tilde{\pi}_{(q_1-1,1,1)}(s) - \mu\tilde{\pi}_{(q_1,1,1)}(s)}{s+\alpha},$$

the functions $\tilde{\pi}_x(s)$ *are of the form,*

$$\tilde{\pi}_{(q_1,1,1)}(s) = \frac{\lambda z(s)\tilde{L}_{q_1}(s)\mathbf{e}_1(2)}{\mu - \lambda z(s)\tilde{M}_{q_1}(s)\mathbf{e}_1(2)}, \tag{12}$$

$$(\tilde{\pi}_{(q_1-1,1,0)}(s), \tilde{\pi}_{(q_1-1,1,1)}(s)) = \tilde{\pi}_{(q_1,1,1)}(s)\tilde{M}_{q_1}(s) + \tilde{L}_{q_1}(s), \tag{13}$$

$$(\tilde{\pi}_{(0,0,0)}(s), \tilde{\pi}_{(0,0,1)}(s)) = \tilde{\pi}_{(q_1,1,1)}(s)\prod_{i=0}^{q_1}\tilde{M}_{q_1-i}(s) + \sum_{i=0}^{q_1}\tilde{L}_{q_1-i}(s)\prod_{j=i+1}^{q_1}\tilde{M}_{q_1-j}(s), \tag{14}$$

the matrices $\tilde{M}_i(s)$ *and* $\tilde{L}_i(s)$ *are evaluated recursively,*

$$\tilde{M}_0(s) = \mu_1\tilde{N}_0(s),\ \tilde{L}_0(s) = \mathbf{e}_0'(2)\tilde{N}_0(s),\ \tilde{N}_0(s) = -(\hat{Q}_{1,0} - sI_2)^{-1}, \tag{15}$$

$$\tilde{M}_q(s) = \mu_1\tilde{N}_q(s),\ \tilde{L}_q(s) = \lambda\tilde{L}_{q-1}(s)\tilde{N}_q(s),\ \tilde{N}_q(s) = -(\hat{Q}_{1,1} - sI_2 + \lambda\tilde{M}_{q-1}(s))^{-1}\ q = \overline{1, q_1-1},$$

$$\tilde{M}_{q_1}(s) = -\mu\mathbf{e}_1'(2)\tilde{N}_{q_1}(s),\ \tilde{L}_{q_1}(s) = -\lambda\tilde{L}_{q_1-1}\tilde{N}_{q_1}(s),\ \tilde{N}_{q_1}(s) = (\hat{Q}_{1,2} - sI_2 + \lambda\tilde{M}_{q_1-1}(s))^{-1},$$

the matrices $\hat{Q}_{1,0}, \hat{Q}_{1,1}$ *and* $\hat{Q}_{1,2}$ *are of the form*

$$\hat{Q}_{1,0} = \begin{pmatrix} -\lambda & 0 \\ \mu_2 & -(\lambda+\alpha_2+\mu_2) \end{pmatrix},\ \hat{Q}_{1,1} = \begin{pmatrix} -(\lambda+\alpha_1+\mu_1) & 0 \\ \mu_2 & -(\lambda+\alpha+\mu) \end{pmatrix},$$

$$\hat{Q}_{1,2} = \begin{pmatrix} -(\lambda+\alpha_1+\mu_1) & \lambda \\ \mu_2 & -(\lambda+\alpha+\mu) \end{pmatrix},$$

the function $z(s)$ *is defined as*

$$z(s) = \frac{s+\alpha+\lambda+\mu}{2\lambda} - \sqrt{\left(\frac{s+\alpha+\lambda+\mu}{2\lambda}\right)^2 - \frac{\mu}{\lambda}}. \tag{16}$$

Proof. The absorbing states of the process $\{\hat{X}_2(t)\}$ are $x = (q, 2, d_2), d_2 \in \{0, 1, 2\}$ and $x = (q, d_1, 2), d_1 \in \{0, 1, 2\}$. Using the same notations as in previous section we can get the following set of Kolmogorov differential equations,

sztrik.janos@inf.unideb.hu

$$\pi'_{(0,0,0)}(t) = -\lambda\pi_{(0,0,0)}(t) + \mu_1\pi_{(0,1,0)}(t) + \mu_2\pi_{(0,0,1)}(t), \tag{17}$$
$$\pi'_{(q,1,0)}(t) = -(\alpha_1+\lambda+\mu_1)\pi_{(q,1,0)} + \lambda\pi_{(q-1,1,0)}(t) + \mu_1\pi_{(q+1,1,0)}(t) + \mu_2\pi_{(q,1,1)}(t),$$
$$0 \le q \le q_1 - 2,$$
$$\pi'_{(q_1-1,1,0)}(t) = -(\alpha_1+\lambda+\mu_1)\pi_{(q_1-1,1,0)} + \lambda\pi_{(q_1-2,1,0)}(t) + \mu_2\pi_{(q_1-1,1,1)},$$
$$\pi'_{(0,0,1)}(t) = -(\alpha_2+\lambda+\mu_2)\pi_{(0,0,1)}(t) + \mu_1\pi_{(0,1,1)}(t),$$
$$\pi'_{(0,1,1)}(t) = -(\alpha_2+\lambda+\mu_2)\pi_{(0,0,1)}(t) + \lambda\pi_{(0,0,1)}(t) + \mu_1\pi_{(0,1,1)}(t),$$
$$\pi'_{(q,1,1)}(t) = -(\alpha+\lambda+\mu)\pi_{(q,1,1)}(t) + \lambda\pi_{(q-1,1,1)}(t) + \mu_1\pi_{(q+1,1,1)}(t),\ 1 \le q \le q_1-2,$$
$$\pi'_{(q_1-1,1,1)}(t) = -(\alpha+\lambda+\mu)\pi_{(q_1-1,1,1)}(t) + \lambda\pi_{(q_1-1,1,0)}(t) + \lambda\pi_{(q_1-2,1,1)}(t) + \mu\pi_{(q_1,1,1)}(t)$$

with initial conditions $\pi_{(0,0,0)}(0) = 1$ and $\pi_x(0) = 0, x \in \hat{E}_2$. By taking Laplace transforms of these equations, where $\tilde{\pi}_x(s) = \int_0^\infty \pi_x(t)e^{-st}dt, Re[s] \ge 0$, and using then their partial generating functions,

$$\tilde{P}_{1,0}(s,z) = \tilde{\pi}_{(0,0,0)}(s) + \sum_{q=0}^{q_1-1} \tilde{\pi}_{(q,1,0)}(s)z^{i+1},$$
$$\tilde{P}_{1,1}(s,z) = \tilde{\pi}_{(0,0,1)}(s) + \sum_{q=0}^{q_1-1} \tilde{\pi}_{(q,1,1)}(s)z^{i+1},$$
$$\tilde{P}_{1,2}(s,z) = \sum_{q=q_1}^{\infty} \tilde{\pi}_{(q,1,1)}(s)z^{i+1}$$

for $|z| < 1$, after some manipulation the system (17) is transformed into the set of equations for the introduced double transforms,

$$\tilde{P}_{1,0}(s,z) = \frac{z + \tilde{\pi}_{(0,0,0)}(s)(\mu_1(z-1)+\alpha_1 z) - \lambda z^{q_1+2}\tilde{\pi}_{(q_1-1,1,0)}(s) + \mu_2 z\tilde{P}_{1,1}(s,z)}{-\lambda z^2 + (s+\alpha_1+\lambda+\mu_1)z - \mu_1},$$
$$\tilde{P}_{1,1}(s,z) = \frac{\tilde{\pi}_{(0,0,0)}(s)(z(\alpha_1+\mu_1)-\mu_1) + \lambda(\tilde{\pi}_{(q_1-1,1,0)}(s) - z^{q_1+1}\pi_{(q_1-1,1,1)}(s)) + \mu\tilde{\pi}_{(q_1,1,1)}(s)}{-\lambda z^2 + (s+\alpha+\lambda+\mu)z - \mu_1},$$
$$\tilde{P}_{1,2}(s,z) = \frac{z^{q_1+1}(\lambda z\tilde{\pi}_{(q_1-1,1,1)}(s) - \mu\tilde{\pi}_{(q_1,1,1)}(s))}{-\lambda z^2 + (s+\alpha+\lambda+\mu)z - \mu}.$$

Denote by $F(s,z) = -\lambda z^2 + (s+\alpha+\lambda+\mu)z - \mu$ the auxiliary function for the denominator of $\tilde{P}_{1,2}(s,z)$. It is easy to see that

$$F(s,0) = -\mu < 0, \quad F(s,1) = s + \alpha \ge 0.$$

Thus the square equation $F(s,z) = 0$ has for any $s > 0$ two roots and the minimal of them takes the value in the interval $[0,1]$. This root we denote by

$$z(s) = \frac{s+\alpha+\lambda+\mu}{2\lambda} - \sqrt{\Big(\frac{s+\alpha+\lambda+\mu}{2\lambda}\Big)^2 - \frac{\mu}{\lambda}}.$$

Since the function $\tilde{P}_{1,2}(s,z)$ is analytical, the numerator of this function must be zero at point $z = z(s)$ as well, i.e.

$$\lambda z(s)\tilde{\pi}_{(q_1-1,1,1)}(s) - \mu\tilde{\pi}_{(q_1,1,1)}(s) = 0. \tag{18}$$

To have a second equation for the boundary transforms $\tilde{\pi}_{(q_1-1,1,1)}(s)$ and $\tilde{\pi}_{(q_1,1,1)}(s)$ denote by

$$\tilde{\boldsymbol{\pi}}_{0,0}(s) = (\tilde{\pi}_{(0,0,0)}(s), \tilde{\pi}_{(0,0,1)}(s)), \tilde{\boldsymbol{\pi}}_{q,1}(s) = (\tilde{\pi}_{(q,1,0)}(s), \tilde{\pi}_{(q,1,1)}(s)),\ 1 \le q \le q_1 - 1.$$

For the system of the Laplace transforms $\tilde{\pi}_x(s)$ obtained from (17) we can get the following relations in matrix form,

$$\tilde{\boldsymbol{\pi}}_{0,0}(s) = -\mu_1 \tilde{\boldsymbol{\pi}}_{0,1}(s)(\hat{Q}_{1,0} - sI_2)^{-1} - \mathbf{e}_0'(2)(\hat{Q}_{1,0} - sI_2)^{-1} = \tilde{\boldsymbol{\pi}}_{0,1}(s)\tilde{M}_0(s) + \tilde{L}_0(s).$$

The substitution of the last expression into the matrix relation for $\tilde{\pi}_{0,1}(s)$ yields

$$\begin{aligned}\tilde{\boldsymbol{\pi}}_{0,1}(s) &= -\mu_1 \tilde{\boldsymbol{\pi}}_{1,1}(s)(\hat{Q}_{1,1} - sI_2 + \lambda\tilde{M}_0(s))^{-1} - \lambda\tilde{L}_0(s)(\hat{Q}_{1,1} - sI_2 + \lambda\tilde{M}_0(s))^{-1} \\ &= \tilde{\boldsymbol{\pi}}_{1,1}(s)\tilde{M}_1(s) + \tilde{L}_1(s).\end{aligned}$$

Sequential application of such forward-elimination-backward-substitution method leads to the following recursive relations

$$\begin{aligned}&\tilde{\boldsymbol{\pi}}_{q-1,1}(s) = \tilde{\boldsymbol{\pi}}_{q,1}(s)\tilde{M}_q(s) + \tilde{L}_q(s),\ 1 \le q \le q_1 - 2, \\ &\tilde{\boldsymbol{\pi}}_{q_1-1,1}(s) = \tilde{\pi}_{(q_1,1,1)}(s)\tilde{M}_{q_1}(s) + \tilde{L}_{q_1}(s),\end{aligned}$$

where $\tilde{M}_q(s)$ and $\tilde{L}_q(s)$ can be calculated by (15). By combining the relation

$$\tilde{\pi}_{(q_1-1,1,1)}(s) = (\pi_{q_1,1,1}(s)\tilde{M}_{q_1}(s) + \tilde{L}_{q_1}(s))\mathbf{e}_1(2)$$

and (18), we may express $\tilde{\pi}_{(q_1,1,1)}(s)$ in form (12). The transforms for the rest of boundary states can be hence evaluated as a functions of $\tilde{\pi}_{(q_1-1,1,1)}(s)$. Finally the double transforms are calculated at point $z = 1$ and substituted into (10).

4 Numerical Results

In this section we present some numerical examples to study the effect of system parameters on proposed reliability measures. First we fix the systemparameters at values

$$\begin{aligned}&\lambda = 1.7,\ \mu_1 = 2.4,\ \mu_2 = 0.4,\ \alpha_1 = 0.1,\ \alpha_2 = 0.2, \\ &\beta_1 = 0.3,\ \beta_2 = 0.3,\ \rho = 0.83,\ q_1 = 9,\ q_2 = 6.\end{aligned}$$

In all cases presented below the parametric values are chosen in such a way that the ergodicity condition holds.

In Figs. 1 and 2 the stationary availabilities $A_i, 1 \le i \le 4$, are plotted against the arrival rate λ versus failure rates α_1, α_2 and repair rates β_1, β_2, respectively. As we expect, A_i decreases with increasing λ. The upper curves correspond to the lower value of α_1 and α_2 and to the higher value of $\beta_1$and β_2. The availabilities A_1, A_2 and A_3 take different values by changing of failure and repair rates of servers. We notice that descriptor A_3 changes by varying α_1 and β_1 but it is insensitive to the change of α_2 and β_2. It happens since the parameters α_1 and

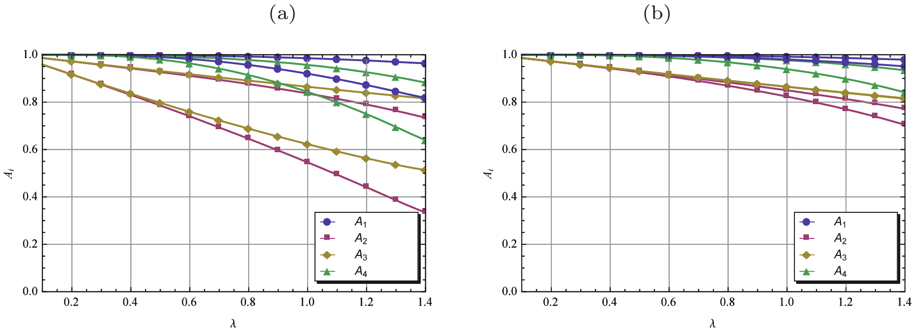

Fig. 1. The availability A_i, $1 \le i \le 4$, for $\alpha_1 = 0.1, 0.3$ (a) and $\alpha_2 = 0.1, 0.3$ (b) vs. λ

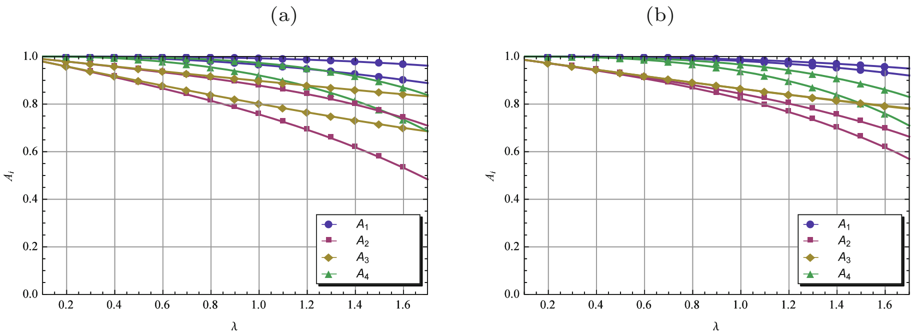

Fig. 2. The availability A_i, $1 \le i \le 4$, for $\beta_1 = 0.2, 0.4$ (a) and $\beta_2 = 0.2, 0.4$ (b) vs. λ

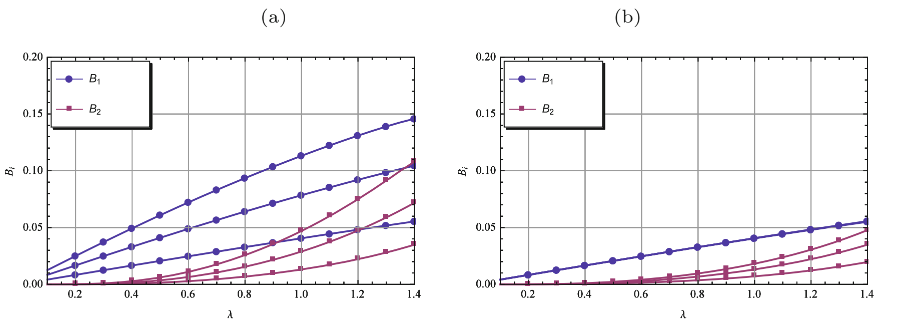

Fig. 3. The failure frequency B_i, $i = 1, 2$, for $\alpha_1 = 0.1, 0.2, 0.3$ (a) and $\alpha_2 = 0.1, 0.2, 0.3$ (b) vs. λ

β_1 influences the busy state of server 2 due to the threshold policy, which in turn makes a contribution to the availability A_3.

In Figs. 3 and 4 we plot the failure frequency B_l for

$$\alpha_l = \{0.1, 0.2, 0.3\} \text{ and } \beta_l = \{0.2, 0.3, 0.4\},\ l = 1, 2,$$

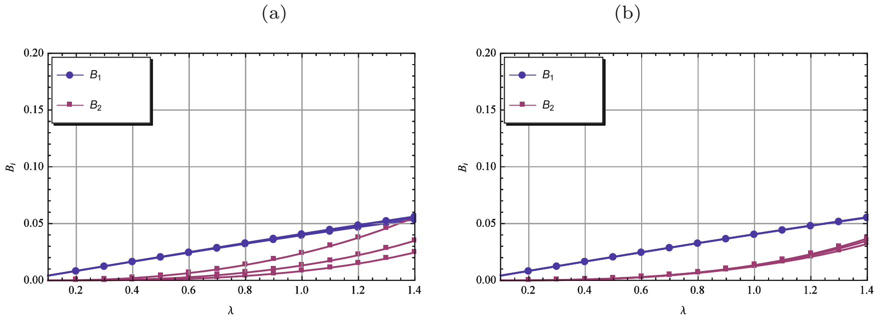

Fig. 4. The failure frequency $B_i, i = 1, 2$, for $\beta_1 = 0.2, 0.3, 0.4$ (a) and $\beta_2 = 0.2, 0.3, 0.4$ (b) vs. λ

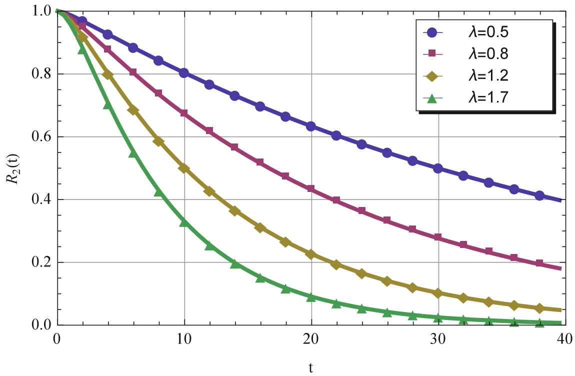

Fig. 5. The function $R(t)$ vs. λ

respectively. These characteristics monotonously increase by increasing of λ. Moreover we notice that $B_1 > B_2$, since the probability to be in state x with $d_1(x) = 1$ is higher than the probability for $d_2(x) = 1$, since server 2 is used according to the threshold control policy. We observe that the function B_1 is insensitive to changes of α_2, β_1 and β_2, and the function B_2 is almost insensitive to change of β_2.

In Fig. 5 we analyze the effect of the arrival rate λ to the reliability function $R(t)$. To evaluate this function we have used a numerical inversion algorithm for the corresponding Laplace transforms $\tilde{R}(s)$, which must be calculated in symbolic form. For the calculations we have used the program *Mathematica* of the Wolfram Research. This program has some limitation on the volume of symbolic representations. Due to this reason and in order to reduce the algorithm's evaluation time, we had to restrict the number of items of the sums in (10) by assuming that $q_1 = 2$ and $q_2 = 1$. We notice that the illustrated function for the higher values of λ exhibit heavier tails.

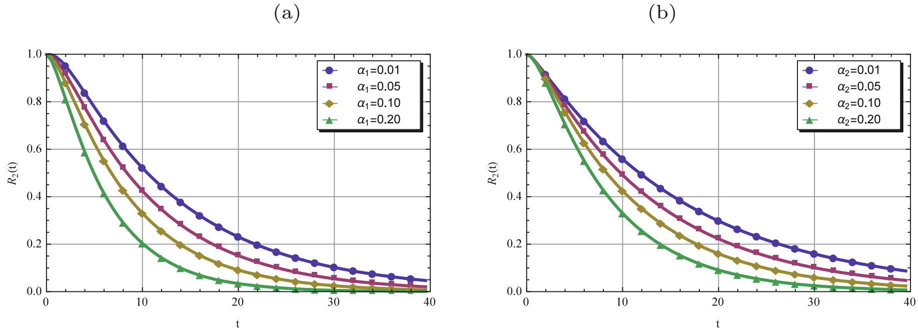

Fig. 6. The function $R(t)$ vs. α_1 (a) and α_2 (b)

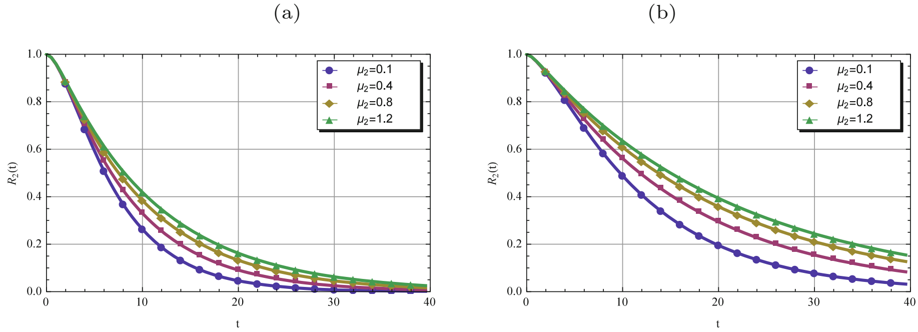

Fig. 7. The function $R(t)$ vs. μ_2 for $\mu_1 = 2.4$ (a) and $\mu_1 = 4.8$ (b)

In Figs. 6 and 7 we illustrate respectively the influence of α_1, α_2, μ_1 and μ_2 on the reliability function $R(t)$. Obviously, for

$$\alpha_1 = 0.01,\ \alpha_2 = 0.01,\ \mu_1 = 4.8,\ \mu_2 = 1.2$$

we observe that the corresponding distribution function exhibits a heavier tail. Finally, we calculate the moment of the life time $\mathbb{E}[T]$ by varying λ,

$$\lambda = \{0.5, 0.8, 1.2, 1.7\},\ \mathbb{E}[T] = \{42.81, 23.51, 13.81, 9.03\}.$$

As is to be expected, the mean life time is decreasing function of λ.

5 Conclusion

The paper provides reliability analysis of a two-server heterogeneous unreliable queueing system with a threshold control policy for the allocation of customers between the servers. The proposed results complement the classical performance analysis of the unreliable queueing models which can be described by the quasi-birth-and-death processes. The matrix-geometric solution method has been used to obtain the stationary state probabilities and some stationary reliability measures like availability and failure frequency. The combination of the forward-elimination-backward-substitution method for the boundary states with generating function approach for the states above the highest threshold level has led to a closed form solution in terms of Laplace transform for the reliability function and as a consequence for the mean time to the first failure. We finally performed numerical experiments to explore the effect of various system parameters on reliability of servers.

References

1. Efrosinin, D.: Controlled queueing systems with heterogeneous servers. In: Dynamic Optimization and Monotonicity Properties, Saarbrücken. VDM (2008)
2. Efrosinin, D.: Queueing model of a hybrid channel with faster link subject to partial and complete failures. Ann. Oper. Res. **202**, 75–102 (2013)
3. Efrosinin, D., Samouylov, K., Gudkova, I.: Busy period analysis of a queueing system with breakdowns and its application to wireless network under licensed shared access regime. In: Galinina, O., Balandin, S., Koucheryavy, Y. (eds.) NEW2AN/ruSMART -2016. LNCS, vol. 9870, pp. 426–439. Springer, Cham (2016). doi:10.1007/978-3-319-46301-8_36
4. Efrosinin, D., Sztrik, J.: Optimal control of a two-server heterogeneous queueing system with breakdowns and constant retrials. In: Dudin, A., Gortsev, A., Nazarov, A., Yakupov, R. (eds.) ITMM 2016. CCIS, vol. 638, pp. 57–72. Springer, Cham (2016). doi:10.1007/978-3-319-44615-8_5
5. Koole, G.: A simple proof of the optimality of a threshold policy in a two-server queueing system. Syst. Control Lett. **26**, 301–303 (1995)
6. Kumar, B.K., Madheswari, S.P.: An $M/M/2$ queueing system with heterogeneous servers and multiple vacations. Math. Comput. Model. **41**, 1415–1429 (2005)
7. Kumar, B.K., Madheswari, S.P., Venkatakrishnan, K.S.: Transient solution of an $M/M/2$ queue with heterogeneous servers subject to catastrophes. Inf. Manag. Sci. **18**(1), 63–80 (2007)
8. Legros, B., Jouini, O.: Routing in a queueing system with two heterogeneous servers in speed and in quality of resolution (2016). http://www.lgi.ecp.fr/~jouini
9. Levy, J., Yehiali, U.: An $M/M/s$ queue with servers' vacations. Inf. Syst. Oper. Res. **14**(2), 153–163 (1976)
10. Lin, W., Kumar, P.R.: Optimal control of a queueing system with two heterogeneous servers. IEEE Trans. Autom. Control **29**, 696–703 (1984)
11. Mitrany, I.L., Avi-Itzhak, B.: A many server queue with service interruptions. Oper. Res. **16**, 628–638 (1967)
12. Neuts, M.F.: Matrix-Geometric Solutions in Stochastic Models. The John Hopkins University Press, Baltimore (1981)
13. Neuts, M.F., Lucantoni, D.M.: A Markovian queue with N servers subject to breakdowns and repairs. Manag. Sci. **25**, 849–861 (1979)
14. Ozkan, E., Kharoufeh, J.: Optimal control of a two-server queueing system with failures. Probab. Eng. Inf. Sci. **28**, 489–527 (2014)
15. Rykov, V., Efrosinin, D.: On the slow server problem. Autom. Remote Control **70**(12), 2013–2013 (2009)
16. Yue, D., Yue, W., Yu, J., Tian, R.: A heterogeneous two-server queueing system with balking and server breakdowns. In: Proceedings of the Eighth International Symposium on Operations Research and Its Applications, China, pp. 230–244 (2009)