Queueing System with Two Unreliable Servers and Backup Server as a Model of Hybrid Communication System

Valentina Klimenok[1], Alexander Dudin[1](✉), and Vladimir Vishnevsky[2]

[1] Department of Applied Mathematics and Computer Science, Belarusian State University, 220030 Minsk, Belarus
{klimenok,dudin}@bsu.by

[2] Institute of Control Sciences of Russian Academy of Sciences and Closed Corporation "Information and Networking Technologies", Moscow, Russia
vishn@inbox.ru

Abstract. In this paper, we analyze a queueing system with two main unreliable servers and backup reliable server. The input flow is defined by the $BMAP$ (Batch Markovian Arrival Process). Heterogeneous breakdowns arrive to the main servers according to a $MMAP$ (Marked Markovian Arrival Process). Service times and repair times have PH (Phase type) distribution. The queueing system under consideration is an adequate model of operation of hybrid communication systems which combine the use of Free Space Optics and radio technologies. We derive a condition for the stable operation of the system, compute its stationary distribution and the key performance measures. Illustrative numerical examples give some insight into the behavior of the system.

Keywords: Unreliable queueing system · Heterogeneous servers · Backup server · Stationary distribution · Stationary performance measures

1 Introduction

The rapid and continuous increase in the number of users on the Internet, the increase in the volume and quality of information transmitted in broadband wireless networks requires a dramatic increase in the performance of multimedia communication channels. In this regard, in recent years, within the frame of the development of next generation networks, intensive research is being carried out to improve wireless performance. One of the directions for creating ultra-high-speed (up to 10 Gbit/s) and reliable wireless communications is the development of hybrid systems based on laser and radio technologies.

The FSO (Free Space Optics) technology has been widely used in recent times. This technology is based on the transmission of data by modulated radiation in the infrared (or visible) part of the spectrum through the atmosphere and their subsequent detection by an optical photo-detector device. The main advantages of atmospheric optical communication lines are as follows.

V. M. Vishnevskiy et al. (Eds.): DCCN 2020, LNCS 12563, pp. 176–195, 2020.
https://doi.org/10.1007/978-3-030-66471-8_15

- High bandwidth and quality of digital communication. Modern FSO-solutions can provide a transmission speed of digital flows up to 10 Gbit/s with a bit error rate of 10^{-12}, which is currently impossible to achieve with any other wireless technologies;
- High security of the channel from unauthorized access and stealth. No wireless transmission technology can offer such confidentiality of communication as laser. Absence of pronounced external signs (basically, this electromagnetic radiation) allows to hide not only the transmitted information, but also the very fact of information exchange. Therefore, laser systems are often used for a variety of applications where high confidentiality of data transmission is required, including financial, medical and military organizations;
- High level of noise immunity and noninterference. FSO-equipment is immune to radio interference and does not create an interference itself;
- Speed and easiness of deployment of the FSO network.

Along with these advantages of wireless optical systems, their main disadvantages are also known: the dependence of the accessibility of the communication channel on weather conditions and the need to provide direct visibility between the optical transmitter and the receiver. Unfavorable weather conditions, such as snow, fog, can significantly reduce the effective range of operation of laser atmospheric communication lines. Thus, the attenuation of a signal in an optical channel in a strong fog can reach a critical value of 50–100 dB/km. Therefore, in order to achieve operator reliability values of the FSO communication channel, it is necessary to resort to the use of hybrid solutions.

Hybrid radio-optical equipment is based on the use of redundant radio channels (centimeter and/or millimeter range of radio waves) together with an optical channel. Note that the operation of the radio channel of the centimeter range of radio waves is practically independent of the weather. The performance of a millimeter-wave wireless channel is not affected by fog. At the same time, the signal/noise ratio, which determines the quality of channel's operation, is greatly reduced with heavy rain. This complementary behavior of optical and broadband radio channels has made it possible to put forward the concept of hybrid carrier-class systems that function reliably in all weather conditions.

Due to the high need for high-speed and reliable communication channels, the following architectures of hybrid systems are currently being used to solve the "last mile" problem (see [1–6]): a high-speed laser channel is reserved by a broadband radio channel operating under the IEEE 802.11n protocol in the centimeter band of radio waves ("cold" or "hot" reserve); The FSO channel is reserved by the radio channel of the millimeter-wave E-band of radio waves (71–76 GHz, 81–86 GHz); The FSO channel and the radio channel of the millimeter band operate in parallel and are reserved by the channel IEEE 802.11n, which is in the cold reserve. In [7] the authors consider a hybrid communication system consisting of FSO links supported by terrestrial optical fiber connections.

Practical needs stimulated theoretical studies on the performance evaluation and the selection of optimal modes for the operation of hybrid systems using queueing theory models with unreliable service channels. Initially, these papers, see, e.g. [8–11] used simplified assumptions about the Poisson character of the input flow, flow of breakdowns, the exponential distribution of packet service time and repair time of communication channels.

Papers from [8] are mainly focused on the study of stationary reliability characteristics, methods and algorithms for optimal channel switching in hybrid communication systems by means of simulation. In the paper [10] the authors consider the hybrid communication system with so called hot redundancy where the FSO channel and backup IEEE 802.11n channel transmit data in parallel. In the paper [9], the hybrid communication system with cold redundancy is investigated. Here the radio-wave link is assumed to be absolutely reliable and backs up the FSO channel in cases when the latter interrupts its functioning because of the unfavorable weather conditions. In the paper a statistical analysis of meteorological data for duration of the periods of favorable and unfavorable weather conditions is also carried out. The paper [11] deals with hybrid communication system consisting of the FSO channel and millimeter-wave radio channel. It is assumed that periods of favorable weather conditions for both channels alternate with periods of unfavorable weather conditions for one of the channels. To model this system, the authors consider two-channel queueing system with unreliable heterogeneous servers which fail alternately.

In further works [12–14], more complicated models of unreliable single-server queues are considered. They generalize models of [9–11] to the case of much more adequate processes describing the operation of corresponding hybrid communication systems. The input flow and the flow of breakdowns are described by Markovian Arrival Processes ($BMAP$ and MAP), see [15], and packet transmission time via communication channels and repair time are assumed to have Phase type (PH)-distributions, see [16]. Although these assumptions complicate the study of models that adequately describe the operation of hybrid systems, but they allow to take into account the non-stationary, correlated nature of information flows in modern and future 5G networks.

Almost in all previous papers, the subject for study were hybrid communication systems consisting of the main FSO channel and backup low speed radio channel. Such systems were modeled by single-server queues with a backup server. One way to increase the reliability and speed of information transmission is to create hybrid system consisting of two main unreliable but high-speed channels (FSO channel and a radio channel of millimeter-wave) which are reserved by reliable but low-speed radio channel IEEE 802.11n which is in the cold reserve. The unreliability of the main channels is due to the lack of favorable weather conditions: the FSO channel can not transmit information in poor visibility conditions and the millimeter-wave channel can not transmit information when precipitation occurs. Such a hybrid system can be modeled by unreliable two-server queueing system with a backup server. In the present paper, we consider such a queueing system. We assume that customers arrive into the system in batch correlated flow $BMAP$, heterogeneous breakdowns arrive to the main servers in $MMAP$ (Marked Markovian Arrival Process, see [17]), service and repair times have PH distribution. We investigate the operation of the system in steady state, derive a stability condition, compute the stationary distribution and performance measures of the system. We also present illustrative numerical examples which give some insight into the behavior of the system.

2 Mathematical Model

We consider a queueing system with infinite waiting room and two unreliable heterogeneous servers, which model FSO and millimeter-wave channels, and backup reliable server which models radio-wave IEEE802 channel. In the following, FSO channel will be called as server 1, millimeter-wave channel as server 2 and radio-wave IEEE802 channel as server 3. Customers arrive into the system in accordance with a $BMAP$. The $BMAP$ is a very general arrival process which is able to capture correlation and burstiness that are commonly seen in the traffic of modern communication networks. The $BMAP$ is defined by the underlying process ν_t, $t \geq 0$, which is an irreducible continuous-time Markov chain with finite state space $\{0, \ldots, W\}$, and the matrix generating function $D(z) = \sum_{k=0}^{\infty} D_k z^k$, $|z| \leq 1$. The batches of customers enter the system only at the epochs of the chain ν_t, $t \geq 0$, transitions. The $(W+1) \times (W+1)$ matrices D_k, $k \geq 1$, and non-diagonal entries of the matrix D_0 define the rates of the process ν_t, $t \geq 0$, transitions which are accompanied by generating the k-size batch of customers, $k \geq 0$. The intensity (fundamental rate) of the $BMAP$ is defined as $\lambda = \boldsymbol{\theta} D'(1)\mathbf{e}$ where the vector $\boldsymbol{\theta}$ is the unique solution of the system $\boldsymbol{\theta} D(1) = \mathbf{0}$, $\boldsymbol{\theta}\mathbf{e} = 1$. Hereinafter $\mathbf{e}$ is a column vector of 1's and $\mathbf{0}$ is a row vector of 0's. For more information about the $BMAP$ see, e.g. [15].

If the arriving customer meets both servers 1 and 2 idle, it starts service at server 1. If the arriving customer meets one of the servers 1 or 2 idle, it starts service at the idle server. If both servers are busy at the customers arrival moment, the customer moves to the buffer.

The service time of a customer by the jth server, $j = 1, 2, 3$, has PH type distribution with irreducible representation $(\boldsymbol{\beta}_j, S_j)$. Such a service time can be interpreted as the time until the continuous-time Markov chain $m_t^{(j)}, t \geq 0$, reaches the single absorbing state $M_j + 1$ if it has the state space $\{1, .., M_j, M_j + 1\}$, initial state is selected according to the vector $\boldsymbol{\beta}_j$, transitions within non-absorbing states are governed by the sub-generator S and the rates of transitions into the absorbing state are given by the vector $\mathbf{S}_0^{(j)} = -S_j\mathbf{e}$. The mean service time is calculated as $b_1^{(j)} = \boldsymbol{\beta}_j(-S_j)^{-1}\mathbf{e}$ and the service rate is equal to $\mu_j = b_1^{(j)^{-1}}$. More detailed description of the PH type process can be found in [16].

Breakdowns of two types arrive to the servers 1, 2 according to a $MMAP$ which is defined by the underlying process η_t, $t \geq 0$, with state space $\{0, \ldots, V\}$ and by the matrices H_0, H_1, H_2. The matrix H_0 defines the rates of the process η_t, $t \geq 0$, transitions which do not lead to generation of a breakdown. The matrix H_j defines the rates of the η_t, $t \geq 0$, transitions which are accompanied by generating a breakdown which is directed to the server j, for $j = 1, 2$. The rate of breakdowns directed to the jth server is calculated as $h_j = \boldsymbol{\vartheta} H_j \mathbf{e}$ where the vector $\boldsymbol{\vartheta}$ is the unique solution of the system $\boldsymbol{\theta}(H_0 + H_1 + H_2) = \mathbf{0}$, $\boldsymbol{\vartheta}\mathbf{e} = 1$.

When a breakdown breaks one of the main servers, the repair period at this server starts immediately and the other main server, if it is available, begins the service of the interrupted customer as a new. If the latter server is busy or under repair, the customer goes to the server 3 and starts its service as a new. However, if during the service time of the customer at the server 3 one of the main servers becomes fault-free, the customer restarts its service on this server. Service at the server 3 is terminated. This server can provide service only when both main servers are broken.

The repair period at the jth main server, $j = 1, 2$, has PH type distribution with an irreducible representation $(\boldsymbol{\tau}_j, T_j)$. The repair process at the jth server is directed by the Markov chain $r_t^{(j)}$, $t \geq 0$, with state space $\{1, \dots, R_j, R_j + 1\}$ where $R_j + 1$ is an absorbing state. The rates of transitions into the absorbing state are given by the vector $\mathbf{T}_0^{(j)} = -T_j \mathbf{e}$. The repair rate is calculated as $\tau_j = (\boldsymbol{\beta}_j(-T_j)^{-1}\mathbf{e})^{-1}$. Breakdowns arriving to a server are ignored it the server is under repair at the moment of the breakdown arrival.

3 Process of the System States

Let, at the moment t,

i_t be the number of customers in the system, $i_t \geq 0$,

$n_t = 0$, if both main servers are fault-free (both ones are busy or idle); $n_t = 0_j$, if both main servers are fault-free, the jth server is busy and the other one is idle, $j = 1, 2$; $n_t = 1$, if the server 1 is under repair; $n_t = 2$, if the server 2 is under repair; $n_t = 3$, if both servers are under repair;

$m_t^{(j)}$ be the state of the directing process of the service at the jth busy server, $j = 1, 2, 3$, $m_t^{(j)} = \overline{1, M_j}$;

$r_t^{(j)}$ be the state of the directing process of the repair time at the jth server, $j = 1, 2$, $r_t^{(j)} = \overline{1, R_j}$;

ν_t and η_t be the states of the directing processes of the $BMAP$ and the $MMAP$, respectively, $\nu_t = \overline{0, W}$, $\eta_t = \overline{0, V}$.

The process of the system states is described by the regular irreducible continuous time Markov chain, $\xi_t, t \geq 0$, with state space

$$X = \{(0, n, \nu, \eta),\ i = 0, n = \overline{0, 3},\ \nu = \overline{0, W},\ \eta = \overline{0, V}\} \bigcup$$

$$\{(i, 0_j, \nu, \eta, m^{(j)}),\ i = 1, j = 1, 2, n = 0_j, \nu = \overline{0, W}, \eta = \overline{0, V},\ m^{(j)} = \overline{1, M_j}\} \bigcup$$

$$\{(i, 0, \nu, \eta, m^{(1)}, m^{(2)}),\ i > 1, \nu = \overline{0, W}, \eta = \overline{0, V}, m^{(k)} = \overline{1, M_k}, k = 1, 2\} \bigcup$$

$$\{(i, 1, \nu, \eta, m^{(2)}, r^{(1)}), i \geq 1, \nu = \overline{0, W}, \eta = \overline{0, V}, m^{(2)} = \overline{1, M_2}, r^{(1)} = \overline{1, R_1}\} \bigcup$$

$$\{(i, 2, \nu, \eta, m^{(1)}, r^{(2)}), i \geq 1, \nu = \overline{0, W}, \eta = \overline{0, V}, m^{(1)} = \overline{1, M_1}, r^{(1)} = \overline{1, R_2}\} \bigcup$$

$$\{(i,3,\nu,\eta,m^{(3)},r^{(1)},r^{(2)}),\ i>0,\nu=\overline{0,W},\eta=\overline{0,V},m^{(3)}=\overline{1,M_3},$$
$$r^{(j)}=\overline{1,R_j},j=1,2\}.$$

Let $Q_{i,j}$, $i,j\geq 0$, be the matrices formed by the rates of the chain transitions from the states corresponding to the value i of the component i_t to the states corresponding to the value j of this component. The following statement is true.

Lemma 1. *Infinitesimal generator of the Markov chain ξ_t, $t\geq 0$, has the following block structure*

$$Q=\begin{pmatrix} Q_{0,0} & Q_{0,1} & Q_{0,2} & Q_{0,3} & Q_{0,4} & \cdots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & Q_{1,3} & Q_{1,4} & \cdots \\ O & Q_{2,1} & Q_1 & Q_2 & Q_3 & \cdots \\ O & O & Q_0 & Q_1 & Q_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where non-zero blocks are of the following form:

$$Q_{0,0}=\begin{pmatrix} D_0\oplus H_0 & I_{\bar W}\otimes H_1\otimes\boldsymbol{\tau}_1 & I_{\bar W}\otimes H_2\otimes\boldsymbol{\tau}_2 & O \\ I_a\otimes\boldsymbol{T}_0^{(1)} & D_0\oplus(H_0+H_1)\oplus T^{(1)} & O & I_{\bar W}\otimes H_2\otimes I_{R_1}\otimes\boldsymbol{\tau}_2 \\ I_a\otimes\boldsymbol{T}_0^{(2)} & O & D_0\oplus(H_0+H_2)\oplus T^{(2)} & I_{\bar W}\otimes H_1\otimes I_{R_2}\otimes\boldsymbol{\tau}_1 \\ O & I_{aR_1}\otimes\boldsymbol{T}_0^{(2)} & I_a\otimes\boldsymbol{T}_0^{(1)}\otimes I_{R_2} & D_0\oplus H\oplus T^{(1)}\oplus T^{(2)} \end{pmatrix},$$

$$Q_{0,1}=$$

$$\begin{pmatrix} D_1\otimes I_{\bar V}\otimes\boldsymbol{\beta}_1 & O_{a\times aM_2} & O & O & O \\ O & O & D_1\otimes I_{\bar V}\otimes\boldsymbol{\beta}_2\otimes I_{R_1} & O & O \\ O & O & O & D_1\otimes I_{\bar V}\otimes\boldsymbol{\beta}_1\otimes I_{R_2} & O \\ O & O & O & O & D_1\otimes I_{\bar V}\otimes\boldsymbol{\beta}_3\otimes I_{R_1R_2} \end{pmatrix},$$

$$Q_{0,k}=diag\{D_k\otimes I_{\bar V}\otimes\boldsymbol{\beta}_1\otimes\boldsymbol{\beta}_2,\ D_k\otimes I_{\bar V}\otimes\boldsymbol{\beta}_2\otimes I_{R_1},$$
$$D_k\otimes I_{\bar V}\otimes\boldsymbol{\beta}_1\otimes I_{R_2},\ D_k\otimes I_{\bar V}\otimes\boldsymbol{\beta}_3\otimes I_{R_1R_2}\},$$

$$Q_{1,0}=\begin{pmatrix} I_a\otimes\boldsymbol{S}_0^{(1)} & O & O & O \\ I_a\otimes\boldsymbol{S}_0^{(2)} & O & O & O \\ O & I_a\otimes\boldsymbol{S}_0^{(2)}\otimes I_{R_1} & O & O \\ O & O & I_a\otimes\boldsymbol{S}_0^{(1)}\otimes I_{R_2} & O \\ O & O & O & I_a\otimes\boldsymbol{S}_0^{(3)}\otimes I_{R_1}\otimes I_{R_2} \end{pmatrix},$$

$$Q_{1,1}=\begin{pmatrix} Q_{1,1}^{(1)} & Q_{1,1}^{(2)} \end{pmatrix},$$

$$Q_{1,1}^{(1)}=\begin{pmatrix} D_0\oplus H_0\oplus S_1 & O & I_{\bar W}\otimes H_1\otimes\mathbf{e}_{M_1}\otimes\boldsymbol{\beta}_2\otimes\boldsymbol{\tau}_1 \\ O & D_0\oplus H_0\oplus S_2 & I_{\bar W}\otimes H_1\otimes I_{M_2}\otimes\boldsymbol{\tau}_1 \\ O & I_a\otimes I_{M_2}\otimes\boldsymbol{T}_0^{(1)} & D_0\oplus(H_0+H_1)\oplus S_2\oplus T_1 \\ I_a\otimes I_{M_1}\otimes\boldsymbol{T}_0^{(2)} & O & O \\ O & O & I_a\otimes\boldsymbol{\beta}_2\otimes\mathbf{e}_{M_3}\otimes I_{R_1}\otimes\boldsymbol{T}_0^{(2)} \end{pmatrix},$$

$$Q_{1,1}^{(2)} = \begin{pmatrix} I_{\bar{W}} \otimes H_2 \otimes I_{M_1} \otimes \boldsymbol{\tau}_2 & O \\ I_{\bar{W}} \otimes H_2 \otimes \mathbf{e}_{M_2} \otimes \boldsymbol{\beta}_1 \otimes \boldsymbol{\tau}_2 & O \\ O & I_{\bar{W}} \otimes H_2 \otimes \mathbf{e}_{M_2} \otimes \boldsymbol{\beta}_3 \otimes I_{R_1} \otimes \boldsymbol{\tau}_2 \\ D_0 \oplus (H_0 + H_2) \oplus S_1 \oplus T_2 & I_{\bar{W}} \otimes H_1 \otimes \mathbf{e}_{M_1} \otimes \boldsymbol{\beta}_3 \otimes \boldsymbol{\tau}_1 \otimes I_{R_2} \\ I_a \otimes \boldsymbol{\beta}_1 \otimes \mathbf{e}_{M_3} \otimes \boldsymbol{T}_0^{(1)} \otimes I_{R_2} & D_0 \oplus H \oplus S_3 \oplus T_1 \oplus T_2 \end{pmatrix},$$

$$Q_{1,k} =$$

$$\begin{pmatrix} D_{k-1} \otimes I_{\bar{V}M_1} \otimes \boldsymbol{\beta}_2 & O & O & O \\ D_{k-1} \otimes I_{\bar{V}} \otimes \boldsymbol{\beta}_1 \otimes I_{M_2} & O & O & O \\ O & D_{k-1} \otimes I_{\bar{V}M_2R_1} & O & O \\ O & O & D_{k-1} \otimes I_{\bar{V}M_1R_2} & O \\ O & O & O & D_{k-1} \otimes I_{\bar{V}M_3R_1R_2} \end{pmatrix},$$

$$Q_{2,1} = \begin{pmatrix} I_a \otimes I_{M_1} \otimes \boldsymbol{S}_0^{(2)} & O \\ O & O \end{pmatrix} + \Big(O \mid diag\{I_a \otimes \boldsymbol{S}_0^{(1)} \otimes I_{M_2},$$

$$I_a \otimes \boldsymbol{S}_0^{(2)} \boldsymbol{\beta}_2 \otimes I_{R_1},\, I_a \otimes \boldsymbol{S}_0^{(1)} \boldsymbol{\beta}_1 \otimes I_{R_2},\, I_a \otimes \boldsymbol{S}_0^{(3)} \boldsymbol{\beta}_3 \otimes I_{R_1R_2}\} \Big),$$

$$Q_0 = diag\{I_a \otimes (\boldsymbol{S}_0^{(1)} \boldsymbol{\beta}_1 \oplus \boldsymbol{S}_0^{(2)} \boldsymbol{\beta}_2),\, I_a \otimes \boldsymbol{S}_0^{(2)} \boldsymbol{\beta}_2 \otimes I_{R_1},$$

$$I_a \otimes \boldsymbol{S}_0^{(1)} \boldsymbol{\beta}_1 \otimes I_{R_2},\, I_a \otimes \boldsymbol{S}_0^{(3)} \boldsymbol{\beta}_3 \otimes I_{R_1R_2}\},$$

$$Q_1 = \begin{pmatrix} Q_1^{(1,1)} & Q_1^{(1,2)} \\ Q_1^{(2,1)} & Q_1^{(2,2)} \end{pmatrix},$$

$$Q_1^{(1,1)} = \begin{pmatrix} D_0 \oplus H_0 \oplus S_1 \oplus S_2 & I_{\bar{W}} \otimes H_1 \otimes \mathbf{e}_{M_1} \otimes I_{M_2} \otimes \boldsymbol{\tau}_1 \\ I_a \otimes \boldsymbol{\beta}_1 \otimes I_{M_2} \otimes \boldsymbol{T}_0^{(1)} & D_0 \oplus (H_0 + H_1) \oplus S_2 \oplus T_1 \end{pmatrix},$$

$$Q_1^{(1,2)} = \begin{pmatrix} I_{\bar{W}} \otimes H_2 \otimes I_{M_1} \otimes \mathbf{e}_{M_2} \otimes \boldsymbol{\tau}_2 & O \\ O & I_{\bar{W}} \otimes H_2 \otimes \mathbf{e}_{M_2} \otimes \boldsymbol{\beta}_3 \otimes I_{R_1} \otimes \boldsymbol{\tau}_2 \end{pmatrix},$$

$$Q_1^{(2,1)} = \begin{pmatrix} I_a \otimes I_{M_1} \otimes \boldsymbol{\beta}_2 \otimes \boldsymbol{T}_0^{(2)} & O \\ O & I_a \otimes \boldsymbol{\beta}_2 \otimes \mathbf{e}_{M_3} \otimes I_{R_1} \otimes \boldsymbol{T}_0^{(2)} \end{pmatrix},$$

$$Q_1^{(2,2)} = \begin{pmatrix} D_0 \oplus (H_0 + H_2) \oplus S_1 \oplus T_2 & I_{\bar{W}} \otimes H_1 \otimes \mathbf{e}_{M_1} \otimes \boldsymbol{\beta}_3 \otimes I_{R_2} \otimes \boldsymbol{\tau}_1 \\ I_a \otimes \boldsymbol{\beta}_1 \otimes \mathbf{e}_{M_3} \otimes \boldsymbol{T}_0^{(1)} \otimes I_{R_2} & D_0 \oplus H \oplus S_3 \oplus T_1 \oplus T_2 \end{pmatrix},$$

$$Q_{k+1} = diag\{D_k \otimes I_{\bar{V}M_1M_2},\, D_k \otimes I_{\bar{V}M_2R_1},\, D_k \otimes I_{\bar{V}M_1R_2},\, D_k \otimes I_{\bar{V}M_3R_1R_2}\},\ k \geq 1,$$

where $H = H_0 + H_1$, $\otimes, \oplus$ *are the symbols of Kronecker's product and sum of matrices,* $diag\{\dots\}$ *denotes the block diagonal matrix with the diagonal blocks listed in the brackets,* $\bar{W} = W + 1$, $\bar{V} = V + 1$, $a = \bar{W}\bar{V}$, $\mathbf{e}_n$ *is a column vector of size* n*, consisting of 1's,* I (O) *is an identity (zero) matrix.*

Corollary 1. *The Markov chain* $\xi_t, t \geq 0$*, belongs to the class of continuous time quasi-Toeplitz Markov chains, see [18].*

Proof. The generator Q of the chain $\xi_t, t \geq 0$, has a block upper-Hessenberg structure and, starting from $i = 3$, the blocks $Q_{i,j}$ depend on i, j only through the difference $j-i$. Then, according the definition given in [18] the chain $\xi_t, t \geq 0$, is a quasi-Toeplitz Markov chain.

In what follows we need expressions for the matrix generating functions $Q^{(n)}(z) = \sum\limits_{k=2}^{\infty} Q_{n,k} z^k$, $n = 0, 1$, and $Q(z) = \sum\limits_{k=0}^{\infty} Q_k z^k$, $|z| \leq 1$. These expressions are given by the following

Corollary 2. *The matrix generation functions* $Q(z)$, $Q^{(n)}(z)$, $n = 0, 1$, *have the following form:*

$$Q^{(0)}(z) = diag\{D(z) - D_0 - D_1 z) \otimes I_{\bar{V}} \otimes \boldsymbol{\beta}_1 \otimes \boldsymbol{\beta}_2,\ (D(z) - D_0 - D_1 z) \otimes I_{\bar{V}} \otimes \boldsymbol{\beta}_2 \otimes I_{R_1},$$
$$(D(z) - D_0 - D_1 z) \otimes I_{\bar{V}} \otimes \boldsymbol{\beta}_1 \otimes I_{R_2},\ (D(z) - D_0 - D_1 z) \otimes I_{\bar{V}} \otimes \boldsymbol{\beta}_3 \otimes I_{R_1 R_2}\}, \tag{1}$$

$$Q^{(1)}(z) = \tag{2}$$
$$z \begin{pmatrix} \bar{D}(z) \otimes I_{\bar{V} M_1} \otimes \boldsymbol{\beta}_2 & O & O & O \\ \bar{D}(z) \otimes I_{\bar{V}} \otimes \boldsymbol{\beta}_1 \otimes I_{M_2} & O & O & O \\ O & \bar{D}(z) \otimes I_{\bar{V} M_2 R_1} & O & O \\ O & O & \bar{D}(z) \otimes I_{\bar{V} M_1 R_2} & O \\ O & O & O & \bar{D}(z) \otimes I_{\bar{V} M_3 R_1 R_2} \end{pmatrix},$$

where $\bar{D}(z) = D(z) - D_0$,

$$Q(z) = Q_0 + \mathcal{Q} z + z diag\{D(z) \otimes I_{\bar{V} M_1 M_2},$$
$$D(z) \otimes I_{\bar{V} M_2 M_3 R_1},\ D(z) \otimes I_{\bar{V} M_1 M_3 R_2},\ D(z) \otimes I_{\bar{V} M_3 R_1 R_2}\}, \tag{3}$$

where the matrix $\mathcal{Q}$ *is of the form*

$$\mathcal{Q} = \begin{pmatrix} \mathcal{Q}^{(1,1)} & \mathcal{Q}^{(1,2)} \\ \mathcal{Q}^{(2,1)} & \mathcal{Q}^{(2,2)} \end{pmatrix}, \tag{4}$$

$$\mathcal{Q}^{(1,1)} = \begin{pmatrix} I_{\bar{W}} \otimes H_0 \oplus S_1 \oplus S_2 & I_{\bar{W}} \otimes H_1 \otimes \mathbf{e}_{M_1} \otimes I_{M_2} \otimes \boldsymbol{\tau}_1 \\ I_{\bar{W}\bar{V}} \otimes \boldsymbol{\beta}_1 \otimes I_{M_2} \otimes \boldsymbol{T}_0^{(1)} & I_{\bar{W}} \otimes (H_0 + H_1) \oplus S_2 \oplus T_1 \end{pmatrix},$$

$$\mathcal{Q}^{(1,2)} = \begin{pmatrix} I_{\bar{W}} \otimes H_2 \otimes I_{M_1} \otimes \mathbf{e}_{M_2} \otimes \boldsymbol{\tau}_2 & O \\ O & I_{\bar{W}} \otimes H_2 \otimes \mathbf{e}_{M_2} \otimes \boldsymbol{\beta}_3 \otimes I_{R_1} \otimes \boldsymbol{\tau}_2 \end{pmatrix},$$

$$\mathcal{Q}^{(2,1)} = \begin{pmatrix} I_{\bar{W}} \otimes I_{\bar{V} M_1} \otimes \boldsymbol{\beta}_2 \otimes \boldsymbol{T}_0^{(2)} & O \\ O & I_{\bar{W}\bar{V}} \otimes \boldsymbol{\beta}_2 \otimes \mathbf{e}_{M_3} \otimes I_{R_1} \otimes \boldsymbol{T}_0^{(2)} \end{pmatrix},$$

$$\mathcal{Q}^{(2,2)} = \begin{pmatrix} I_{\bar{W}} \otimes (H_0 + H_2) \oplus S_1 \oplus T_2 & I_{\bar{W}} \otimes H_1 \otimes \mathbf{e}_{M_1} \otimes \boldsymbol{\beta}_3 \otimes I_{R_2} \otimes \boldsymbol{\tau}_1 \\ I_{\bar{W}\bar{V}} \otimes \boldsymbol{\beta}_1 \otimes \mathbf{e}_{M_3} \otimes \boldsymbol{T}_0^{(1)} \otimes I_{R_2} & I_{\bar{W}} \otimes H \oplus S_3 \oplus T_1 \oplus T_2 \end{pmatrix}.$$

4 Stationary Distribution

Let $\mathcal{Q}^-$ be a matrix obtained from the matrix $\mathcal{Q}$ given by (4) by formally deleting in its blocks the Kronecker cofactor $I_{\bar{W}}$ and Q_0^- be a matrix obtained from the matrix Q_0 by formally replacement in its blocks the cofactor I_a by the cofactor $I_{\bar{W}}$. Let also

$$\Psi = Q_0^- + \mathcal{Q}^-.$$

Theorem 1. *The necessary and sufficient condition for ergodicity of the Markov chain ξ_t, $t \geq 0$, is the fulfillment of the inequality*

$$\lambda < -\boldsymbol{\pi}_0(S_1 \oplus S_2)\mathbf{e} + \boldsymbol{\pi}_1\mathbf{S}_0^{(2)} + \boldsymbol{\pi}_2\mathbf{S}_0^{(1)} + \boldsymbol{\pi}_3\mathbf{S}_0^{(3)} \tag{5}$$

where $\boldsymbol{\pi}_0 = \mathbf{x}_0(\mathbf{e}_{V+1} \otimes I_{M_1M_2})$, $\boldsymbol{\pi}_1 = \mathbf{x}_1(\mathbf{e}_{V+1} \otimes I_{M_2} \otimes \mathbf{e}_{R_1})$, $\boldsymbol{\pi}_2 = \mathbf{x}_2(\mathbf{e}_{V+1} \otimes I_{M_1} \otimes \mathbf{e}_{R_2})$, $\boldsymbol{\pi}_3 = \mathbf{x}_3(\mathbf{e}_{V+1} \otimes I_{M_3} \otimes \mathbf{e}_{R_1R_2})$ and the vectors $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ are sub-vectors of the vector $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$, which is the unique solution of the system of linear algebraic equations

$$\mathbf{x}\Psi = 0, \ \mathbf{xe} = 1. \tag{6}$$

Remark 1. Note that the ratio of the left part of inequality (5) and the right part of this inequality is the system load factor ρ, i.e.

$$\rho = \frac{\lambda}{-\boldsymbol{\pi}_0(S_1 \oplus S_2)\mathbf{e} + \boldsymbol{\pi}_1\mathbf{S}_0^{(2)} + \boldsymbol{\pi}_2\mathbf{S}_0^{(1)} + \boldsymbol{\pi}_3\mathbf{S}_0^{(3)}}.$$

Proof. It can be verified that the matrix $\sum\limits_{k=0}^{\infty} Q_k$ is irreducible. Hence, from [18], the necessary and sufficient condition for ergodicity of the chain ξ_t is the fulfillment of the inequality

$$\mathbf{y}Q'(z)|_{z=1}\mathbf{e} < 0 \tag{7}$$

where the vector $\mathbf{y}$ is the unique solution of the system

$$\mathbf{y}Q(1) = \mathbf{y}, \quad \mathbf{ye} = 1. \tag{8}$$

The theorem will be proven if we show that inequality (7) is equivalent to inequality (5). Let the vector $\mathbf{y}$ be of the form

$$\mathbf{y} = (\boldsymbol{\theta} \otimes \mathbf{x_0}, \boldsymbol{\theta} \otimes \mathbf{x_1}, \boldsymbol{\theta} \otimes \mathbf{x_2}, \boldsymbol{\theta} \otimes \mathbf{x_3}). \tag{9}$$

Substituting the vector $\mathbf{y}$ in the form (9) into (8) and using relation $\boldsymbol{\theta}D(1) = \mathbf{0}$, we reduce system (8) to the form (6).

Now we substitute into the inequality (7) the vector $\mathbf{y}$ in the form (9) and the expression for $Q^{'}(1)$ calculated by formula (3). Taking into account that $\boldsymbol{\theta}D'(1)\mathbf{e} = \lambda$, we transform inequality (7) to the form

$$\lambda + \mathbf{x}\mathcal{Q}^-\mathbf{e} < 0. \tag{10}$$

Using the known relations $H\mathbf{e} = (H_0 + H_1 + H_2)\mathbf{e} = \mathbf{0}$, $S_n\mathbf{e} + \boldsymbol{S}_0^{(n)} = \mathbf{0}$, $T_n\mathbf{e} + \boldsymbol{T}_0^{(n)} = \mathbf{0}$, $n = 1, 2$, we reduce inequality (10) to the following form:

$$\lambda < \mathbf{x}_0(\mathbf{e}_{\bar{V}} \otimes I_{M_1M_2})(\mathbf{S}_0^{(1)} \oplus \mathbf{S}_0^{(2)})\mathbf{e} + \mathbf{x}_1(\mathbf{e}_{\bar{V}} \otimes I_{M_2} \otimes \mathbf{e}_{R_1})\mathbf{S}_0^{(2)}\mathbf{e}$$

$$+\mathbf{x}_2(\mathbf{e}_{\bar{V}} \otimes I_{M_1} \otimes \mathbf{e}_{R_2})\mathbf{S}_0^{(1)}\mathbf{e} + \mathbf{x}_3(\mathbf{e}_{\bar{V}} \otimes I_{M_3} \otimes \mathbf{e}_{R_1R_2})\mathbf{S}_0^{(3)}\mathbf{e}.$$

After using the notation introduced in the statement of the theorem, this inequality takes the form (5). □

Remark 2. In the physical interpretation of the inequality (5), we take into account that this inequality is derived under the system overload condition. Let us consider the physical meaning of the first term in the right-hand side of (5). The component $\boldsymbol{\pi}_0(m^{(1)}, m^{(2)})$ of the row vector $\boldsymbol{\pi}_0$ is the probability that servers 1 and 2 are fault-free and serve customers on the phases $m^{(1)}$ and $m^{(2)}$, respectively. The corresponding component of the column vector $(\mathbf{S}_0^{(1)} \oplus \mathbf{S}_0^{(2)})\mathbf{e}$ is the total service rate by servers 1 and 2 provided that the service on these servers is in the phases $m^{(1)}$ and $m^{(2)}$, respectively. Then the product $\boldsymbol{\pi}_0(\mathbf{S}_0^{(1)} \oplus \mathbf{S}_0^{(2)})\mathbf{e}$ represents the rate of the output flow in periods when customers are served by servers 1 and 2. The other summands of the sum on the right-hand side of inequality (5) are interpreted similarly: the second term is the rate of the output flow when customers are only served by server 2 (server 1 is under repair), the third term is the rate of the output flow when customers are only served by server 1 (server 2 is under repair), the fourth term is the rate of the output flow when customers are only served by server 3 (servers 1 and 2 are under repair). Then the right-hand side of inequality (5) expresses the total rate of the output flow under overload condition. Obviously, for the existence of a steady-state regime in the system, it is necessary and sufficient that the input rate λ be less than the rate of the output flow.

Corollary 3. *In the case of stationary Poisson flow of breakdowns and exponential distribution of service and repair times, ergodicity condition (5) is reduced to the following inequality:*

$$\lambda < \pi_0(\mu_1 + \mu_2) + \pi_1\mu_2 + \pi_2\mu_1 + \pi_3\mu_3, \tag{11}$$

where the vector $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \pi_3)$ *is the unique solution to the system of linear algebraic equation*

$$\boldsymbol{\pi}\begin{pmatrix} -(h_1+h_2) & h_1 & h_2 & 0 \\ \tau_1 & -(h_2+\tau_1) & 0 & h_2 \\ \tau_2 & 0 & -(h_1+\tau_2) & h_1 \\ 0 & \tau_2 & \tau_1 & -(\tau_1+\tau_2) \end{pmatrix} = \mathbf{0}, \quad \boldsymbol{\pi}\mathbf{e} = 1.$$

In what follows we assume that the ergodicity condition given by Theorem 1 is satisfied, which ensures that there exist the stationary probabilities of the system states

$$p_0^{(n)}(\nu, \eta) = \lim_{t\to\infty} P\{i_t = 0, n_t = n, \nu_t = \nu, \eta_t = \eta\},\ n = \overline{0,3},\ \nu = \overline{0,W}, \eta = \overline{0,V};$$

$$p_1^{(0_n)}\{(\nu,\eta,m^{(k)})\} = \lim_{t\to\infty} P\{i_t = i, n_t = 0_n, \nu_t = \nu, \eta_t = \eta, m_t^{(n)} = m^{(n)}\},$$
$$n = 1,2,\ \nu = \overline{0,W},\ \eta = \overline{0,V},\ m^{(n)} = \overline{1,M^{(n)}};$$
$$p_i^{(0)}(\nu,\eta,m^{(1)},m^{(2)}) = \lim_{t\to\infty} P\{i_t = i, n_t = 0, \nu_t = \nu, \eta_t = \eta, m_t^{(1)} = m^{(1)},$$
$$m_t^{(2)} = m^{(2)}\}, i > 1,\ \nu = \overline{0,W},\ \eta = \overline{0,V},\ m^{(n)} = \overline{1,M^{(n)}}.$$
$$p_i^{(1)}(\nu,\eta,m^{(2)},r^{(1)}) = \lim_{t\to\infty} P\{i_t = i, n_t = 1,\ \nu_t = \nu, \eta_t = \eta,\ m_t^{(2)} = m^{(2)},$$
$$r_t^{(1)} = r^{(1)}\},\ i \geq 1, \nu = \overline{0,W},\ \eta = \overline{0,V},\ m^{(2)} = \overline{1,M^{(2)}},\ r^{(1)} = \overline{1,R^{(1)}};$$
$$p_i^{(2)}(\nu,\eta,m^{(1)},r^{(2)}) = \lim_{t\to\infty} P\{i_t = i, n_t = 2,\ \nu_t = \nu, \eta_t = \eta,\ m_t^{(1)} = m^{(1)},$$
$$r_t^{(2)} = r^{(2)}\},\ i \geq 1,\ \nu = \overline{0,W},\ \eta = \overline{0,V},\ m^{(1)} = \overline{1,M^{(1)}},\ r^{(2)} = \overline{1,R^{(2)}};$$
$$p_i^{(3)}(\nu,\eta,m^{(3)},r^{(1)},r^{(2)}) = \lim_{t\to\infty} P\{i_t = i, n_t = 3,\ \nu_t = \nu, \eta_t = \eta,\ m_t^{(3)} = m^{(3)},$$
$$r_t^{(1)} = r^{(1)}, r_t^{(2)} = r^{(2)}\}, i \geq 1,\ \nu = \overline{0,W},\ \eta = \overline{0,V},\ m^{(3)} = \overline{1,M^{(3)}},$$
$$r^{(1)} = \overline{1,R^{(1)}},\ r^{(2)} = \overline{1,R^{(2)}}.$$

Within each selected group, we order the probabilities in the lexicographic order of the components and form the vectors of these probabilities

$$\mathbf{p}_0^{(n)}, n = \overline{0,3};\ \mathbf{p}_1^{(0_1)},\ \mathbf{p}_1^{(0_2)},\ \mathbf{p}_i^{(n)}, n = \overline{0,3},\ i \geq 1.$$

Next, we form the vectors $\mathbf{p}_i$ of stationary probabilities corresponding to the values i of the denumerable component as follows:

$$\mathbf{p}_0 = (\mathbf{p}_0^{(0)},\ \mathbf{p}_0^{(1)},\ \mathbf{p}_0^{(2)},\ \mathbf{p}_0^{(3)}),\ \mathbf{p}_1 = (\mathbf{p}_1^{(0_1)},\ \mathbf{p}_1^{(0_2)},\ \mathbf{p}_1^{(1)},\ \mathbf{p}_1^{(2)},\ \mathbf{p}_1^{(3)}),$$
$$\mathbf{p}_i = (\mathbf{p}_i^{(0)},\ \mathbf{p}_i^{(1)},\ \mathbf{p}_i^{(2)},\ \mathbf{p}_i^{(3)}),\ i \geq 2.$$

These vectors are calculated using algorithm for calculating the stationary distribution of quasi-Toeplitz Markov chains, see [18].

5 Vector Generating Function of the Stationary Distribution. System Performance Characteristics

Calculating the stationary probability vectors $\mathbf{p}_i$, $i \geq 0$, we can also calculate various characteristics of the system performance. In the calculation process, the following result will be useful.

Lemma 2. *Vector generating function* $\boldsymbol{P}(z) = \sum\limits_{i=0}^{\infty} \mathbf{p}_i z^i$, $|z| \leq 1$, *satisfies the following equation:*

$$P(z)Q(z) = \mathcal{B}(z) \tag{12}$$

where

$$\mathcal{B}(z) = \mathbf{p}_0 Q(z) + z[\mathbf{p}_1 Q(z) - \mathbf{p}_0 Q^{(0)}(z) - \mathbf{p}_1 Q^{(1)}(z)] + z^2 \mathbf{p}_2 Q_0. \tag{13}$$

Formula (12) can be used to calculate the values of the function $\boldsymbol{P}(z)$ and its derivatives at the point $z = 1$ without calculating infinite sums. The obtained values allow to find the moments of the number of customers in the system and some other characteristics of the system. Note that it is not possible to calculate directly the value of $\boldsymbol{P}(z)$ and its derivatives at the point $z = 1$ from Eq. (12) since the matrix $Q(1)$ is singular. This difficulty can be overcome by using the recursion formulas given below in Corollary 4.

Let us denote $f^{(m)}(z)$ the mth derivative of the function $f(z)$, $m \geq 1$, and $f^{(0)}(z) = f(z)$.

Corollary 4. *The mth, $m \geq 0$, derivatives of the vector generating function $\boldsymbol{P}(z)$ at the point $z = 1$ are recursively calculated from the following system of linear algebraic equations:*

$$\begin{cases} \boldsymbol{P}^{(m)}(1)Q(1) = \mathcal{B}^{(m)}(1) - \sum\limits_{l=0}^{m-1} C_m^l \boldsymbol{P}^{(l)}(1)Q^{(m-l)}(1), \\ \boldsymbol{P}^{(m)}(1)Q'(1)\mathbf{e} = \frac{1}{m+1}[\mathcal{B}^{(m+1)}(1) - \sum\limits_{l=0}^{m-1} C_{m+1}^l \boldsymbol{P}^{(l)}(1)Q^{(m+1-l)}(1)]\mathbf{e}, \end{cases}$$

where the derivatives $\mathcal{B}^{(m)}(1)$ are calculated using formula (13) and expressions (1)–(3) for the vector generator functions $Q(z), Q^{(0)}(z), Q^{(1)}(z)$.

The proof of the corollary is parallel to the one outlined in [19] and is omitted here.

Having the stationary distribution $\boldsymbol{p}_i$, $i \geq 0$, been calculated we find a number of important stationary performance measures of the system and examine their behavior through the numerical experiments.

- Throughput of the system (the maximum rate of the flow that can be processed by the system)

$$\varrho = -\boldsymbol{\pi}_0(S_1 \oplus S_2)\mathbf{e} + \boldsymbol{\pi}_1\mathbf{S}_0^{(2)} + \boldsymbol{\pi}_2\mathbf{S}_0^{(1)} + \boldsymbol{\pi}_3\mathbf{S}_0^{(3)}.$$

- Mean number of customers in the system $L = \boldsymbol{P}'(1)\mathbf{e}$.
- Variance of the number of customers in the system $V = \boldsymbol{P}^{''}(1)\mathbf{e} - L^2 + L$.
- Probability that i customers stay in the system $p_i = \mathbf{p}_i\mathbf{e}$.
- Probability $P_i^{(0)}$ that i customers stay in the system and both servers are fault- free

$$P_0^{(0)} = \mathbf{p}_0 \begin{pmatrix} \mathbf{e}_a \\ \mathbf{0}^T \end{pmatrix}, \; P_1^{(0)} = \mathbf{p}_1 \begin{pmatrix} \mathbf{e}_{a(M_1+M_2} \\ \mathbf{0}^T \end{pmatrix}, \; P_i^{(0)} = \mathbf{p}_i \begin{pmatrix} \mathbf{e}_{aM_1M_2} \\ \mathbf{0}^T \end{pmatrix}, \; i \geq 2.$$

- Probability $P_i^{(1)}(P_i^{(2)})$ that i customers stay in the system and server 1 (server 2) is under repair

$$P_0^{(1)} = \mathbf{p}_0 \begin{pmatrix} \mathbf{0}_a^T \\ \mathbf{e}_{aR_1} \\ \mathbf{0}^T \end{pmatrix}, \quad P_0^{(2)} = \mathbf{p}_0 \begin{pmatrix} \mathbf{0}_{a(1+R_1)}^T \\ \mathbf{e}_{aR_2} \\ \mathbf{0}^T \end{pmatrix}$$

$$P_1^{(1)} = \mathbf{p}_1 \begin{pmatrix} \mathbf{0}^T_{a(M_1+M_2)} \\ \mathbf{e}_{aM_2R_1} \\ \mathbf{0}^T \end{pmatrix}, \quad P_0^{(2)} = \mathbf{p}_1 \begin{pmatrix} \mathbf{0}^T_{a(M_1+M_1+M_2R_1)} \\ \mathbf{e}_{aM_1R_2} \\ \mathbf{0}^T \end{pmatrix}.$$

$$P_i^{(1)} = \mathbf{p}_i \begin{pmatrix} \mathbf{0}^T_{aM_1M_2} \\ \mathbf{e}_{aM_2R_1} \\ \mathbf{0}^T \end{pmatrix}, \quad P_0^{(2)} = \mathbf{p}_0 \begin{pmatrix} \mathbf{0}^T_{aM_2(M_1+R_1)} \\ \mathbf{e}_{aM_1R_2} \\ \mathbf{0}^T \end{pmatrix}, i \geq 2.$$

- Probability $P_i^{(3)}$ that i customers stay in the system and both main servers are under repair

$$P_0^{(3)} = \mathbf{p}_0 \begin{pmatrix} \mathbf{0}^T \\ \mathbf{e}_{aR_1R_2} \end{pmatrix}, P_1^{(3)} = \mathbf{p}_1 \begin{pmatrix} \mathbf{0}^T \\ \mathbf{e}_{aM_3R_1R_2} \end{pmatrix}, P_i^{(3)} = \mathbf{p}_i \begin{pmatrix} \mathbf{0}^T \\ \mathbf{e}_{aM_3R_1R_2} \end{pmatrix}, i \geq 2.$$

- Probability that at an arbitrary time the servers are in the state n

$$P^{(n)} = \sum_{i=0}^{\infty} P_i^{(n)}, \; n = \overline{0,3}.$$

- Probability $P_{i,k}^{(0)}$ that an arriving batch of size k finds i customers in the system and both servers fault-free

$$P_{0,k}^{(0)} = \lambda^{-1}\mathbf{p}_0 \begin{pmatrix} I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}} \\ O \end{pmatrix} D_k, \quad P_{1,k}^{(0)} = \lambda^{-1}\mathbf{p}_1 \begin{pmatrix} I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}(M_1+M_2))} \\ O \end{pmatrix} D_k,$$

$$P_{i,k}^{(0)} = \lambda^{-1}\mathbf{p}_i \begin{pmatrix} I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}M_1M_2} \\ O \end{pmatrix} D_k, \; i \geq 2.$$

- Probability $P_{i,k}^{(1)}(P_{i,k}^{(2)})$ that an arriving batch of size k finds i customers in the system and server 1 (server 2) under repair

$$P_{0,k}^{(1)} = \lambda^{-1}\mathbf{p}_0 \begin{pmatrix} O_{a\times\bar{W}} \\ I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}R_1} \\ O \end{pmatrix} D_k, \quad P_{1,k}^{(1)} = \lambda^{-1}\mathbf{p}_1 \begin{pmatrix} O_{a(M_1+M_2)\times\bar{W}} \\ I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}M_2R_1} \\ O \end{pmatrix} D_k,$$

$$P_{i,k}^{(1)} = \lambda^{-1}\mathbf{p}_i \begin{pmatrix} O_{aM_1M_2\times\bar{W}} \\ I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}M_2R_1} \\ O \end{pmatrix} D_k, \; i \geq 2,$$

$$P_{0,k}^{(2)} = \lambda^{-1}\mathbf{p}_0 \begin{pmatrix} O_{a(1+R_1)\times\bar{W}} \\ I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}R_2} \\ O \end{pmatrix} D_k, \quad P_{1,k}^{(2)} = \lambda^{-1}\mathbf{p}_1 \begin{pmatrix} O \\ I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}M_1R_2} \\ O_{aM_3R_1R_2\times\bar{W}} \end{pmatrix} D_k,$$

$$P_{i,k}^{(2)} = \lambda^{-1}\mathbf{p}_0 \begin{pmatrix} O_{aM_2(M_1+R_1)\times\bar{W}} \\ I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}M_1R_2} \\ O \end{pmatrix} D_k, \; i \geq 2.$$

- Probability $P_{i,k}^{(3)}$ that an arriving batch of size k finds i customers in the system and both main server under repair

$$P_{0,k}^{(3)} = \lambda^{-1}\mathbf{p}_0 \begin{pmatrix} O \\ I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}R_1R_2} \end{pmatrix} D_k,\ P_{1,k}^{(3)} = \lambda^{-1}\mathbf{p}_1 \begin{pmatrix} O \\ I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}M_3R_1R_2} \end{pmatrix} D_k,$$

$$P_{i,k}^{(3)} = \lambda^{-1}\mathbf{p}_i \begin{pmatrix} O \\ I_{\bar{W}} \otimes \mathbf{e}_{\bar{V}M_3R_1R_2} \end{pmatrix} D_k,\ i \geq 2.$$

6 Numerical Results

In this section, we present results of four numerical experiments. The purpose of the experiments is to study the behavior of the main performance characteristics of the system as functions of its parameters and illustrate the influence of the correlation in the input flow and variation in the repair process.

Experiment 1. In the experiment, we investigate the dependence of the mean number of customers in the system, L, on the input rate λ for different values of the breakdown rate h.

We suppose that the maximum size of batch in the $BMAP$ is 3. To specify the $BMAP$, we first define the matrices D_0 and D

$$D_0 = \begin{pmatrix} -1.349076 & 10^{-6} \\ 10^{-6} & -0.043891 \end{pmatrix}, \quad D = \begin{pmatrix} 1.340137 & 0.008939 \\ 0.0244854 & 0.0194046 \end{pmatrix}.$$

Now we express the matrices $D_k, k = \overline{1,3}$, in terms of the matrix D using the formula $D_k = Dq^{k-1}(1-q)/(1-q^3), k = \overline{1,3}$, where $q = 0.8$.

This $BMAP$ has the coefficient of correlation $c_{cor} = 0.407152$. The squared coefficient of variation is equal to $c_{var}^2 = 9.621426$.

$MMAP$ of breakdowns is defined by the matrices

$$H_0 = \begin{pmatrix} -8.110725 & 0 \\ 0 & -0.26325 \end{pmatrix},$$

$$H_1 = \frac{1}{3}\begin{pmatrix} 8.0568 & 0.053925 \\ 0.146625 & 0.116625 \end{pmatrix}, H_2 = \frac{2}{3}\begin{pmatrix} 8.0568 & 0.053925 \\ 0.146625 & 0.116625 \end{pmatrix}.$$

For this $MMAP$ $c_{cor} = 0.200504557$, $c_{var}^2 = 12.34004211$. We denote by h the total breakdown rate defined as $h = h_1 + h_2$ where, as defined above, h_1 and h_2 are the rates of breakdowns arriving at server 1 and server 2, respectively. It is evident from the form of the matrices H_1 and H_2 that $h_1 = \frac{1}{3}h$ and $h_2 = \frac{2}{3}h$.

PH service time distributions at server 1, server 2 and server 3 will be denoted as $PH_1^{(serv)}, PH_2^{(serv)}$, $PH_3^{(serv)}$, respectively. They are assumed to be Erlangian of order 2 with parameter $20, 15$ and 4. These distributions are defined by the vectors $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)} = \boldsymbol{\beta}^{(3)} = (1, 0)$ and the matrices

$$S^{(1)} = \begin{pmatrix} -20 & 20 \\ 0 & -20 \end{pmatrix}, \quad S^{(2)} = \begin{pmatrix} -15 & 15 \\ 0 & -15 \end{pmatrix}, \quad S^{(3)} = \begin{pmatrix} -4 & 4 \\ 0 & -4 \end{pmatrix}.$$

PH repair time distributions at server 1 and server 2 coincide. They are assumed to be hyper-exponential of order 2 with the squared variation coefficient $c_{var}^2 = 25.07248$ and are defined by the following vector and matrix:

$$\boldsymbol{\tau}^{(1)} = \boldsymbol{\tau}^{(2)} = (0.05,\, 0.95),\;\; T^{(1)} = T^{(2)} = \begin{pmatrix} -0.003 & 0 \\ 0 & -0.245 \end{pmatrix}.$$

Figure 1 depicts the mean number L of customers in the system as a function of λ for different values of breakdown rate, $h = 0.0001, h = 0.001, h = 0.001$.

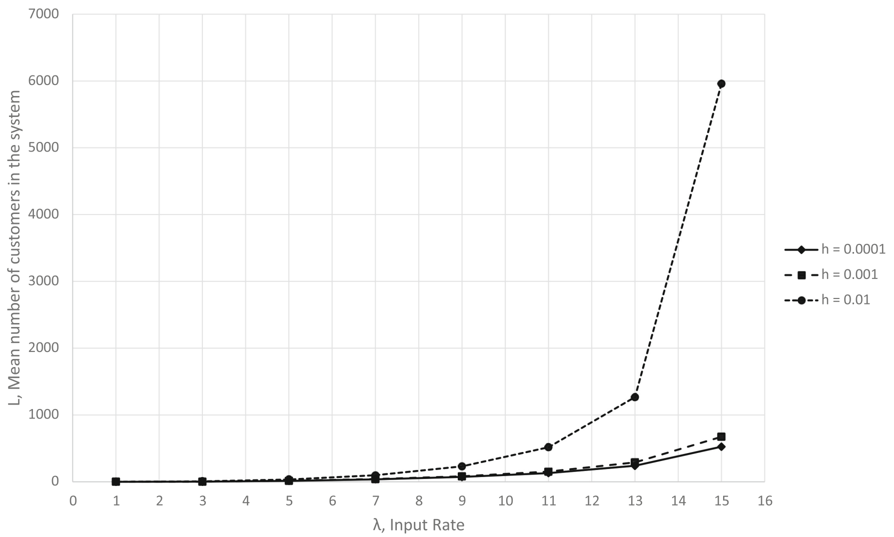

Fig. 1. Mean number L of customers in the system as a function of input rate λ for different values of breakdown rate: $h = 0.0001; h = 0.001; h = 0.001$.

It can be seen from Fig. 1 that the mean number of customers in the system, as expected, increases with increasing input rate λ and the rate of increase grows with increasing the breakdown rate h. You can also see that the mean queue length is rapidly increasing with the growth of load factor ρ.

Experiment 2. In the experiment, we investigate the dependence of the mean value L of the number of customers in the system on the input rate λ for different values of the coefficient of correlation in the $BMAP$. We consider three $BMAP$s having the same mean arrival rate but different coefficients of correlation. These $BMAP$s will be denoted as $BMAP_1, BMAP_2, BMAP_3$.

$BMAP_1$ is a stationary Poisson process with $D_0 = -\lambda$, $D_1 = \lambda$. For this process $c_{cor} = 0$, $c_{var} = 1$.

$BMAP_2$ is defined by the matrices D_0 and $D_k = Dq^{k-1}(1-q)/(1-q^3), k = \overline{1,3}$, where $q = 0.8$,

$$D_0 = \begin{pmatrix} -6.34080 & 10^{-6} \\ 10^{-6} & -0.13888 \end{pmatrix}, \quad D = \begin{pmatrix} 6.32140 \; 0.01939 \\ 0.10822 \; 0.03066 \end{pmatrix}$$

For this $BMAP$ $c_{var} = 3.5$, $c_{cor} = 0.1$.

$BMAP_3$ is the $BMAP$ defined in the Experiment 1. For this $BMAP$ $c_{var}^2 = 9.621425623$, $c_{cor} = 0.407152089$. We also fix the $MMAP$ of breakdowns and PH distributions of service and repair times the same as in Experiment 1.

Figures 2 depicts the mean number of customers in the system as a functions of λ for different $BMAP$s. We see that values of L depend, in some cases strongly, on the coefficient of correlation in the $BMAP$. Under the same value of input rate, λ, the mean number of customers in the system increases when the correlation increases. In addition, the difference in the values of L for different coefficients of correlation increases with increasing λ. In particular, we see that approximation of the $BMAP$ with coefficient correlation $c_{cor} = 0.4$ by the stationary Poisson process leads to a huge error in the calculation of L. Thus, under such an approximation, the mean estimates are too optimistic.

Experiment 3. In the experiment, we investigate the dependence of the number of customers in the system, L, on the breakdowns rate h for PH distributions of repair time with different coefficients of variation.

We take $BMAP$, $MMAP$ of breakdowns and PH distributions of service times from Experiment 1. We assume that PH distributions of repair times of server 1 and server 2 coincide and consider three different PH distributions (PH_1, PH_2, PH_3). The average repair time is equal to 20, however, the coefficients of variation of repair time are different.

PH_1 is the exponential distribution with the parameter 0.05 and coefficient of variation $c_{var} = 1$.

PH_2 is the hyper-exponential distribution of order 2. This distribution is defined by the following vector and matrix:

$$\boldsymbol{\tau} = (0.05,\, 0.95),\ T = \begin{pmatrix} -0.003 & 0 \\ 0 & -0.245 \end{pmatrix}.$$

In this case the repair time has coefficient of variation $c_{var} = 5$.

PH_3 is the hyper-exponential distribution of order 2. This distribution is defined by the following vector and matrix:

$$\boldsymbol{\tau} = (0.05,\, 0.95),\ T = \begin{pmatrix} -250000 & 0 \\ 0 & -0.05 \end{pmatrix}.$$

In this case the repair time has coefficient of variation $c_{var} = 9.9$.

Since in this experiment we choose the breakdown rates such that the following value differs from the previous one by an order of magnitude, it is reasonable to use a logarithmic scale (with a base of ten) on the X-axis. The resulting graph is shown in Fig. 3. It is seen from Fig. 3 that, under the same value of repair rate, h, the mean number of customers in the system L essentially varies for repair times with different coefficient of variation. In this example, the value of L increases when the variation increases and the difference in the value of L increases with increasing of h.

Experiment 4. In the experiment, we investigate the dependence of the throughput ϱ on the repair rate τ.

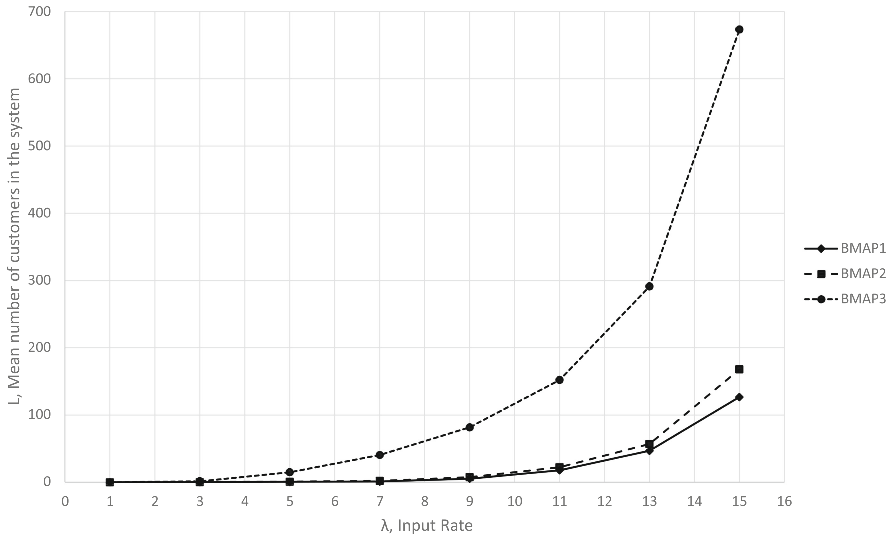

Fig. 2. Mean number of customers in the system as a function of input rate λ for $BMAP$s with different coefficients of correlation ($c_{cor} = 0; 0.1; 0.4$)

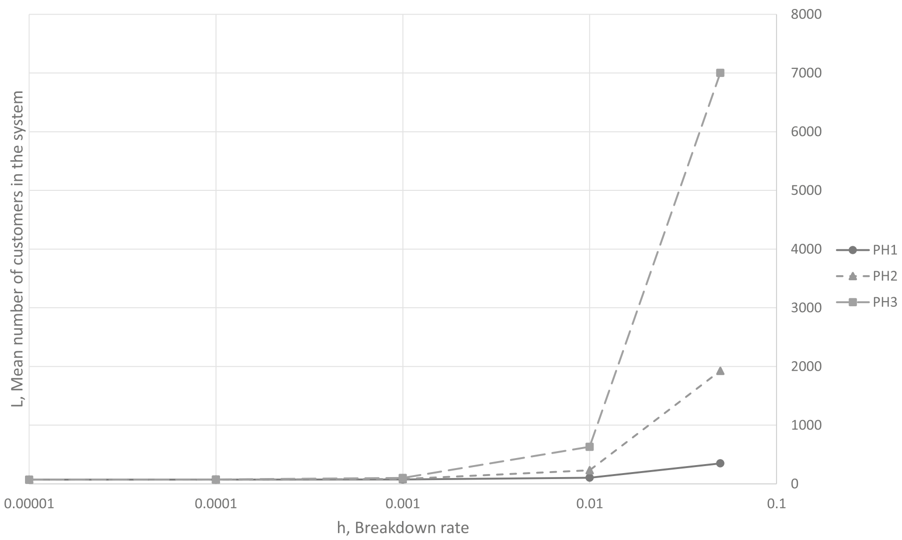

Fig. 3. Mean number of customers in the system as a function the breakdowns rate for repair times with different coefficients of variation

Here we use the same input data as in Experiment 1. Denote the identical repair rates of server 1 and server 2 as τ.

By analogy with the previous experiment, in this experiment we choose the breakdown rates such that the following value differs from the previous one by

Table 1. Values of throughput ϱ obtained in Experiment 4

$\tau \setminus h$	0.0001	0.0005	0.001	0.005	0.01	0.05	0.1	0.5
0.000001	2.627	2.177	2.094	2.020	2.010	2.002	2.001	2.000
0.00001	4.865	2.995	2.627	2.177	2.094	2.020	2.010	2.002
0.0001	11.95	6.49	4.865	2.994	2.627	2.176	2.094	2.020
0.001	16.68	14.06	11.95	6.494	4.864	2.994	2.626	2.176
0.01	17.42	17.08	16.68	14.06	11.94	6.489	4.858	2.986
0.1	17.49	17.46	17.42	17.08	16.67	14.05	11.92	6.439

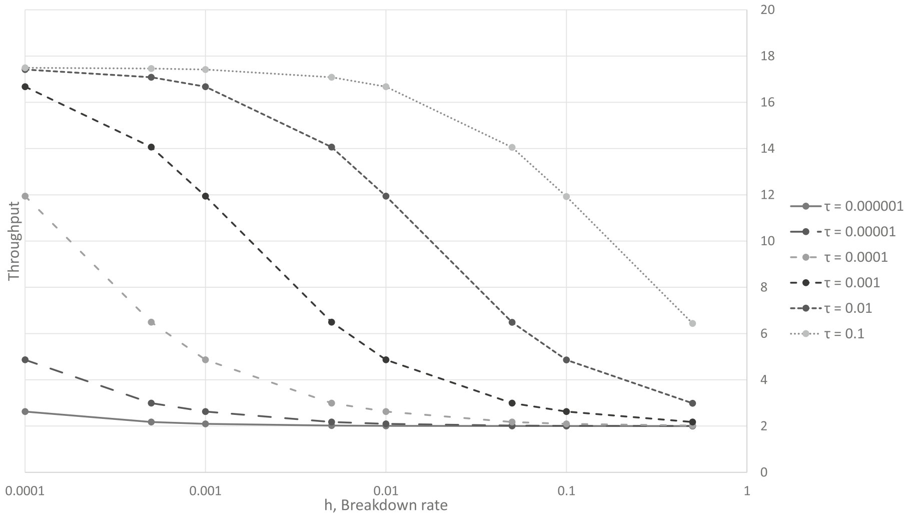

Fig. 4. Throughput of the system as a function of the breakdown rate h for different repair rates τ.

an order of magnitude. Thus, it is reasonable to use a logarithmic scale (with a base of ten) on the X-axis. The resulting graph is shown in Fig. 4. The data for the graph are also given in Table 1. As expectable, under the same value of repair rate τ, the throughput decreases when the breakdown rate increases. The behavior of the curves is more interesting. Let us consider the case when breakdowns rarely occur and repairs are fast. In this case we can assume that $\varrho \approx \mu_1 + \mu_2$, where μ_n is the service rate on server $n, n = 1, 2$. As it is seen from Table 1 this intuitive assumption is confirmed numerically: for $\tau = 0.1$ and $h = 0.0001$ $\varrho = 17.49$ while $\mu_1 + \mu_2 = 17.5$ in this experiment. It can also be seen from the Figure and the Table that, as expected, the throughput cannot exceed the horizontal asymptote $\varrho = \mu_1 + \mu_2$. Further, it can be noted that as repair rate is low and the breakdown rate increases, the curves tend to the horizontal asymptote $\varrho = \mu_3 = 2$, where μ_3 is the service rate on the backup server. Such a behavior coincides with the expected behavior of the system since, under the high breakdown rate, both main servers are almost always broken and only the backup reliable server serves customers.

7 Conclusion

In this paper, we have investigated the unreliable queueing system with backup server that can be used for modelling the hybrid communication system consisting of two main unreliable but high-speed channels (FSO channel and a radio channel of millimeter-wave) which are reserved by reliable but low-speed radio channel. We make quite general assumptions about the process of arrival of customers and breakdowns as well as about service and repair processes. We investigate the system behavior in steady-state. To this end, we derive nontrivial condition for existence of the stationary regime, calculate the steady-state probabilities of the system states and derive formulas for a number of importance performance measures of the system. We give the numerical examples that illustrate the computational tractability of the presented results and investigate the main performance measures of the system as functions of input and breakdown rates, correlation in the $BMAP$ and variation in the repair process.

References

1. Douik, A., Dahrouj, H., Al-Naffouri, T.Y., Alouini, M.S.: Hybrid radio/free-space optical design for next generation backhaul systems. IEEE Trans. Commun. **64**, 2563–2577 (2016)
2. Esmail, M.A., Fathallah, H., Alouini, M.S.: Outdoor FSO communications under fog: attenuation modeling and performance evaluation. IEEE Photonics J. **8**, 1–22 (2016)
3. Makki, B., Svensson, T., Eriksson, T., Alouini, M.S.: On the performance of RF-FSO links with and without hybrid ARQ. IEEE Trans. Wirel. Commun. **15**, 4928–4943 (2016)
4. Wu, Y., Yang, Q., Park, D., Kwak, K.S.: Dynamic link selection and power allocation with reliability guarantees for hybrid FSO/RF systems. IEEE Access **5**, 13654–13664 (2017)
5. Zhang, V, Chu, Y.J., Nguyen, T.: Coverage algorithms for WiFO: a hybrid FSO-WiFi femtocell communication system. In: Proceedings of the 26th International Conference on Computer Communication and Networks (ICCCN), pp. 1–6. IEEE (2017)
6. Zhou, K., Gong, C., Wu, N., Xu, Z.: Distributed channel allocation and rate control for hybrid FSO/RF vehicular ad hoc networks. J. Opt. Commun. Netw. **9**, 669–681 (2017)
7. Shehaj, M., Nace, D., Kalesnikau, I., Pioro, M.: Link dimensioning of hybrid FSO/fiber networks resilient to adverse weather conditions. Comput. Netw. **161**, 1–13 (2019)
8. Advanced Optical Wireless Communication Systems. Cambridge University Press (2012)

9. Sharov, S.Yu., Semenova, O.V.: Simulation model of wireless channel based on FSO and RF technologies. In: Proceedings of the 10th Distributed Computer and Communication Networks. Theory and Applications (DCCN-2010), pp. 368–374 (2010)
10. Vishnevsky, V.M., Semenova, O.V., Sharov, S.Y.: Modeling and analysis of a hybrid communication channel based on free-space optical and radio-frequency technologies. Autom. Remote Control **72**, 345–352 (2013)
11. Vishnevsky, V.M., Kozyrev, D.V., Semenova, O.V.: Redundant queuing system with unreliable servers. In: Proceedings of the 6th International Congress on Ultra Modern Telecommunications and Control Systems, pp. 383–386. IEEE Xplore Digital Library (2014)
12. Dudin, A., Klimenok, V., Vishnevsky, V.: Analysis of unreliable single server queuing system with hot back-up server. Commun. Comput. Inf. Sci. **499**, 149–161 (2015)
13. Klimenok, V.I.: Two-server queueing system with unreliable servers and Markovian arrival process. Commun. Comput. Inf. Sci. **800**, 42–55 (2017)
14. Vishnevsky, V.M., Klimenok, V.I.: Unreliable queueing system with cold redundancy. Commun. Comput. Inf. Sci. **522**, 336–347 (2015)
15. Lucantoni, D.M.: New results on the single server queue with a batch Markovian arrival process. Commun. Stat. Stoch. Model. **7**, 1–46 (1991)
16. Neuts, M.: Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach. Johns Hopkins University Press, Baltimore (1981)
17. He, Q.M.: Queues with marked customers. Adv. Appl. Probab. **28**, 567–587 (1996)
18. Klimenok, V.I., Dudin, A.N.: Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory. Queueing Syst. **54**(245–259), 42–55 (2006)
19. Dudin, A., Klimenok, V., Lee, M.H.: Recursive formulas for the moments of queue length in the $BMAP/G/1$ queue. IEEE Commun. Lett. **13**, 351–353 (2009)