

# Optimal Control of $M(t)/M/K$ Queues with Homogeneous and Heterogeneous Servers

Dmitry Efrosinin<sup>1,2(✉)</sup> and Michael Feichtenschlager<sup>3</sup>

<sup>1</sup> Institute of Control Sciences, Profsoyuznaya str., 65, 117997 Moscow, Russia  
dmitry.efrosinin@mail.ru

<http://www.ipu.ru>

<sup>2</sup> RUDN University, Miklukho-Maklaya str., 6, 117198 Moscow, Russia  
<http://www.rudn.ru>

<sup>3</sup> Johannes Kepler University Linz, Altenbergerstrasse 69, 4040 Linz, Austria  
michael-feichtenschlager@gmx.at  
<http://www.jku.at>

**Abstract.** The paper deals with a multi-server controllable queueing system  $M(t)/M/K$  with time-dependent and, in particular, with periodic arrival rates. The models with homogeneous and heterogeneous servers are of interest. In latter case the fastest free server allocation mechanism is assumed and the preemption is allowed. The control problem consists in evaluation of the optimal number of servers during some specified stages and is solved by finite horizon dynamic programming approach. To calculate the transient solutions we use a forth-order Runge-Kutta method for the system with a truncated queue length. The results are compared with corresponding queues operating in a stationary regime. It is shown that the optimal control policies are also time dependent and periodic as arrival rates and heterogeneous systems are superior in performance comparing to the homogeneous ones.

**Keywords:** Time-dependent arrival rate · Controllable queueing system · Dynamic programming approach · Forth-order Runge-Kutta method

## 1 Introduction

Many queueing systems are subject to time-dependent changes in system parameters. This feature is very important to cover the problems with seasonality and periodicity of stochastic processes. Particularly it happens with an arrival rate which is used for modelling of arrivals of calls and inquires at call centres, arrival of packets at routers of the telecommunication systems, of time changing air traffic at airports, different arrival rates of trucks to the warehouses, goods

---

D. Efrosinin—This work was funded by the Russian Foundation for Basic Research, Project No. 16-37-60072 mol\_a.dk. The reported study was funded within the Agreement No. 02.a03.21.0008 dated 24.11.2016 between the Ministry of Education and Science of the Russian Federation and RUDN University.

depot or seaports and so on. A very good literature overview on this subject can be found in [4]. This paper surveys and classifies the results on performance evaluation approaches for time-dependent queueing systems and their applications and identifies the links between different approaches. The performance analysis of multi-server queueing system subject to breakdowns was studied in [1]. There are several approaches to analyse such systems. The dynamic behaviour of Markovian queueing systems is described by a system of Kolmogorov differential equations (KDEs). Analytical solution of such systems exists only for special cases. Numerical approaches are based on a Runge-Kutta method. The systems with an infinite buffer are normally approximated by using a finite buffer system. The numerical solution of KDEs is used for example for the performance evaluation of a  $M(t)/M/1/N$  system in [3].

In many cases the time-dependent system parameters must be combined with some controllable problems. The paper [5] deals with optimal allocation of such resources as beds in emergency departments of a hospital. For the mathematical modelling the queueing system with losses of the type is used. It was shown that the periodic variation of arrival rates makes a hysteretic policy time-dependent and periodic with the same period. To find the optimal decisions dynamic programming is used. The same approach was used in [2] for the multi-server queueing system with a controllable number of homogeneous servers.

The contribution of the present work is an evaluation of the optimal number of servers in multi-server queueing system with homogeneous and heterogeneous servers and time-dependent arrival rate. The discretization of a continuous-time Markov process is performed to apply the Runge-Kutta method and the iterative dynamic programming algorithm over a finite horizon. The decisions are chosen at specified moments of time which divide the observation time interval into so called stages. The number of available servers is assumed to be a constant within each stage. The paper provides comparison analysis of stationary and transient solutions as well as homogeneous and heterogeneous systems.

The rest of the paper is organized as follows. In Sect. 2 we describe a mathematical model and formulate a optimization problem for transient and stationary case. Section 3 deals with a description of a time-dependent arrival rate. In Sect. 4 a forth-order Runge-Kutta method is adopted for the model under study. The recursive dynamic programming algorithm is shown in Sect. 5. Some illustrative numerical examples are discussed in Sect. 6. Conclusions are given in Sect. 7.

## 2 Mathematical Model

Consider the controllable multi-server queueing system  $M(t)/M/K$  with  $K$  servers. This system features Poisson arrival stream with time dependent arrival rate  $\lambda(t)$ . The servers are assumed to be heterogeneous with servers intensities  $\mu_j, j = 1, 2, \dots, K$ . In special case when all intensities are equal,  $\mu_j = \mu, j = 1, 2, \dots, K$ , we get the homogeneous system. The control consists in specification of the number of servers  $K(t)$  at any decision epoch which will be specified later.

Denote by  $N(t)$  the number of customers in the system at time  $t$ . The dynamics of the system is described by means of the controllable continuous-time inhomogeneous Markov chain  $\{N(t)\}_{t \geq 0}$  with a set of states  $E = \{n; n \in \mathbb{N}_0\}$  and set

of control actions  $A = \{1, 2, \dots, K\}$ . Define additional cost structure with the following components:  $c_1$  – the waiting cost per unit of time for each customer in the queue,  $c_{2,j}$  – the idle state cost per unit of time when the server  $j$  is idle. In homogeneous case it is assumed that  $c_{2,j} = c_2$ ,  $j = 1, 2, \dots, K$ . The servers are enumerated in such a way that

$$\mu_1 \geq \mu_2 \geq \dots \geq \mu_K, \quad c_{2,1} \geq c_{2,2} \geq \dots \geq c_{2,K}. \tag{1}$$

In accordance with the given cost structure, the mean total cost criterion is denoted by

$$J^f(n) = \mathbb{E}^f \left[ \int_0^T c(N(t), K(t), \lambda(t), t) dt \mid N(0) = n \right]. \tag{2}$$

Here  $f$  is a Markov control policy which depends on the current state and time only, i.e.  $K(t) = f(t, n(t))$ , the expectation  $\mathbb{E}^f$  is taken with respect to the probability distribution  $\mathbb{P}^f$  over the state-action sequence under control policy  $f$ . The immediate cost function  $c(N(t), K(t), \lambda(t), t)$  is defined as

$$\begin{aligned} c(N(t), K(t), \lambda(t), t) = c_1 \sum_{k=K(t)+1}^{\infty} (k - K(t)) \mathbb{1}_{\{N(t)=k\}} \\ + \sum_{k=0}^{K(t)} \sum_{j=k+1}^{K(t)} c_{2,j} \mathbb{1}_{\{N(t)=k\}}. \end{aligned} \tag{3}$$

The substitution of (3) into (2) yields the relation for the mean total cost in form

$$\begin{aligned} J^f(n) &= \int_0^T \eta(n, K(t), \lambda(t), t) dt \\ &= \int_0^T \left[ c_1 \bar{Q}(n, K(t), \lambda(t), t) + \sum_{k=0}^{K(t)} \sum_{j=k+1}^{K(t)} c_{2,j} \pi_k(n, K(t), \lambda(t), t) \right] dt. \end{aligned} \tag{4}$$

The first term by  $c_1$  at the right hand side of (4) stands for the mean number of customers in the queue at time  $t$  with  $K(t)$  servers and initial state  $N(0) = n$ , the second term stands for the mean idle state costs. We wish to minimize the functional  $J^f(n)$  over all control policies and find optimal policy  $f^*$  that achieves the minimal cost  $J^*(n)$ , i.e.

$$J^*(n) := J^{f^*}(n) = \min_f J^f(n). \tag{5}$$

The solution of proposed optimization problem can be performed numerically. To realize some iterative algorithm the continuous time model must be converted to a discrete one. We divide a time interval  $[0, T]$  into  $I$  equally spaced periods. The mean total cost functional in this case can be rewritten as follows,

$$\begin{aligned}
 J^f(n) &= \sum_{i=1}^I \eta(n, K(i), \lambda(i), i) \\
 &= \sum_{i=1}^I \left[ c_1 \bar{Q}(n, K(i), \lambda(i), i) + \sum_{k=0}^{K(i)} \sum_{j=k+1}^{K(i)} c_{2,j} \pi_k(n, K(i), \lambda(i), i) \right],
 \end{aligned} \tag{6}$$

where  $K(i)$  is a number of servers at period  $i$ ,  $\bar{Q}(n, K(i), \lambda(i), i)$  is a mean number of customers in the queue at period  $i$ ,  $\pi_k(n, K(i), i)$  – probability of  $k$  customers in the system with  $K(i)$  servers at period  $i$  with initial state  $n$ .

The transient solution of the problem will be compared with a stationary one. In this case the long-run average cost per unit of time

$$\eta(K, \lambda(t)) = c_1 \bar{Q}(K, \lambda(t)) + \sum_{k=0}^K \sum_{j=k+1}^K c_{2,j} \pi_k(K, \lambda(t)) \tag{7}$$

must be minimized over  $K(t)$  for any fixed value  $\lambda(t)$ . The substitution of the stationary state probabilities of the infinite buffer system into (7) yields the relation

$$\begin{aligned}
 \eta(K(t), \lambda(t)) &= \left[ c_1 \prod_{j=1}^{K(t)} \frac{\lambda(t)}{\sum_{k=1}^j \mu_k} \frac{\lambda(t) \sum_{k=1}^{K(t)} \mu_k}{(\lambda(t) - \sum_{k=1}^{K(t)} \mu_k)^2} \right. \\
 &\quad \left. + \sum_{k=0}^{K(t)} \sum_{j=k+1}^{K(t)} c_{2,j} \frac{\lambda(t)^k}{\prod_{l=1}^k \sum_{j=1}^l \mu_j} \right] \pi_0(K(t), \lambda(t)), \\
 \pi_0(K(t), \lambda(t)) &=
 \end{aligned} \tag{8}$$

$$\left[ \sum_{k=0}^{K(t)-1} \frac{\lambda(t)^k}{\prod_{l=1}^k \sum_{j=1}^l \mu_j} + \prod_{j=1}^{K(t)} \frac{\lambda(t)}{\sum_{k=1}^j \mu_k} \frac{\sum_{k=1}^{K(t)} \mu_k}{\sum_{k=1}^{K(t)} \mu_k - \lambda(t)} \right]^{-1}. \tag{9}$$

### 3 Arrival Rate

We have chosen a similar arrival rate  $\lambda(t)$  as in the paper from [2]. The authors have studied there incoming and service of airplanes of an airport modelled via a  $M(t)/M/K/N$  queuing system. In the queuing system under study the condition

$$\lambda(t) < \sum_{j=1}^K \mu_j \tag{10}$$

is a necessary one, since there is no cost relation for customers who get rejected. That means that the maximum number of server  $K$  can handle the average arrival rate of users. The data for  $\lambda(t)$  is given in Table 1.

The arrival rate will be divided into three equidistant stages. It means that each stage lasts eight hours, which is a normal working shift cycle. At the beginning of a stage, the number of customers  $n$  in the system is known. This value

**Table 1.** Values for the arrival rate  $\lambda(t)$

Time in hour	Input intensity
1–5	$8.75 + 4.25 \cos(t/1.6)$
6	4.5
7	5.0
8	5.5
9	6.5
10–17	7.0
18–21	10.0
22–24	13.0

will be called initial state of the current stage. A natural question, one can ask is, how many servers (in this context workers) should be hired at current and following stage(s) so that the expected costs are minimized. Obviously the number of necessary operators are depending on the initial state  $n$ .

### 4 Fourth-Order Explicit Runge-Kutta Method

Since there is no way to solve the system of Kolmogorov forward equations

$$\pi'(t) = \pi(t)A(t) \tag{11}$$

analytically a numerical algorithm is needed to get an approximate solution. For this task we have used the standard fourth order explicit Runge-Kutta procedure which is a widely used one-step method. It considers differential equations of the form

$$y'(t) = f(t, y(t)) \quad \forall t \in (0, T) \tag{12}$$

with given initial condition

$$y(t_0) = y_0. \tag{13}$$

**Algorithm 1.** *The explicit fourth-order Runge-Kutta method.*

*Step 1. Computation of five parameters  $\kappa_1, \kappa_2, \kappa_3, \kappa_4$  and  $\kappa$ :*

$$\begin{aligned} \kappa_1 &= f(t_n, y_n)\Delta t \\ \kappa_2 &= f\left(t_n + \frac{\Delta t}{2}, y_n + \frac{\kappa_1}{2}\right)\Delta t \\ \kappa_3 &= f\left(t_n + \frac{\Delta t}{2}, y_n + \frac{\kappa_2}{2}\right)\Delta t \\ \kappa_4 &= f(t_n + \Delta t, y_n + \kappa_3)\Delta t \\ \kappa &= \frac{1}{6}(\kappa_1 + 2\kappa_2 + 2\kappa_3 + \kappa_4) \end{aligned}$$

Step 2. Evaluation of  $y_{n+1}$  by the recursive relation,

$$y_{n+1} = y_n + \kappa.$$

For solving the Kolmogorov forward equations (11) interpret  $f(t)\pi(t)$  as  $f(\pi(t), t)$ , choose a appropriate step size  $\Delta t$  and apply this Runge-Kutta method directly on the function  $f(\pi(t), t)$ . Obviously the error between the calculated and the real solution gets less if one selects a smaller step size. On the other hand a greater step size means less computing time. We have chosen  $\Delta t = 0.005$ . This value for the step size seems to have a reasonable balance between computing time and computation error.

To solve the system (11) we use a truncation of the buffer capacity by assuming that  $N$  is the maximum allowable number of customers in the system.

## 5 Optimisation Problem

Let  $T = 24$  h be an observation cycle and a finite horizon for the dynamic programming. The decision epochs occur each 8 h, hence we get  $S = 3$  stages with  $\frac{T}{S}$  periods  $i$  within each stage  $s$ . A strategy at a decision epoch  $d = (S-s)I/S+1$  which depends on a current stage  $s$  is denoted by  $f(d, n)$ , where  $n$  stands as before for the initial state. A strategy is equal for any period  $i$  from the interval

$$d \leq i \leq d + \frac{I}{S} - 1.$$

Denote by  $V_n(s)$  the optimal cost function for  $s$  stages left which we refer to as value function:

$$V_n(s) = \min_f \mathbb{E}^f \left[ \sum_{i=1}^I c(N(i), K(i), \lambda(i), i) \mid N(s) = n \right], \quad s = 1, 2, \dots, S, \quad n \in E. \tag{14}$$

The minimum must be taken over tail policies

$$(f(1, K(1)), f(I/S + 1, K(I/S + 1)), \dots, f((S - 1)I/S + 1, K((S - 1)I/S + 1))).$$

Obviously,  $V_n(S) = J^*(n)$ .

**Algorithm 2.** *The finite horizon dynamic programming:*

Step 1. *Backward recursion:*  $V_n(0) = 0, n \in E$ , and

$$V_n(s) = \min_{1 \leq k \leq K} \left\{ r(n, k, s) + \sum_{m=0}^N p_{nm}(k, s) V_m(s - 1) \right\}, \tag{15}$$

where  $r(n, k, s)$  is the total average cost per stage,

$$r(n, k, s) = \sum_{i=d}^{d + \frac{I}{S} - 1} \eta(n, k, \lambda(i), i), \quad s = 1, 2, \dots, S, \tag{16}$$

and the transition probabilities between the stages are defined as

$$p_{nm}(k, s) = \mathbb{P}\left[N(d + I/S) = m \mid N(d) = n, f(d, n) = k\right], \quad (17)$$

$$s = 2, \dots, S. \quad (18)$$

Step 2. Any Markov policy  $f^*$  that satisfies

$$f^*(d, n) = \arg \min_{1 \leq k \leq K} \left\{ r(n, k, s) + \sum_{m=0}^N p_{nm}(k, s) V_m(s-1) \right\} \quad (19)$$

is an optimal control policy.

The values  $p_{nm}(k, s)$  in (17) have to be interpreted in the following way. They stand for the probability to be in state  $m$  at the beginning of the next stage under the condition that the initial state of the previous stage was  $n$ .

**Algorithm 3.** The following basic steps are involved into the computation procedure:

Step 1. Compute the state probabilities  $\pi(n, k, \lambda(i), i)$  for each  $n, k$  and  $i$  via the Runge-Kutta fourth order method.

Step 2. Compute the cost function  $\eta(n, k, \lambda(i), i)$  for each  $n, k$  and  $i$ .

Step 3. Compute  $r(n, k, s)$  for each  $n, k$  and  $s$  by accumulating  $\eta(n, k, \lambda(i), i)$  over all periods  $i$  within the corresponding stage.

Step 4. Compute the transition probabilities  $p_{nm}(k, s)$  for each  $n, m, k$  and  $s$  via Runge-Kutta fourth order method.

Step 5. Evaluation of the optimal strategy for any  $s$  and  $n$  by means of Algorithm 2.

The number of servers  $k^* = f(s, n)$  defined by (19) for which the expression in the right hand side of (15) is minimal is called the best or optimal strategy at stage  $s$  given the initial state is  $n$ .

*Remark 1.* Notice that in this queueing model rejecting of customers is not a valid option. If, for example, the initial state  $n$  at the current stage is the capacity of the buffer plus 2, the best strategy cannot be one server. However if the one choose the waiting room capacity high enough, restrict  $n$  up to this value and condition (10) is clearly fulfilled there will be no dropping of users.

## 6 Numerical Realisation and Results

The main goal in this paper is to compare the operating costs for the  $M(t)/M/K$  queue between homogeneous and heterogeneous servers when optimal policies are used. Further the difference between transient and stationary solution will be contrasted for both philosophies. For this computations the following assumptions are used:

1. The maximum number of servers is  $K = 6$ .
2. The buffer capacity is  $N - K = 10$ .
3. The waiting cost  $c_1 = 10$ .
4. The service rate of the server in homogeneous case is  $\mu = 4$ .
5. The service intensities  $\mu_j, j = 1, 2, \dots, K$ , of heterogeneous servers are listed in Table 2.
6. The idle state cost in homogeneous case is  $c_2 = 2.1$ .
7. The idle state costs  $c_{2,j}, j = 1, 2, \dots, K$ , for heterogeneous servers are listed in the Table 2.

To get comparability with the homogeneous case we have chosen the service rates and operator idle costs for heterogeneous servers so that they satisfy the following conditions,

$$\sum_{j=1}^K \mu_j = K\mu, \quad \sum_{j=1}^K \frac{c_{2,j}}{\mu_j} = \frac{c_2}{\mu} \tag{20}$$

together with the ordering (1). That means that the server 1 is the fastest one but has the highest mean standing costs. Followed by server 2 and so forth. This is a quite reasonable assumption because a faster operator needs more resources. If this server is idle more costs are generated as for a slower one. In heterogeneous case the fastest free server policy is used for the allocation mechanism. If more then one operator is free and a new customer enters the queue the fastest free server will be entrusted with this task. This must be considered in the calculation of  $\eta(n, k, \lambda(i), i)$ . To calculate the best stationary policy the long-run average cost  $\eta(k, \lambda(i))$  must be minimized for  $k, 1 \leq k \leq K$ . For homogeneous servers the service intensities  $\mu_j$  as well as idle state costs  $c_{2,j}$  in homogeneous case must be set to be equal as was discussed before.

**Table 2.** Values of  $\mu_j$  and  $c_{2,j}$  for heterogeneous servers

Server	Idle state cost	Service rate
1	$c_{2,1} = 0.4371$	$\mu_1 = 480/49$
2	$c_{2,2} = 0.9 c_{2,1}$	$\mu_2 = \mu_1/2$
3	$c_{2,3} = 0.85 c_{2,2}$	$\mu_3 = \mu_1/3$
4	$c_{2,4} = 0.8 c_{2,3}$	$\mu_4 = \mu_1/4$
5	$c_{2,5} = 0.75 c_{2,4}$	$\mu_5 = \mu_1/5$
6	$c_{2,6} = 0.7 c_{2,5}$	$\mu_6 = \mu_1/6$

In Figs. 1(a) and (b) we illustrate the mean number of customers in the buffer  $\bar{Q}(n, K(t), \lambda(t), t)$  in homogeneous and heterogeneous cases for different number of servers  $K(t)$ . The queue length for  $K(t) = 4, 5, 6$  servers are not shown in these figures, since the values are very small (especially when heterogeneous servers



are used). In Fig. 1(a) and (b) the initial state  $N(t) = n$  at time  $t = 0$  is set to be zero. That means the queueing system is at start empty. The mean number of waiting customers, which is needed in (6), is calculated via the formula

$$\bar{Q}(n, K(i), \lambda(i), i) = \sum_{k=K(i)+1}^N (k - K(i))\pi_k(n, K(i), \lambda(i), i), \quad (21)$$

where  $K(i)$  is fixed in period  $i$ ,  $k$  is a state at period  $i$  and  $N$  is the maximum number of customers in the system. The state probabilities in this expression are the solution of the system (11) performed by the fourth-order Runge-Kutta method for each given number of servers  $K(i)$  at each period  $i$ .

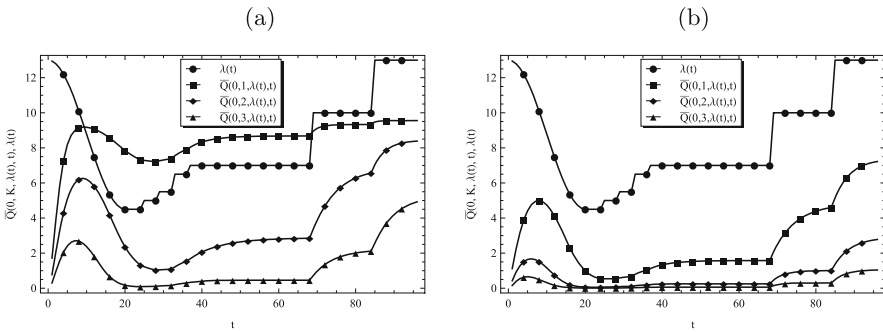


Fig. 1.  $\bar{Q}(n, K, \lambda(t), t)$  for (a) homogeneous and (b) heterogeneous system

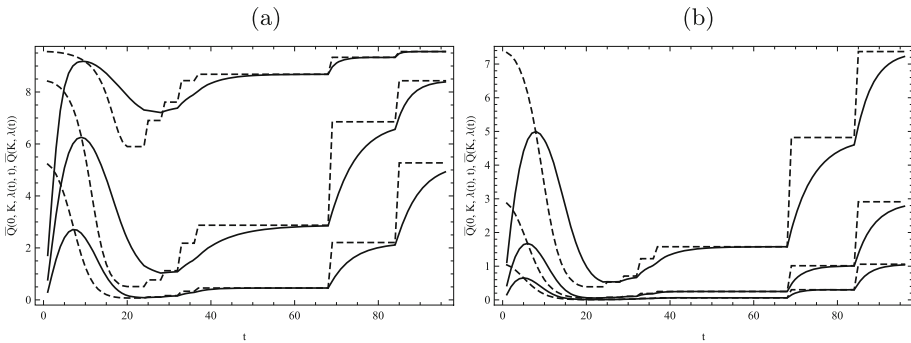


Fig. 2. Transient/stationary  $\bar{Q}(n, K, \lambda(t), t)/\bar{Q}(K, \lambda(t))$  for (a) homogeneous and (b) heterogeneous system

Figures 2(a) and (b) illustrate how the mean queue length (21) differs from the stationary solution in the homogeneous and heterogeneous case which can

be calculated by expression from (8). Again the mean queue length for 4, 5 and 6 servers are not shown because of the small values. The continuously plotted lines belong to the transient and the dashed lines to the stationary solution.

These pictures illustrate clearly the behaviour of the queuing system. When the transient solution is off and  $\lambda(t)$  is a constant value for a certain period it converges to the stationary result as time goes by. This is not astonishing because a stationary queuing system can be interpreted as a long running transient system with a constant arrival rate.

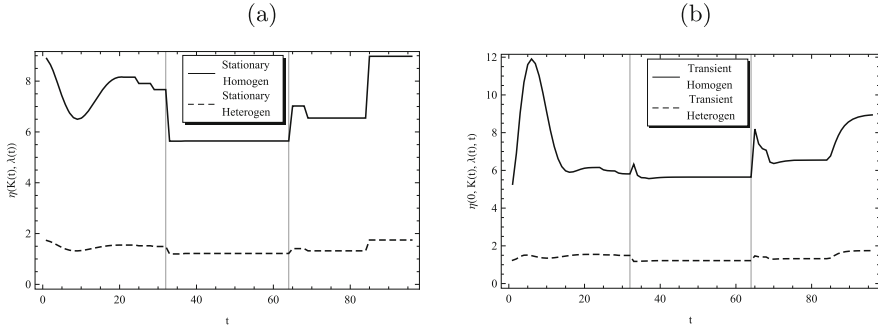
To compare the minimum expected costs between homogeneous/heterogeneous servers in stationary/transient case simply evaluate  $\eta(n, k^*, \lambda(i), i)$  and  $\eta(k^*, \lambda(i))$  for each period  $i$  and corresponding optimal number of servers  $k^*$ . The calculated values of the optimal policy are listed in Table 3. To get

**Table 3.** Optimal policy  $f$

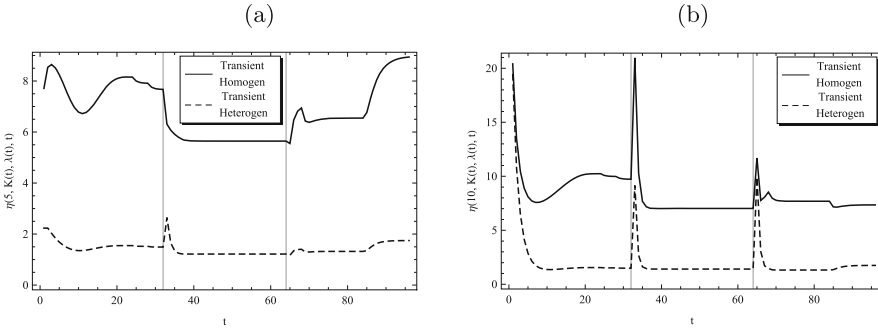
Initial state at stage one	Homogen transient	Homogen stationary	Heterogen transient	Heterogen stationary
0	4	5	6	6
1	4	5	6	6
2	5	5	6	6
3	5	5	6	6
4	5	5	6	6
5	5	5	6	6
6	5	5	6	6
7	5	5	6	6
8	5	5	6	6
9	5	5	6	6
10	6	5	6	6
Stage two				
0	4	4	4	4
1	4	4	4	4
2	4	4	4	4
3	4	4	4	4
4	4	4	4	4
5	4	4	4	4
6	4	4	5	4
7	4	4	5	4
8	4	4	6	4
9	4	4	6	4
10	5	4	6	4
Stage three				
0	5	5	6	6
1	5	5	6	6
2	5	5	6	6
3	5	5	6	6
4	5	5	6	6
5	5	5	6	6
6	5	5	6	6
7	5	5	6	6
8	5	5	6	6
9	6	5	6	6
10	6	5	6	6

more server variety in the transient solutions one can increase the waiting room capacity and adjust  $c_1$  and  $c_2$ .

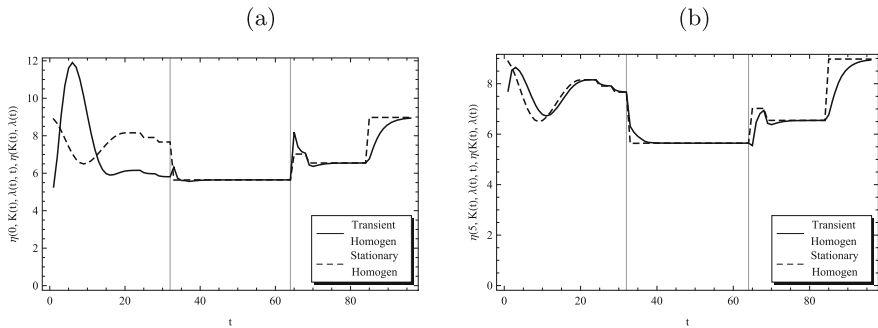
Figures 3(a), (b) and 4(a), (b) show the minimum expected costs  $\eta(n, K(i), \lambda(i), i)$  in homogeneous and heterogeneous case. The first one deals with a



**Fig. 3.** (a)  $\eta(K(t), \lambda(t))$  in stationary case and (b)  $\eta(n, K(t), \lambda(t), t)$  in transient case for  $n = 0$



**Fig. 4.**  $\eta(n, K(t), \lambda(t), t)$  in transient case for (a)  $n = 5$  and (b)  $n = 10$

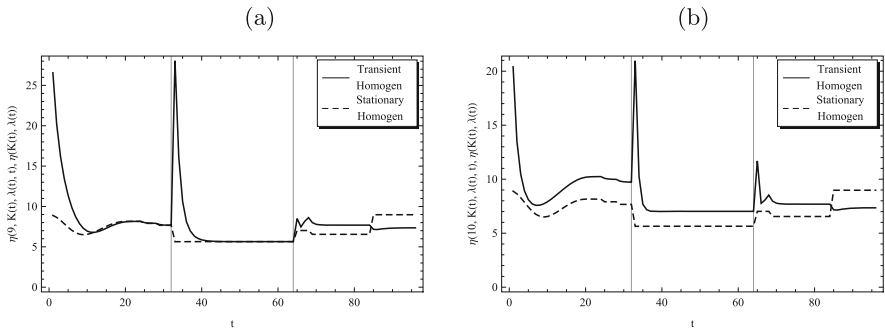


**Fig. 5.**  $\eta(n, K(t), \lambda(t), t)$  with homogeneous servers for (a)  $n = 0$  and (b)  $n = 5$

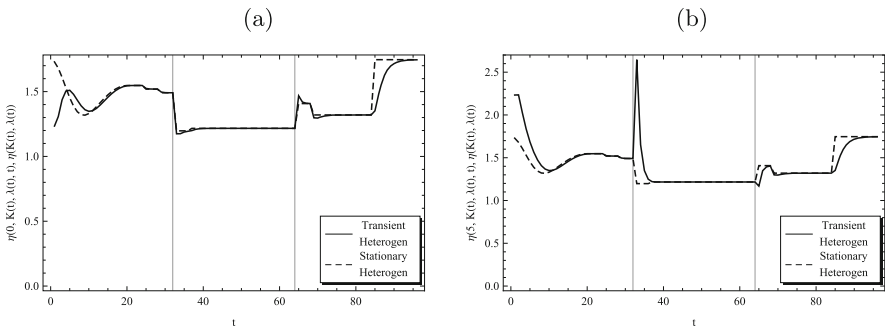
stationary case. Here the initial state  $n$  is without any significance. The second, third, fourth one is dedicated to the transient solution of the optimizing problem with initial state  $n$  at each stage equal to 0, 5 and 10. In the following pictures the stages should be interpreted severally.

As one would expect, the queuing system with heterogeneous servers is superior in terms of running costs. A similar gap to the homogeneous costs as in Figs. 3(b), 4(a) and (b) can be seen for different initial states  $n$  at the end of every stage.

The Figs. 5(a), (b) and 6(a), (b) deal with homogeneous and number Figs. 7(a), (b) and 8(a) and (b) with heterogeneous operators. They picture the minimum expected costs  $\eta(n, K(i), \lambda(i), i)$  and  $\eta(K(i), \lambda(i))$  which were induced in the stationary and transient case respectively for chosen initial states.



**Fig. 6.**  $\eta(n, K(t), \lambda(t), t)$  with homogeneous servers for (a)  $n = 9$  and (b)  $n = 10$



**Fig. 7.**  $\eta(n, K(t), \lambda(t), t)$  with heterogeneous servers for (a)  $n = 0$  and (b)  $n = 5$

At a close look at Figs. 5(a) up to 8(b) one can see that the costs induced by the transient queuing system converges to the expenses of the stationary model if and only if the computed optimal policies for this stage are the same.

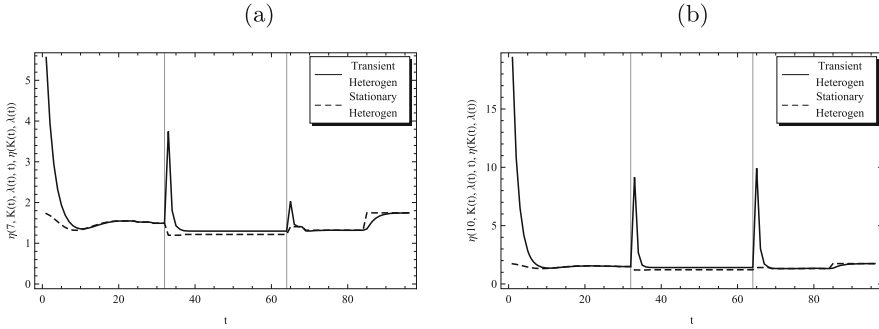


Fig. 8.  $\eta(n, K(t), \lambda(t), t)$  with heterogeneous servers for (a)  $n = 7$  and (b)  $n = 10$

## 7 Conclusions

In this paper we have provided performance analysis of the Markovian controllable queueing system with a time-dependent arrival rate. The control policy prescribes the number of allowable servers and is time-dependent as well. The transient and stationary analysis for homogeneous and heterogeneous systems with preemption is provided. It is shown that the optimal policy differs in transient and stationary case. For the same control policy the corresponding cost functions take very close values. It is confirmed that the heterogeneous queueing systems are superior in performance comparing to the homogeneous case.

## References

1. Ghimire, S., Ghimire, R.P., Thapa, G.B.: Performance evaluation of unreliable  $M(t)/M(t)/n/n$  queueing system. *Br. J. Appl. Sci. Technol.* **7**(4), 412–422 (2015)
2. Jung, M., Lee, E.S.: Numerical optimization of a queueing system by dynamic programming. *J. Math. Anal. Appl.* **141**, 84–93 (1989)
3. Koopman, B.O.: Air-terminal queues under time-dependent conditions. *Oper. Res.* **20**(6), 1089–1114 (1972)
4. Schwarz, J.A., Selinka, G., Stolletz, R.: Performance analysis of time-dependent queueing systems: survey and classification. *Omega* (2016). doi:[10.1016/j.omega.2015.10.013](https://doi.org/10.1016/j.omega.2015.10.013)
5. Tirdad, A., Grassmann, W.K., Tavakoli, J.: Optimal policies of  $M(t)/M/c/c$  queues with two different levels of servers. *Eur. J. Oper. Res.* **249**, 1124–1130 (2016)