Feedback in Infinite-Server Queuing Systems

Svetlana Moiseeva[(✉)] and Lyubov Zadiranova

Tomsk State University, Tomsk, Russia
{smoiseeva,zhidkovala}@mail.ru

Abstract. Queuing systems with infinite servers $M|M|\infty$, $MMPP|M|\infty$, $GI|M|\infty$ with feedback are considered. Asymptotic characteristic functions of the flow of repeating requests in the systems are obtained by using method of asymptotic analysis under a condition of increasing service time. Results of numeric experiments are presented.

Keywords: Queuing system with infinite servers · Asymptotic analysis · Feedback

1 Introduction

Due to development of the telecommunication and information-computing systems, assessment of transmission capacity and of service quality is a very important design stage. Queuing systems (QS) with feedback are adequate mathematic models of many real situations in which part of processed requests needs to return to systems for repeating service.

In queuing theory there are two types of the models with feedback: models with delayed feedback (with an orbit) [1,2]; models with instantaneous feedback [3,4].

Article by D'Avignon G.R., Disney R.L [5] has covered queuing system that has one server, two independent Poisson input flows which give two types of requests, and an instantaneous feedback. Each type of requests has its own distribution for service time. Whether there is feedback (allowing for a repeating service) or there isn't decided depending on request type. This work has covered conditions of existence of stationary mode and conclusion about joint and marginal distributions for the length of a queue.

Articles by I.S. Zaryadov [6,7] have covered the research of single-channel QS with a recurrent input flow of requests, the service time for each call distributed by an exponential law, unlimited capacity storage system and the renewal mechanism with repeating service.Request that is in the server will either leave system with probability of p, or stay with probability of 1-p, while discarding all other requests are rejected.

Real technical systems have limited number of servers, but in the case when probability of request loss could be disregarded such systems could also be approximated by infinite-server QS [8].

QS with infinite number of servers, stationary Poisson arrival process and repeating requests were considered in articles [9].

It's worth noting that using of Poisson arrival process follows a greater inaccuracy when calculating characteristics of service quality [10,11], thus it is necessary to complicate models by using non-Poisson arrival process.

This paper has covered models of queuing with infinite number of servers, Poisson process (M), Markov-modulated Poisson process ($MMPP$) and a renewal process (GI) as arrival process. A flow of requests which are returned to the system for repeating service is an object of study.

2 Problem Statement

Let us consider with the queuing systems with infinite number of servers, Poisson process (M), Markov-modulated Poisson process ($MMPP$) and a renewal process (GI) as models of arrivals.

Service time of a request is a random variable and is distributed by an exponential law with parameter μ. The incoming call occupies any of free server, and after finishing the service leaves the system with probability 1-r, or stays for additional service with it probability r.

The object of the study is a number of requests that have come into reviewed systems for additional service during time interval t (the flow of repeating requests).

3 The Flow of Repeating Requests in $M|M|\infty$ System with Feedback

Let us start with the queuing system with infinite number of servers, and Poisson arrival process with an intensity λ.

Let us denote the following: $i(t)$ is a number of occupied servers at the instant t, $n(t)$ is a number of requests that have come into system for an additional service during time interval t. Two-dimensional flow $\{i(t), n(t)\}$ is Markovian.

For its probability distribution $P(i, n, t) = P\{i(t) = i, n(t) = n\}$ we write down Kolmogorov differential equation system in the form

$$\frac{\partial P(i, n, t)}{\partial t} = -\lambda P(i, n, t) - i\mu P(i, n, t) + \lambda P(i - 1, n, t)$$
$$+ i\mu r P(i, n - 1, t) + \mu(1 - r)(1 + i)P(i + 1, n, t), \qquad (1)$$

where i=0,1,..., n=0,1,....

Let us introduce characteristic functions in the form of

$$H(u, w, t) = \sum_{i=0}^{\infty} \sum_{n=0}^{\infty} e^{jui} e^{jwn} P(i, n, t),$$

then we get the following differential equation from (1)

$$\frac{\partial H(u,w,t)}{\partial t} = \lambda(e^{ju}-1)H(u,w,t)+j\mu(1-re^{jw}-(1-r)e^{-ju})\frac{\partial H(u,w,t)}{\partial u}. \quad (2)$$

The resulting equation lets us determine main statistically distributed characteristics of the considered system, as well as for the flow of repeating requests into the system. Let us research the flow of repeating requests into the system during time interval t, by using method of asymptotic analysis under condition of increasing service time [12]. To do this, let us denote the following

$$\mu = \epsilon, u = \epsilon y, H(u,w,t) = F(y,w,t,\epsilon).$$

We can rewrite (2) in the following form

$$\frac{\partial F(y,w,t,\epsilon)}{\partial t} = j(1-re^{jw}-(1-r)e^{-j\epsilon y})\frac{\partial F(y,w,t,\epsilon)}{\partial y} +\lambda(e^{j\epsilon y}-1)F(y,w,t,\epsilon). \quad (3)$$

Theorem 1. *Limit value, where* $\epsilon \to 0$, *of the function* $F(y,w,t)$ *of solution* $F(y,w,t,\epsilon)$ *of the Eq. (3) has the following form*

$$F(y,w,t) = exp\left\{\frac{\lambda rt(e^{jw}-1)}{1-r}+\frac{j\lambda y}{1-r}\right\}. \quad (4)$$

Proof. After executing the transmission to the limit in Eq. (3) we get the first-order partial differential equation

$$\frac{\partial F(y,w,t,\epsilon)}{\partial t} = jr(1-e^{jw})\frac{\partial F(y,w,t,\epsilon)}{\partial y}. \quad (5)$$

General solution of that equation is

$$F(y,w,t) = \varphi\left(t+\frac{jy}{r(e^{jw}-1)}\right),$$

where $\varphi(y)$ is some function the form of which is determinable by using initial condition.

Let us consider function $F(y,w,t)$ at time zero, obviously, this function doesnt depend on number of repeating requests w, so the initial condition will have the form

$$F(y,w,0) = \Phi(y), \quad (6)$$

where $\Phi(y)$ is an asymptotic approximation of characteristic function of distribution for a number of occupied servers in the system under condition of increasing service time of requests. Its easy to derive that the function has the following form

$$\Phi(y) = exp\left\{\frac{j\lambda y}{1-r}\right\}.$$

Thus, solution of the Eq. (3) has the following form

$$F(y,w,t) = exp\left\{\frac{\lambda rt(e^{jw}-1)}{1-r}+\frac{j\lambda y}{1-r}\right\}.$$

which is the same as the Eq. (4). Theorem is proved.

By assuming the number of occupied devices at (4) as $y = 0$, we have an asymptotic approximation of characteristic function of the number of repeating requests that have come into system during time interval t under a condition of increasing service time

$$h(w, t) = exp \left\{ \frac{\lambda rt(e^{jw} - 1)}{1 - r} \right\}.$$

4   The Flow of Repeating Requests in $MMPP|M|\infty$ System with Feedback

Let us consider the queuing system with infinite number of servers, and Markov-modulated Poisson arrival process ($MMPP$), underlying by Markov chain with a finite number of states $k(t) = 1, 2, \ldots, K$, given matrix of infinitesimal characteristics $\mathbf{Q}$, $i, j = 1, 2, \ldots, K$, and a matrix of conditional densities $\boldsymbol{\Lambda}$. Let us denote $i(t)$ is a number of occupied devices at the instant t, $n(t)$ is a number of repeating requests that have come during the time interval t, $k(t)$ is a state of the Markov chain. Three-dimensional process $\{k(t), i(t), n(t)\}$ is Markovian. For probability distribution $P(k, i, n, t) = P\{k(t) = k, i(t) = i, n(t) = n\}$ we can write down Kolmogorov differential equation system for partial characteristic function in form of differential matrix equation

$$\frac{\partial \mathbf{H}(u, w, t)}{\partial t} + j\mu(re^{jw} - 1 + (1 - r)e^{-ju}) \frac{\partial \mathbf{H}(u, w, t)}{\partial u} = \mathbf{H}(u, w, t)[(e^{ju} - 1)\boldsymbol{\Lambda} + \mathbf{Q}], \quad (7)$$

where
$$\mathbf{H}(u, w, t) = [H(1, u, w, t), H(2, u, w, t), ..., H(K, u, w, t)],$$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{pmatrix}, \mathbf{Q} = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1K} \\ q_{21} & q_{22} & \cdots & q_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ q_{K1} & q_{K2} & \cdots & q_{KK} \end{pmatrix}.$$

Let us find the asymptotic characteristic function of the number of repeating requests in $MMPP|M|\infty$ system during time interval t under condition of increasing time. We will denote

$$\mu = \epsilon, u = \epsilon y, \mathbf{H}(u, w, t) = \mathbf{F}(y, w, t, \epsilon),$$

Let us rewrite the Eq. (7) considering all introduced designations

$$\frac{\partial \mathbf{F}(y, w, t, \epsilon)}{\partial t} + j(re^{jw} - 1 + (1 - r)e^{-jy\epsilon}) \frac{\partial \mathbf{F}(y, w, t, \epsilon)}{\partial y} = \mathbf{F}(y, w, t, \epsilon)[(e^{jy\epsilon} - 1)\boldsymbol{\Lambda} + \mathbf{Q}]. \quad (8)$$

Let us formulate the theorem, proof of which could be carried out the same as the one of the Theorem 1.

Theorem 2. *Sum of the components of limit value where $\epsilon \to 0$ of the vector-function $\boldsymbol{F}(y, w, t)$ of the solution $\boldsymbol{F}(y, w, t, \epsilon)$ of the Eq. (8) has the following form*

$$\boldsymbol{F}(y, w, t)\boldsymbol{E} = exp\left\{\frac{\lambda r t(e^{jw} - 1)}{1 - r} + \frac{j\lambda y}{1 - r}\right\}, \tag{9}$$

where the value of λ is determined by expression

$$\lambda = \boldsymbol{R}\boldsymbol{\Lambda}\boldsymbol{E},$$

vector $\boldsymbol{R}$ is a stationary probability distribution of states of the Markov chain, which is determined by the linear equation system $\begin{cases} \boldsymbol{R}\boldsymbol{Q} = 0, \\ \boldsymbol{R}\boldsymbol{E} = 1. \end{cases}$

By assuming that $y = 0$ at (9), we obtain the asymptotic approximation of characteristic function of the number of repeating requests that have come into the system during time interval t under a condition of increasing service time:

$$h(w, t) = exp\left\{\frac{\lambda r t(e^{jw} - 1)}{1 - r}\right\}.$$

5   The Flow of Repeating Requests in $GI|M|\infty$ System with Feedback

Now, let us start with the queuing system with infinite number of servers and renewal arrival process with the distribution function $A(x)$ for the lengths of intervals between arrivals.

Let us denote the following: $i(t)$ is a number of occupied servers in the system at the moment of time t, $n(t)$ is a number of requests that have come into system for repeating service during time interval t. Because of the resulting random process $\{i(t), n(t)\}$ being non-Markovian, we have to Markovize it by introducing additional variable $z(t)$ which is equal to the length of interval between the instant t and the epoch the next. Then the three-dimensional process $\{z(t), i(t), n(t)\}$ will be a Markov process.

Let us denote the probability distribution for values of the process $P(z, i, n, t) = P\{z(t) < z, i(t) = i, n(t) = n\}$. Then we can write the Kolmogorov differential equation system for partial characteristic function

$$\frac{\partial H(z, u, w, t)}{\partial t} = \frac{\partial H(z, u, w, t)}{\partial z} - j\mu(re^{jw} - 1 + (1 - r)e^{-ju})\frac{\partial H(z, u, w, t)}{\partial u}$$
$$+ (e^{ju}A(z) - 1)\frac{\partial H(0, u, w, t)}{\partial z}. \tag{10}$$

Let us define characteristics of the flow of repeating requests in the system by means of method of asymptotic analysis. We consider the $GI|M|\infty$ system with feedback under a condition of increasing service time. Denote the following:

$$\mu = \epsilon, u = \epsilon y, H(z, u, w, t) = F(z, y, w, t, \epsilon),$$

then we can rewrite (10) considering all introduced designations

$$\frac{\partial F(z, u, w, t, \epsilon)}{\partial t} = \frac{\partial F(z, u, w, t, \epsilon)}{\partial z} - j(re^{jw} - 1 + (1-r)e^{-jy\epsilon})\frac{\partial F(z, u, w, t, \epsilon)}{\partial y}$$

$$+ (e^{jy\epsilon}A(z) - 1)\frac{\partial F(0, u, w, t, \epsilon)}{\partial z}. \tag{11}$$

Let us formulate the theorem, proof of which could be carried out the same as the one of the Theorem 1.

Theorem 3. *The limit value where of the $\epsilon \to 0$, of the function $F(y, w, t)$ of the solution $F(y, w, t, \epsilon)$ of the Eq. (11) has the following form*

$$F(y, w, t) = exp\left\{\frac{\lambda rt(e^{jw} - 1)}{1 - r} + \frac{j\lambda y}{1 - r}\right\}, \tag{12}$$

where

$$\lambda = R'(0),$$

$R(z)$ is a stationary probability distribution of values of the random process $z(t)$.

By assuming that $y = 0$ at (12), we have the asymptotic approximation of the characteristic function of the number of repeating requests that have come into system during time interval t under condition of increasing service time

$$h(w, t) = exp\left\{\frac{\lambda rt(e^{jw} - 1)}{1 - r}\right\}.$$

6 Numerical Results

In order to evaluate the assessment of error of the probability distribution for the reviewed process we have to compare the results of simulation with the results obtained by using method of asymptotic analysis.

The domain of applicability of the results of asymptotic analysis will be defined by means of Kolmogorov distance

$$\Delta = \max_i |\widehat{F(i)} - F(i)|.$$

Where $\widehat{F(i)}$ – empirical probability distribution function obtained by means of imitational modeling, $F(i)$ is a Poisson probability distribution function with the parameter $\lambda rt/(1 - r)$.

Example 1. Consider the system with $MMPP$ arrival process with the following parameters:

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 10 \end{pmatrix}, \mathbf{Q} = \begin{pmatrix} -0.5 & 0.3 & 0.2 \\ 0.2 & -0.6 & 0.4 \\ 0.5 & 0.3 & -0.8 \end{pmatrix}.$$

Let probability of the return of a request back into the system for repeating service is $r = 0.1$. Let us notice that for these parameters the rate of arrivals is $\lambda = 4,73$, and the parameter of Poisson distribution will be equal to $\lambda r/(1-r) = 0.525$.

By increasing the value of an average service time $1/\mu$ we can define when the Kolmogorov distance will become acceptable.

Table 1 has covered values of Kolmogorov distances for the system at various values of a service parameter μ.

Table 1. Error of approximation of the flow of repeating requests into the system by the Poisson process at various values of μ

μ	1	0.5	0.1	0.01
Δ	0.081	0.038	**0.008**	**0.002**

Example 2. Let us consider $GI|M|\infty$ QS with renewal arrival process. Lengths of intervals between the epochs of arrivals have Gamma-distribution with a parameter of form $\alpha = 0,5$ and a parameter of scale $\beta = 2,5$, probability of return of a request into the system $r = 0,1$. Then the rate of the arrival process is $\lambda = 2$, and a parameter of the Poisson distribution is equal to $\lambda r/(1-r) = 0.22$.

Values of Kolmogorov distances for the system at various values of a service parameter μ are given in a Table 2.

Table 2. Error of approximation of the flow of repeating requests into the system by the Poisson flow at various values of μ

μ	1	0.5	0.1	0.01
Δ	0.0540	**0.0252**	**0.0090**	**0.0082**

The analysis demonstrates that probability distribution for a number of requests in the flow of repeating requests in the system with increasing service time is coming close to Poisson distribution. Even at $\mu = 0.5$ the Kolmogorov distances is under 0.03 which is an acceptable error.

The same results were obtained for other values of parameters of arrival process.

7   Conclusion

Thus, this work has covered mathematical models of $M|M|\infty$, $MMPP|M|\infty$ and $GI|M|\infty$ systems with feedback. Using the method of asymptotic analysis under condition of increasing service time, we obtain asymptotic approximation

of characteristic functions of the flow of repeating requests for each system. Its proven that the flow of repeating requests could be approximated by the Poisson process with a parameter $\lambda rt/(1-r)$, where λ is a rate of the arrival process, r is a probability of repeating service. The numerical analysis of the obtained results is presented; the domain of applicability of approximation is defined.

References

1. Pekoz, E.A., Joglekar, N.: Poisson traffic flow in a general feedback. J. Appl. Probab. **39**(3), 630–636 (2002)
2. Foley, R.D., Disney, R.L.: Queues with delayed feedback. Adv. Appl. Probab. **15**(1), 162–182 (1983)
3. Neuts, F.: Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. John Hopkins University Press, Baltimore (1981)
4. Melikov, A.Z., Ponomarenko, L.A., Kuliyeva, K.H.N.: Calculation of the characteristics of multichannel queuing system with pure losses and feedback. J. Autom. Inf. Sci. **47**(6), 19–29 (2015)
5. D'Avignon, G.R., Disney, R.L.: Queues with instantaneous feedback. Manag. Sci. **24**(2), 168–180 (1977)
6. Zaryadov, I.S., Scherbanskaya, A.A.: Time characteristics of queuing system with renovation and reservice. Physics **2**, 61–65 (2014). (in Russian), Bulletin of PFUR. Serires Mathematics. Information Sciences
7. Zaryadov, I.S., Korolkov, A.V., Milovanov, T.A., Sherbanskaya, A.A.: Mathematical model of calculating and analysis of characteristics of systems with generalized and repeating service. T-Comm Telecommun. Transp. **6**(8), 16–20 (2014). (in Russian)
8. Moiseeva, S.P., Zakhorol'naya, I.A.: Mathematical model of parallel retrial queueing of multiple requests. Optoelectronics Instrum. Data Process. **47**(6), 567–572 (2011)
9. Morozova, A.S., Moiseeva, S.P., Nazarov, A.A.: Research of QS with repeating requests and uunlimited number of servers by means of a method of limit decomposing. Comput. Technol. **13**(5), 88–92 (2005). (in Russian)
10. Leland, W.E., Willinger, W., Taqqu, M.S., Wilson, D.V.: On the self-similar nature of ethernet traffic. ACM SIGCOMM Comput. Commun. Rev. **47**(6), 202–213 (1995)
11. Klemm, A., Lindemann, C., Lohmann, M.: Modeling IP traffic using the batch Markovian arrival process (extendend version). Perform. Eval. **54**, 149–173 (2003)
12. Nazarov, A.A., Moiseeva, S.P.: Methods of Asymptotic Analysis in Queuing Theory. NTL, Tomsk (2006). (In Russian)