Study of the Queuing Systems $M|GI|N|\infty$

Ekaterina Lisovskaya and Svetlana Moiseeva(✉)

Tomsk State University, Lenina, 36, Tomsk, Russia
{ekaterina_lisovs,smoiseeva}@mail.ru
http://www.tsu.ru

Abstract. In this paper, we study the queuing system with unlimited queue and with N servers. We obtain the approximation of probability distribution of the number of customers in the system. We obtain the formula of the probability of immediate service and the characteristic function of a positive waiting time. The optimal number of servers can be determined by the obtained characteristics.

Keywords: Queuing systems · Waiting time · Approximation of the probability distribution · Queue

1 Introduction

Mathematical models of queuing systems (QS) is widely used in the solution of important practical problems arising in connection with the rapid development of communication systems, the emergence of information systems, the emergence of a variety and complexity of technological systems, the creation of automated control systems.

Multiserver QS are mathematical models of real systems and processes in the area of telecommunications, communication networks, etc. There are papers by modeling call-centers [1,2].

In this paper we consider queuing system $M|GI|N|\infty$. The system arrival process is distributed by Poisson law with rate λ. The system has N servers. Service times on each servers are i.i.d. with distribution function $A(x)$. The arriving customer occupies any free server or goes to the queue in case of all servers are busy.

Its known that for the system $M|M|N|K$ Erlang formulas have been obtained [3]. However, for general service time and infinity queue the obtained problem theoretically didnot solve, and an analytical solution is not possible, therefore, the task of obtaining the approximation of stationary probability distribution $P(i)$, $0 \leq i < \infty$ of the number of customers in the system $M|GI|N|\infty$. We obtain the formulas for the probability distribution of positive waiting time using the approximation.

A. Dudin et al. (Eds.): ITMM 2015, CCIS 564, pp. 175–184, 2015.
DOI: 10.1007/978-3-319-25861-4_15

2 Approximation of Probability Distribution of the Number of Customers in the System $M|GI|N|\infty$

Denote the number of customers in the system at time t by $i(t)$. Then $P(i) = P\{i(t) = i\}$ is the probability distribution of the number of customers in the system at time t.

Let π_i be an approximation of the probability distribution which is defined as a composite distribution [4]

$$\pi_i = \begin{cases} C_1 P_1(i), 0 \le i \le N, \\ C_2 P_2(i - N + 1), i \ge N. \end{cases} \tag{1}$$

The probabilities $P_1(i)$, where $0 \le i < N$, are the probabilities of the number of occupied servers in N-server QS with customers losses ($M|GI|N|0$), when all servers are busy. Erlang formula defines the probability $P_1(i)$ [5]

$$P_1(i) = \frac{\frac{(\lambda a)^i}{i!}}{\sum\limits_{k=0}^{N} \frac{(\lambda a)^k}{k!}}, \tag{2}$$

where $a = \int\limits_0^\infty (1 - A(x))dx$ is the average service time.

The probabilities $P_2(i)$ are defined when all servers are busy. In this case, the block of occupied servers is considered as a single and its service has distribution function $B(x)$. Therefore, the probabilities $P_2(i)$, where $i = 0, 1, ...$ are defined as the probabilities of the number of customers in the single-server system $M|GI|1|\infty$ with waiting.

In this case, the Pollaczek-Khinchin formula for the generating function can be use:

$$G(x) = \sum_{n=0}^{\infty} x^n P_2(n) = (1 - \lambda b) \frac{(1 - x)B^*(\lambda - \lambda x)}{B^*(\lambda - \lambda x) - x}. \tag{3}$$

To determine the distribution function $B(x)$ we consider the output process of serviced customers when all N servers are occupied.

3 Distribution of Sum of Independent Recurrent Process

Consider the sum of N independent recurrent process, with the same distribution functions $A(x)$.

Let τ be the value of the jump [5] for the total process. Then it is obvious that $\tau = \min(\tau_1, \tau_2, ..., \tau_N)$, where $\tau_1, \tau_2, ..., \tau_N$ are independent value of jump for total process.

Therefore

$$P\{\tau > x\} = 1 - \frac{1}{b}\int_0^x (1 - B(z))dz = P\{\min\{\tau_1, \tau_2, ..., \tau_N\} > x\}$$
$$= P\{\tau_1 > x\}P\{\tau_2 > x\}...P\{\tau_N > x\} = \left(1 - \frac{1}{a}\int_0^x (1 - A(z))dz\right)^N.$$

Hence, for the total process we have the following equation:

$$1 - \frac{1}{a}\int_0^x (1 - B(z))dz = \left(1 - \frac{1}{a}\int_0^x (1 - A(z))dz\right)^N,$$

then we differentiate equation by x and obtain the following formula:

$$B(x) = 1 - Nb\left(1 - \frac{1}{a}\int_0^x (1 - A(z))dz\right)^{N-1}\frac{1}{a}(1 - A(x)).$$

Knowing $\frac{1}{b} = N\frac{1}{a}$ [5], then in the system $M|GI|N|\infty$, the distribution function of the lengths of the intervals of the total process has the form:

$$B(x) = 1 - (1 - A(x))\left(1 - \frac{1}{a}\int_0^x (1 - A(z))dz\right)^{N-1}. \tag{4}$$

and the density distribution has the following form:

$$b(x) = \left\{A'(x)\left(1 - \frac{1}{a}\int_0^x (1 - A(z))dz\right) + \frac{N-1}{a}(1 - A(x))^2\right\}$$
$$\times\left(1 - \frac{1}{a}\int_0^x (1 - A(z))dz\right)^{N-2}.$$

4 Expansion of the Function Pollaczek-Khinchin

Probabilities $P_2(i)$ can be found using the inverse Fourier transform, or expending function (3) to a power series in x.

To determine the probability by the second method, we write the following expansion

$$B^*(\lambda - \lambda x) = \int_0^\infty e^{-(\lambda - \lambda x)z}dB(z) = \int_0^\infty e^{-\lambda z}e^{-\lambda xz}dB(z)$$
$$= \int_0^\infty e^{-\lambda z}\sum_{n=0}^\infty \frac{(\lambda z)^n}{n!}dB(z) = \sum_{n=0}^\infty x^n \int_0^\infty e^{-\lambda z}\frac{(\lambda z)^n}{n!}dB(z).$$

Denote

$$\beta_n = \int\limits_0^\infty e^{-(\lambda z)} \frac{(\lambda z)^n}{n!} dB(z),$$

we obtain the expansion

$$B^*(\lambda - \lambda x) = \sum_{n=0}^{\infty} x^n \beta_n.$$

Hence

$$(1-x)B^*(\lambda - \lambda x) = \sum_{n=0}^{\infty} x^n \beta_n - \sum_{n=0}^{\infty} x^{n+1} \beta_n = \sum_{n=0}^{\infty} x^n \beta_n - \sum_{n=1}^{\infty} x^n \beta_{n-1}$$
$$= \beta_0 + \sum_{n=1}^{\infty} x^n (\beta_n - \beta_{n-1}) = \sum_{n=1}^{\infty} x^n b_n.$$

where $b_0 = \beta_0$, $b_n = \beta_n - \beta_{n-1}$.

The denominator of the expression (3) is writen in the form

$$(B^*(\lambda - \lambda x) - x)^{-1} = \sum_{n=0}^{\infty} x^n \alpha_n. \tag{5}$$

To determine α_n we rewrite the formula (5) as:

$$1 = (B^*(\lambda - \lambda x) - x) \sum_{n=0}^{\infty} x^n \alpha_n = \sum_{n=0}^{\infty} x^n \alpha_n \sum_{n=0}^{\infty} x^n \beta_n - \sum_{n=0}^{\infty} x^{n+1} \alpha_n$$
$$= \sum_{n=0}^{\infty} x^n \sum_{k=0}^{n} \alpha_k \beta_{n-k} - \sum_{n=1}^{\infty} x^n \alpha_{n-1} = \alpha_0 \beta_0 + \sum_{n=1}^{\infty} x^n \sum_{k=0}^{n} \alpha_k \beta_{n-k} - \alpha_{n-1}.$$

Equating coefficients of same powers of x in this expression, we obtain recurrence formulas:

$$\alpha_0 = \frac{1}{\beta_0},\ \alpha_n = \frac{1}{\beta_0} \left[\alpha_{n-1} - \sum_{k=0}^{n-1} \alpha_k \beta_{n-k} \right].$$

Considering the expansion of the function $G(x)$, can be written expression

$$G(x) = \sum_{n=0}^{\infty} x^n P_2(n) = (1-\lambda b) \frac{(1-x)B^*(\lambda - \lambda x)}{B^*(\lambda - \lambda x) - x} = (1-\lambda b) \sum_{i=0}^{\infty} x^i b_i \sum_{i=0}^{\infty} x^i \alpha_i$$
$$= (1-\lambda b) \sum_{i=0}^{\infty} x^i \sum_{k=0}^{i} \alpha_k b_{i-k}.$$

Thus

$$P_2(i) = (1-\lambda b) \sum_{k=0}^{i} \alpha_k b_{i-k}.$$

5 Finding of Constants

Constants C_1 and C_2 can be found from the normalization condition and the conditions of "stitching":

$$\begin{cases} \sum\limits_{i=0}^{\infty} \pi_i = 1, \\ C_1 P_1(N) = C_2 P_2(1), \end{cases}$$

we obtain

$$1 = \sum_{i=0}^{\infty} \pi_i = C_1 \sum_{i=0}^{N} P_1(i) + C_2 \sum_{i=N+1}^{\infty} P_2(i-N+1) = C_1 + C_2 \sum_{n=2}^{\infty} P_2(n)$$
$$= C_1 + C_2(1 - (P_2(0) + P_2(1))),$$

thus

$$C_1 = \frac{P_2(1)}{P_2(1) + P_1(N)(1 - (P_2(0) + P_2(N)))},$$
$$C_2 = \frac{P_1(N)}{P_2(1) + P_1(N)(1 - (P_2(0) + P_2(N)))}. \tag{6}$$

So expression (1) has the form:

$$\pi_i = \begin{cases} \frac{P_2(1)}{P_2(1)+P_1(N)(1-(P_2(0)+P_2(N)))} P_1(i), 0 \le i \le N, \\ \frac{P_1(N)}{P_2(1)+P_1(N)(1-(P_2(0)+P_2(N)))} P_2(i-N+1), i > N. \end{cases}$$

6 Probability of Immediate Service

Let τ be the waiting time of customer service start. Using (1) the probability of immediate service can be written as

$$P_0 = \sum_{i=0}^{N-1} \pi_i = C_1 \sum_{i=0}^{N-1} P_1(i) = C_1(1 - P_1(i))$$
$$= \frac{P_2(1)(1 - P_1(N))}{P_2(1) + P_1(N)[1 - (P_2(0) + P_2(1))]}, \tag{7}$$

where considering expressions (2) and (3) the following equalities

$$P_2(0) = G(0) = 1 - \lambda b,$$

$$P_2(1) = G^{'}(0) = (1 - \lambda b) \frac{1 - B^*(\lambda)}{B^*(\lambda)},$$

$$P_1(N) = \frac{\frac{(\lambda a)^N}{N!}}{\sum\limits_{i=0}^{N} \frac{(\lambda a)^i}{i!}}.$$

7 Probability Distribution of a Positive Waiting Time

If the customer arrives in the system at time when all servers are busy, then its waiting time $\tau > 0$ and this value we call as a positive waiting time τ^+.

We find the conditional probability distribution $P_q(m)$, where $m > 0$ that there are m customers in the queue considering that all servers are busy.

Using expression (1), we written:

$$\begin{aligned} P_q(m) &= \frac{\pi_{N+m}}{\sum\limits_{i=0}^{\infty} \pi_{N+i}} = \frac{C_2 P_2(1+m)}{C_2 \sum\limits_{i=0}^{\infty} P_2(1+m)} \\ &= \frac{P_2(1+m)}{(1-P_2(0))} = \frac{1}{\lambda b} P_2(1+m). \end{aligned} \tag{8}$$

Expression (9) is the conditional probability distribution that there are m customers in the queue considering that all servers in the system are busy

$$P_q(m) = \frac{1}{\lambda b} P_2(m+1). \tag{9}$$

We find the generating function $G_q(x)$ of this distribution

$$\begin{aligned} G_q(x) &= \sum_{m=0}^{\infty} x^m P_q(m) = \sum_{m=0}^{\infty} x^m \frac{1}{\lambda b} P_2(m+1) \\ &= \frac{1}{\lambda b x} \sum_{\nu=1}^{\infty} x^{\nu} P_q(\nu) = \frac{1}{\lambda b x} \left[G(x) - P_0\right] \\ &= \frac{1}{\lambda b x} \left[G(x) - (1-\lambda b)\right] \\ &= \frac{1}{\lambda b x} \left[(1-\lambda b) \frac{(1-x) B^*(\lambda - \lambda x)}{B^*(\lambda - \lambda x) - x} - (1-\lambda b)\right] \\ &= \frac{1-\lambda b}{\lambda b x} \frac{(1-x) B^*(\lambda - \lambda x) - B^*(\lambda - \lambda x) + x}{B^*(\lambda - \lambda x) - x} \\ &= \frac{1-\lambda b}{\lambda b} \frac{1 - B^*(\lambda - \lambda x)}{B^*(\lambda - \lambda x) - x}, \end{aligned}$$

then

$$G_q(x) = \frac{1-\lambda b}{\lambda b} \frac{1 - B^*(\lambda - \lambda x)}{B^*(\lambda - \lambda x) - x}. \tag{10}$$

This generating function is obtained on the period when all servers are busy. In this condition, the N-server block of servers is defined by the distribution function $A(x)$ is permissible to replace the single-server with the distribution function $B(x)$ from formula (4).

Customer arriving in the system when all servers are busy finds m customers in the queue with probability $P_q(m)$. So, the waiting time consists of the total time service of customers, each having the distribution function $B(x)$ from formula (4) and residual service time of one customer with the distribution function

$$B_0(x) = \frac{1}{b}\int\limits_0^x (1 - B(z))\,dz.$$

We denote residual service time by ξ_0 and service times of the first, the second and the m-th customers in the queue by $\xi_1, \xi_2, ..., \xi_m$ respectively. Then the waiting time can be determined by

$$\tau^+ = \xi_0 + \xi_1 + \xi_2 + ... + \xi_m.$$

We find the characteristic function $h(u)$ of the positive waiting time customers in the system $M|GI|N|\infty$. Using total probability law for mean, we can write

$$h(u) = M\left\{e^{jut}\right\} = \sum_{m=0}^{\infty} M\left\{exp\left\{ju(\xi_0 + ... + \xi_m)\right\}|m(t) = m\right\} P_q(m)$$
$$= \sum_{m=0}^{\infty} M\left\{e^{ju\xi_0}\right\}\left(M\left\{e^{ju\xi_1}\right\}\right)^m P_q(m) = \phi(u)^m P_q(m),$$

where $\phi_0(u)$ and $\phi(u)$ are characteristic functions of the residual and total times service of one customer, here

$$\phi(u) = \int\limits_0^{\infty} e^{jux} dB(x). \tag{11}$$

The last equation for $h(u)$ we rewrite as

$$h(u) = \phi_0(u)G_q(\phi(u)) = \phi_0(u)\frac{1-\lambda b}{\lambda b}\frac{1 - B^*(\lambda - \lambda\phi(u))}{B^*(\lambda - \lambda\phi(u)) - \phi(u)}. \tag{12}$$

So

$$B_0(x) = \frac{1}{b}\int\limits_0^x (1 - B(z))\,dz,$$

then

$$\phi_0(u) = \int\limits_0^{\infty} e^{jux} dB_0(x) = \frac{1}{b}\int\limits_0^x e^{jux}(1 - B(x))dx = \frac{1}{jub}(\phi(u) - 1),$$

therefore, the characteristic function $h(u)$ of formula (12) is written as

$$h(u) = \frac{1}{jub}(1 - \lambda b)\frac{\phi(u) - 1}{\lambda b}\frac{1 - B^*(\lambda - \lambda\phi(u))}{B^*(\lambda - \lambda\phi(u)) - \phi(u)}. \tag{13}$$

Here $\phi(u)$ has the form (11).

Formulas (5) and (13) completely characterize the waiting time of customer in the queue N-server system $M|GI|N|\infty$.

8 Mean of the Positive Waiting Time

Applying the characteristic function $h(u)$ from formula (13), the mean of positive waiting time customer in the queue is written in the form

$$\bar{\tau}^{+} = \frac{1}{j}h^{'}(u)|_{u=0} = \frac{b_2}{2b(1-\lambda b)}, \tag{14}$$

where b is the mean, and b_2 is the second initial moment, defined by distribution function $B(x)$ from (4).

9 Optimal Number of Servers in the Multiserver System

In order for steady-state regime to exists in N-server queuing system with the waiting, it is necessary that the system load $\rho = \lambda b = \lambda \frac{a}{N}$ is less than one.

Therefore, N has to satisfy inequality

$$N > \lambda a. \tag{15}$$

The optimal value N_{opt} of the number of servers is defined by criteria:

$$N_{opt}\left[\min_{N}\{N : P(\tau > \tau_{\max}) \leq \delta\}\right].$$

Let τ be waiting time of customer service start, $\tau_{\max}$ is the upper limit waiting time customer service start, δ is allowable share of customers who will wait for the start of service longer than $\tau_{\max}$. The condition $P(\tau > \tau_{\max}) \leq \delta$ can be replaced by the following equivalent condition

$$(1 - P_0)P(\tau^{+} > \tau_{\max}) \leq \delta, \tag{16}$$

where the probability of immediate service is given by (5).

Applying inverse Fourier transform to the function $h(u)$, the probability $P(\tau^{+} > \tau_{\max})$ can be written as the following integration formula:

$$P(\tau^{+} > \tau_{\max}) = 1 - \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1 - e^{-ju\tau_{\max}}}{ju}h(u)du. \tag{17}$$

Here $h(u)$ has the form (13), and

$$\phi(u) = \int_{0}^{\infty} e^{jux}dB(x) = 1 + ju\int_{0}^{\infty} e^{jux}[1 - B(x)]dx,$$

$$B^{*}(\alpha) = \int_{0}^{\infty} e^{-\alpha x}dx = 1 - \alpha\int_{0}^{\infty} e^{-\alpha x}[1 - B(x)]dx = \phi(j\alpha), \tag{18}$$

and hence

$$\phi(u) = B^*(-ju). \tag{19}$$

The most consuming is the finding of probability $P(\tau^+ > \tau_{\max})$ in integral formula (17), because it requires numerical calculation of three-dimensional integrals: firstly, finding the function $B(x)$ by the formula (4), secondly, finding the Fourier and Laplace transform by the formulas (18)–(19), thirdly, the calculations probability by the formula (17).

This problem is solved by considering for this task under heavy load.

10 Asymptotic Analysis of a Positive Waiting Time Under Heavy Load

The characteristic function of a positive waiting time $h(u)$ has the form (13). We find its limit value under heavy load, when $1 - \lambda b = \varepsilon$ and $\varepsilon \to 0$. For this, in expression (13) performs substitutions

$$u = \varepsilon\omega,\ h(u) = H(\omega, \varepsilon),$$

we obtain the following function

$$H(\omega, \varepsilon) = \frac{(\phi(\varepsilon\omega) - 1)(1 - \lambda b)}{\lambda b^2 j\varepsilon\omega} \frac{1 - B^*(\lambda - \lambda\phi(\varepsilon\omega))}{B^*(\lambda - \lambda\phi(\varepsilon\omega)) - \phi(\varepsilon\omega)}. \tag{20}$$

Let us find its limit for $\varepsilon \to 0$. We can write

$$\phi(\varepsilon\omega) = Me^{j\varepsilon\omega\xi} = 1 + j\varepsilon\omega\xi b + \frac{(j\varepsilon\omega\xi)^2}{2} b_2 + o(\varepsilon^3),$$

$$\lambda - \lambda\phi(\varepsilon\omega) = -j\varepsilon\omega\xi b - \frac{(j\varepsilon\omega\xi)^2}{2}\lambda b_2 + o(\varepsilon^3),$$

$$B^*(\alpha) = Me^{-\alpha\xi} = 1 - \alpha b + \frac{\alpha^2}{2} b_2 + o(\alpha^3),$$

$$\begin{aligned} B^*(\lambda - \lambda\phi(\varepsilon\omega)) &= B^*\left(j\varepsilon\omega b - \frac{(j\varepsilon\omega)^2}{2}\lambda b_2\right) = B^*\left(-j\varepsilon\omega(1-\varepsilon) - \frac{(j\varepsilon\omega)^2}{2}\frac{b_2}{b}\right) \\ &= 1 + b\left(j\varepsilon\omega(1-\varepsilon) + \frac{(j\varepsilon\omega)^2}{2}\frac{b_2}{b}\right) + \frac{(j\varepsilon\omega)^2}{2} b_2 + o(\varepsilon^3). \end{aligned}$$

Substituting these expressions in expression (20), we obtain:

$$\begin{aligned} H(\omega, \varepsilon) &= \frac{1 + j\varepsilon\omega b - 1}{\lambda b^2 j\varepsilon\omega}\varepsilon \frac{1 - j\varepsilon\omega b}{1 + j\varepsilon\omega b - j\varepsilon\omega b\varepsilon + (j\varepsilon\omega)^2 b_2 - \left[1 + j\varepsilon\omega b + \frac{(j\varepsilon\omega)^2}{2} b_2\right]} \\ &= \frac{1}{\lambda b}\varepsilon \frac{-j\varepsilon\omega b}{-j\varepsilon\omega b\varepsilon + \frac{(j\varepsilon\omega)^2}{2} b_2} = \frac{1}{1 - j\omega\frac{b_2}{2b}}. \end{aligned}$$

Performing here the following inverse transformation

$$\omega = \frac{u}{\varepsilon} = \frac{u}{1-\lambda b},$$

we obtain the approximate expression for $\lambda b \uparrow 1$

$$h(u) \approx \frac{1}{1 - ju\frac{b_2}{2b(1-\lambda b)}}.$$

for the characteristic function, which has the form of the characteristic function of the exponential distribution with rate

$$\gamma = \frac{2b(1-\lambda b)}{b_2}. \tag{21}$$

Note that the mean $1/\gamma$ of the asymptotic distribution is equal to the mean $\bar{\tau}^+$ of positive waiting time (14).

Therefore, the probability $P(\tau^+ > \tau_{\max})$ under heavy load can be defined from the formula:

$$P(\tau^+ > \tau_{\max}) = e^{-\gamma\tau_{\max}} = \exp\left\{-\frac{2b(1-\lambda b)}{b_2}\tau_{\max}\right\}. \tag{22}$$

11 Conclusion

In this paper, we study the queuing system $M|GI|N|\infty$. We obtain the approximation of probability distribution of the number of customers in the system. We derive the formula of the probability of immediate service and the characteristic function of a positive waiting time. The optimal number of servers can be determined by the obtained characteristics.

Acknowledgments. The work is performed under the state order of the Ministry of Education and Science of the Russian Federation (No. 1.511.2014/K).

References

1. Brown, L., Gans, N., Mandelbaum, A., et al.: Statistical analysis of a telephone call center: a queueing-science perspective. J. Am. Stat. Assoc. **100**, 36–50 (2005)
2. Jouini, O., Aksin, Z., Dallery, Y.: Call centers with delay information: models and insights. Manuf. Serv. Operat. Manag. **13**, 534–548 (2011)
3. Kleinrock, L., Grushko, I.I., Neiman, V.I.: Queueing Theory. Translated from English. Mechanical Engineering (1979)
4. Lisovskaya, E.Yu., Moseeva, S.P.: Study of the process the number of customers in the system $M|GI|N|\infty$. Probability theory, stochastic processes, mathematical statistics and applications. In: Proceedings of the International Scientific Conference Devoted to the 80th Anniversary of Professor, Doctor of Physical and Mathematical Sciences Gennady Alekseevich Medvedev, pp. 123–127 (2015). (in Russian)
5. Nazarov, A., Terpugov, A.F.: Queueing Theory: Textbook. Publishing House of the NTL, Tomsk (2010). (in Russian)