ИЗВЕСТИЯ ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЙ
Т. 58, № 11/2 ФИЗИКА 2015

УДК 519.872

Е.Ю. ЛИСОВСКАЯ, С.П. МОИСЕЕВА

РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ ПОЛОЖИТЕЛЬНОГО ВРЕМЕНИ ОЖИДАНИЯ В СИСТЕМЕ M|GI|N|∞ [1]

Для системы массового обслуживания M|GI|N|∞ показан вид аппроксимации распределения вероятностей числа заявок. Получены формулы для вероятности немедленного обслуживания и для распределения вероятностей положительного времени ожидания в очереди, а также асимптотическое распределение вероятностей положительного времени ожидания в условии большой загрузки.

Ключевые слова: *системы массового обслуживания, время ожидания, аппроксимация распределения вероятностей, очередь.*

Введение

Математические модели систем массового обслуживания (СМО) широко применяют при решении важных практических задач, возникающих в связи с бурным развитием систем коммуникаций, возникновением информационно-вычислительных систем, появлением и усложнением разнообразных технологических систем, созданием автоматизированных систем управления.

Многолинейные СМО могут являться математическими моделями реальных систем и процессов в области телекоммуникаций, сетях связи и т.д. Известны работы по моделированию call-центров [1, 2].

В данной работе рассматривается система с произвольным временем обслуживания на приборах и бесконечной очередью для ожидания обслуживания (M|GI|N|∞).

Известно, что для системы M|M|N|K получены формулы Эрланга [3]. Вместе с тем для произвольного времени обслуживания и бесконечной очереди полученная задача теоретически не решена и представить аналитическое решение невозможно, поэтому ставится задача построения аппроксимации стационарного распределения вероятностей $P(i)$, $0 \le i < \infty$, числа заявок в рассматриваемой системе M|GI|N|∞ и получения формул для распределения вероятностей положительного времени ожидания с помощью полученной аппроксимации.

Математическая модель

На вход системы поступает простейший поток с параметром λ. Поступающая заявка занимает любой из свободных приборов или становится в очередь в случае, когда все приборы заняты. Система имеет *N* обслуживающих приборов, время обслуживания каждой заявки является случайной величиной с произвольной функцией распределения $A(x)$, одинаковой для всех приборов.

Аппроксимация распределения вероятностей числа заявок в системе M|GI|N|∞

Обозначим $i(t)$ – число заявок в системе в момент времени t. Тогда $P(i) = P\{i(t) = i\}$ – распределение вероятностей числа заявок в системе в момент времени t.

Пусть π_i – аппроксимация распределения вероятностей $P(i)$, которая определяется в виде составного распределения [4]:

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

Вероятности $P_1(i), 0 \le i \le N$, – вероятности числа занятых приборов в N-линейной СМО (M|GI|N|0) с потерями заявок, когда все приборы заняты. Тогда их можно определить формулами Эрланга [5]

[1] Работа выполнена в рамках госзаказа № ~1.511.2014/К Министерства образования и науки Российской Федерации.

$$P_1(i) = \frac{(\lambda a)^i}{i!} \Bigg/ \left(\sum_{k=0}^{N} \frac{(\lambda a)^k}{k!} \right), \tag{2}$$

где $a = \int_0^\infty (1 - A(x))dx$ – среднее время обслуживания.

Вероятности $P_2(i)$ определяются для случая, когда все приборы заняты. В таком случае блок занятых приборов представляется как один прибор, который обслуживает случайное время с функцией распределения $B(x)$. Поэтому вероятности $P_2(i), i = 0,1,...$ определим как вероятности числа заявок в однолинейной системе M|GI|1|∞ с ожиданием.

В таком случае, воспользуемся функцией Поллачека – Хинчина для производящей функции:

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

Вероятности $P_2(i)$ можно найти с помощью обратного преобразования Фурье или раскладывая функцию (3) в ряд по степеням x. Константы C_1 и C_2 найдем из условия нормировки и условия «сшивания»:

$$C_1 = \frac{P_2(1)}{P_2(1) + P_1(N)\big(1 - (P_2(0) + P_2(N))\big)}, \quad C_2 = \frac{P_1(N)}{P_2(1) + P_1(N)\big(1 - (P_2(0) + P_2(N))\big)}. \tag{4}$$

Вероятность немедленного обслуживания

Пусть τ – время ожидания заявки до начала обслуживания. Вероятность немедленного обслуживания, учитывая (1), можно записать в виде

$$P_0 = \sum_{i=0}^{N-1} \pi_i = C_1 \sum_{i=0}^{N-1} P_1(i) = C_1\big(1 - P_1(N)\big) = \frac{P_2(1)\big(1 - P_1(N)\big)}{P_2(1) + P_1(N)\big[1 - (P_2(0) + P_2(1))\big]}, \tag{5}$$

где в силу (2) и (3) выполняются равенства

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

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

$$P_1(N) = \frac{(\lambda a)^N}{N!} \Bigg/ \sum_{i=0}^{N} \frac{(\lambda a)^i}{i!}.$$

Распределение вероятностей положительного времени ожидания

Если заявка поступает в систему в тот момент времени, когда все приборы заняты, тогда ее время ожидания $\tau > 0$, и эту величину будем называть положительным временем ожидания τ^+.

Найдем условное распределение вероятностей $P_{\text{ож}}(m)$, $m \geq 0$, того, что в очереди m заявок, при условии, что все приборы системы заняты.

Применяя равенство (1), запишем

$$P_{\text{ож}}(m) = \pi(N+m) \Big/ \sum_{i=0}^{\infty} \pi(N+i) = C_2 P_2(1+m) \Big/ C_2 \sum_{i=0}^{\infty} P_2(1+m) =$$

$$=P_2(1+m)\Big/\big(1-P_2(0)\big)=\frac{1}{\lambda b}P_2(1+m). \tag{6}$$

То есть условное распределение вероятностей $P_{\text{ож}}(m)$ того, что в очереди m заявок, при условии, что все приборы системы заняты, имеет вид

$$P_{\text{ож}}(m)=\frac{1}{\lambda b}P_2(m+1). \tag{7}$$

Найдем производящую функцию $G_{\text{ож}}(x)$ этого распределения

$$G_{\text{ож}}(x)=\sum_{m=0}^{\infty}x^m P_{\text{ож}}(m)=\sum_{m=0}^{\infty}x^m\frac{1}{\lambda b}P_2(m+1)=\frac{1}{\lambda bx}\sum_{\nu=1}^{\infty}x^{\nu}P_{\text{ож}}(\nu)=\frac{1}{\lambda bx}\big[G(x)-P_0\big]=$$

$$=\frac{1}{\lambda bx}\big[G(x)-(1-\lambda b)\big]=\frac{1}{\lambda bx}\left[(1-\lambda b)\frac{(1-x)B^*(\lambda-\lambda x)}{B^*(\lambda-\lambda x)-x}-(1-\lambda b)\right]=$$

$$=\frac{1-\lambda b}{\lambda bx}\frac{(1-x)B^*(\lambda-\lambda x)-B^*(\lambda-\lambda x)+x}{B^*(\lambda-\lambda x)-x}=\frac{1-\lambda b}{\lambda b}\frac{1-B^*(\lambda-\lambda x)}{B^*(\lambda-\lambda x)-x},$$

то есть

$$G_{\text{ож}}(x)=\frac{1-\lambda b}{\lambda b}\frac{1-B^*(\lambda-\lambda x)}{B^*(\lambda-\lambda x)-x}. \tag{8}$$

Эта производящая функция получена на периоде занятости рассматриваемой системы, то есть когда заняты все ее приборы. В этом условии N-линейный блок обслуживания, определенный функцией распределения $A(x)$, допустимо заменить однолинейным с функцией распределения времени обслуживания $B(x)$ вида

$$B(x)=1-\big(1-A(x)\big)\left(1-\frac{1}{a}\int_0^x\big(1-A(z)\big)dz\right)^{N-1}. \tag{9}$$

Заявка, поступившая в систему на периоде ее занятости, с вероятностью $P_{\text{ож}}(m)$ обнаруживает в очереди m заявок, поэтому ее время ожидания τ^+ складывается из суммарного времени обслуживания m заявок, каждое из которых имеет функцию распределения $B(x)$ из (9) и времени дообслуживания одной заявки, имеющего функцию распределения

$$B_0(x)=\frac{1}{b}\int_0^x\big(1-B(z)\big)dz.$$

Обозначив ξ_0 – остаточное время обслуживания, а ξ_1, $\xi_2,\ldots,$ ξ_m – времена обслуживания первой, второй и m-й заявок в очереди, время ожидания можно определить как

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

Найдем характеристическую функцию $h(u)$ положительного времени ожидания заявок в системе M|GI|N|∞. Применяя формулу полной вероятности для математических ожиданий, можно записать

$$h(u)=M\left\{e^{ju\tau^+}\right\}=\sum_{m=0}^{\infty}M\left\{\exp\left\{ju(\xi_0+\xi_1+\ldots+\xi_m)\right\}\mid m(t)=m\right\}P_{ож}(m)=$$

$$=\sum_{m=0}^{\infty}M\left\{e^{ju\xi_0}\right\}\left(M\left\{e^{ju\xi_m}\right\}\right)^m P_{ож}(m)=\varphi_0(u)\sum_{m=0}^{\infty}\varphi(u)^m P_{ож}(m),$$

где $\varphi_0(u)$ и $\varphi(u)$ – характеристические функции остаточного и полного времени обслуживания одной заявки, здесь

$$\varphi(u)=\int_0^\infty e^{jux}dB(x). \tag{10}$$

Последнее равенство для $h(u)$ перепишем в виде

$$h(u)=\varphi_0(u)G_0(\varphi(u))=\varphi_0(u)\frac{1-\lambda b}{\lambda b}\frac{1-B^*(\lambda-\lambda\varphi(u))}{B^*(\lambda-\lambda\varphi(u))-\varphi(u)}. \tag{11}$$

Так как

$$B_0(x)=\frac{1}{b}\int_0^x(1-B(z))dz,$$

то

$$\varphi_0(u)=\int_0^\infty e^{jux}dB_0(x)=\frac{1}{b}\int_0^x e^{jux}(1-B(x))dx=\frac{1}{jub}(\varphi(u)-1),$$

поэтому характеристическую функцию $h(u)$ из (11) можно окончательно записать в виде

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

Здесь $\varphi(u)$ имеет вид (10).

Формулы (5) и (12) полностью характеризуют время ожидания заявки в очереди N-линейной системы M|GI|N|∞. Формула (5) для P_0 определяет вероятность того, что время ожидания будет нулевым, а $h(u)$ из (12) является характеристической функцией распределения вероятностей положительного времени ожидания τ^+ заявки в очереди.

Среднее значение положительного времени ожидания

Применяя характеристическую функцию $h(u)$ из (12), среднее значение $\bar{\tau}^+$ положительного времени ожидания заявки в очереди запишем в виде

$$\bar{\tau}^+=\frac{1}{j}h'(u)\bigg|_{u=0}=\frac{b_2}{2b(1-\lambda b)}, \tag{13}$$

где b и b_2 – математическое ожидание и второй начальный момент случайной величины, определяемой функцией распределения $B(x)$ из (9).

Выбор оптимального числа приборов в многолинейной системе

Для того чтобы в N-линейной системе массового обслуживания с ожиданием существовал стационарный режим, необходимо, чтобы ее загрузка $\rho=\lambda b=\lambda\frac{a}{N}$ была меньше единицы. Поэтому для величины N необходимо выполнение неравенства

$$N>\lambda a. \tag{14}$$

Искомое оптимальное значение N_{opt} числа обслуживающих приборов определяется следующим критерием:

$$N_{\text{opt}}=\left[\min_N\{N:\mathrm{P}(\tau>\tau_{\max})\le\delta\}\right].$$

Здесь τ – время ожидания клиентом начала обслуживания; $\tau_{\max}$ – верхний порог времени ожидания клиентом начала обслуживания; δ – допустимая доля клиентов, которые будут ожидать начала обслуживания дольше, чем $\tau_{\max}$. Условие $\mathrm{P}(\tau>\tau_{\max})\le\delta$ можно заменить эквивалентным условием

$$\left(1-P_0\right)\cdot \mathrm{P}\left(\tau^+ > \tau_{\max}\right) \le \delta \,, \tag{15}$$

где вероятность немедленного обслуживания P_0 определяется формулой (5).

Применяя к функции $h(u)$ обратное преобразование Фурье, вероятность $\mathrm{P}\left(\tau^+ > \tau_{\max}\right)$ запишем в виде следующей интегральной формулы:

$$\mathrm{P}\left(\tau^+ > \tau_{\max}\right) = 1 - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1-e^{-ju\tau_{\max}}}{ju} h(u)\,du \,. \tag{16}$$

Здесь $h(u)$ определяется формулой (12), причем

$$\varphi(u) = \int_0^{\infty} e^{jux} dB(x) = 1 + ju\int_0^{\infty} e^{jux}\left[1-B(x)\right]dx \,,$$

$$B^*(\alpha) = \int_0^{\infty} e^{-\alpha x} dB(x) = 1 - \alpha\int_0^{\infty} e^{-\alpha x}\left[1-B(x)\right]dx = \varphi(j\alpha) \tag{17}$$

и, следовательно,

$$\varphi(u) = B^*(-ju) \,. \tag{18}$$

Наиболее трудоемким является нахождения вероятности $\mathrm{P}\left(\tau^+ > \tau_{\max}\right)$ по интегральной формуле (16), так как он требует численного нахождения трехмерного интеграла: первый – при нахождении функции $B(x)$ по формуле (9), второй – при нахождении преобразований Фурье и Лапласа по формулам (17)–(18), третий – при вычислениях непосредственно по формуле (16).

Эта проблема кардинально решается рассмотрением достаточно естественного для данной задачи условия большой загрузки.

Асимптотическое распределение вероятностей положительного времени ожидания в условии большой загрузки

Характеристическая функция положительного времени ожидания $h(u)$ имеет вид (12). Найдем ее предельное значение в условии большой загрузки, когда $1-\lambda b = \varepsilon$ и $\varepsilon \to 0$. Для этого в выражении (12) выполним замены

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

получим функцию

$$H(w,\varepsilon) = \frac{\left(\varphi(\varepsilon w)-1\right)\left(1-\lambda b\right)}{\lambda b^2 j\varepsilon w} \frac{1-B^*\left(\lambda-\lambda\varphi(\varepsilon w)\right)}{B^*(\lambda-\lambda\varphi(\varepsilon w))-\varphi(\varepsilon w)} \,. \tag{19}$$

Найдем ее предел при $\varepsilon \to 0$. Можно записать

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

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

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

$$B^*\left(\lambda-\lambda\varphi(\varepsilon w)\right) = B^*\left(-j\varepsilon w\lambda b - \frac{(j\varepsilon w)^2}{2}\lambda b_2\right) = B^*\left(-j\varepsilon w(1-\varepsilon) - \frac{(j\varepsilon w)^2}{2}\frac{b_2}{b}\right) =$$

$$= 1 + b\left(j\varepsilon w(1-\varepsilon) + \frac{(j\varepsilon w)^2}{2}\frac{b_2}{b}\right) + \frac{(j\varepsilon w)^2}{2} b_2 + o(\varepsilon^3) \,.$$

Подставляя эти выражения в (19), получаем

$$H(w,\varepsilon)=\frac{1+j\varepsilon wb-1}{\lambda b^2 j\varepsilon w}\varepsilon\frac{1-1-bj\varepsilon w}{1+j\varepsilon wb-bj\varepsilon w\varepsilon+(j\varepsilon w)^2 b_2-\left[1+j\varepsilon wb+\frac{(j\varepsilon w)^2}{2}b_2\right]}=$$

$$=\frac{1}{\lambda b}\varepsilon\frac{-bj\varepsilon w}{-bj\varepsilon w\varepsilon+\frac{(j\varepsilon w)^2}{2}b_2}=\frac{1}{1-jw\frac{b_2}{2b}}.$$

Выполняя здесь обратное преобразование

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

получим приближенное (предельное при $\lambda b\uparrow 1$) выражение

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

для характеристической функции $h(u)$, которое имеет вид характеристической функции экспоненциального распределения с параметром

$$\gamma=\frac{2b(1-\lambda b)}{b_2}. \tag{20}$$

Отметим, что среднее значение $1/\gamma$ этого асимптотического распределения совпадает с найденным ранее средним значением $\bar{\tau}^{+}$ положительного времени ожидания (13).

Следовательно, искомая вероятность $\mathrm{P}\left(\tau^{+}>\tau_{\max}\right)$ в условии большой загрузки может быть найдена по формуле

$$\mathrm{P}\left(\tau^{+}>\tau_{\max}\right)=e^{-\gamma\tau_{\max}}=\exp\left\{-\frac{2b(1-\lambda b)}{b_2}\tau_{\max}\right\}. \tag{21}$$

Заключение

Таким образом, в настоящей работе показан вид аппроксимации распределения вероятностей числа заявок в системе M|GI|N|∞. Получены формулы для вероятности немедленного обслуживания и для распределения вероятностей положительного времени ожидания в очереди, а также асимптотическое распределение вероятностей положительного времени ожидания в условии большой загрузки.

СПИСОК ЛИТЕРАТУРЫ

1. Brown L., Gans N., Mandelbaum A., et al. Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective // Journal of the American Statistical Association. – 2005. – V. 100. – No. 469. – P. 100.
2. Jouini O., Aksin Z., Dallery Y. Call Centers with Delay Information: Models and Insights // Manufacturing & Service Operations Management. – 2011. – V. 13. – P. 534–548.
3. Клейнрок Л. Теория массового обслуживания: пер с англ. И.И. Грушко; ред. В.И. Нейман. – М.: Машиностроение, 1979. – 432 с.
4. Лисовская Е.Ю., Моисеева С.П. Исследование процесса числа заявок в системе M|GI|N|∞ // Теория вероятностей, случайные процессы, математическая статистика и приложения: материалы Международной научной конференции, посвященной 80-летию профессора, доктора физико-математических наук Геннадия Алексеевича Медведева. – Минск, 2015. – С. 123–127.
5. Назаров А.А., Терпугов А.Ф. Теория массового обслуживания: учеб. пособие. – Томск: Изд-во НТЛ, 2010. – 228 с.

Национальный исследовательский Томский государственный университет,
г. Томск, Россия
E-mail: ekaterina_lisovs@mail.ru

Поступила в редакцию 14.10.15.

Лисовская Екатерина Юрьевна, магистрантка;
Моисеева Светлана Петровна, д.ф.-м.н., доцент.

E.Yu. LISOVSKAYA, S.P. MOISEEVA

M|GI|N|∞ POSITIVE WAITING TIME PROBABILITY DISTRIBUTION

Abstract. In this paper, we study the queuing system M|GI|N|∞. 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.*

REFERENCES

1. Brown L., Gans N., Mandelbaum A. et al. Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective. *Journal of the American Statistical Association*, 2005, vol. 100, no. 469, p. 100.
2. Jouini O., Aksin Z., Dallery Y. Call Centers with Delay Information: Models and Insights. *Manufacturing Service Operations Management*, 2011, vol. 13, pp. 534–548.
3. Kleinrock L. *Queueing Theory*. Translated from English. I.I. Grushko; V.I. Neiman (ed.). Moscow, Mechanical Engineering Publ., 1979, 432 p.
4. Lisovskaya E.Yu., Moiseeva S.P. Study of the process the number of customers in the system MjGIjNj1. *Probability theory, stochastic processes, mathematical statistics and applications: proceedings of the International scientic conference devoted to the 80th anniversary of professor, doctor of physical and mathematical sciences Gennady Alekseevich Medvedev*. Minsk, 2015, pp. 123–127.
5. Nazarov A., Terpugov A.F. *Queueing Theory*: textbook. Tomsk, Publishing house of the NTL, 2010, 228 p.