

Анализ характеристик замкнутой системы массового обслуживания с относительными приоритетами

03, март 2014

DOI: 10.7463/0314.0702664

Нестеров Ю. Г.

УДК 519.872

Россия, МГТУ им. Н.Э. Баумана

ugn@bmstu.ru

Введение

Замкнутые системы массового обслуживания (СМО) широко используются в качестве моделей для оценки временных характеристик информационных систем, сетей передачи данных, а также процессов массового обслуживания в производственных, транспортных, торговых, логистических и сервисных системах [1- 4]. В работе [5] аналитические решения для средних значений времен пребывания в узлах замкнутых СМО типа «модели ремонтника» (в обозначениях Кендалла [3] - $Mr|G|Ir|m||N$), используются в качестве базовых моделей для расчета средних значений времен пребывания заявок в узлах широкого класса замкнутых сетей массового обслуживания с приоритетами и консервативными дисциплинами обслуживания. На сегодня известны аналитические результаты для марковских СМО типа $Mr|Mr|m||N$ [2, 4], а также решения для частных случаев СМО такой структуры с одним обслуживающим аппаратом (ОА) [11-14]. Для СМО с одним ОА и произвольной функцией распределения вероятностей (ФРВ) времени обслуживания в работе [8] предложен общий подход к анализу подобных систем, однако переход к собственно распределению вероятностей состояний и от них к средним значениям времен пребывания связан со значительными вычислительными трудностями.

Настоящая работа посвящена получению эффективного в вычислительном отношении аналитического решения для распределения вероятностей состояний и средних значений времен пребывания и загрузок ОА для заявок различных классов в СМО типа «модели ремонтника» с приоритетами $Mr|G|Ir|1||N$.

Решение основано на построении и анализе вложенной цепи Маркова [6], переходе от распределения вероятностей состояний этой цепи к распределению стационарных состояний исходной СМО, а также доказательстве возможности такого перехода.

Постановка задачи.

Рассмотрим замкнутую СМО типа $Mr|G|Ir|1||N$, состоящую из одного обслуживающего аппарата, очереди перед ним и R независимых источников заявок, каждый емкостью N_r ($r=1,R$). ФРВ времени пребывания любой заявки класса “ r ” в источнике имеет экспоненциальное

распределение с параметром $\lambda_r, r = 1, \dots, R$, ФРВ времени обслуживания любой заявки класса "r" в ОА имеет произвольный вид $G\Gamma$. Поступающие из источника в очередь заявки обслуживаются в ОА в соответствии с правилом $d \in D$, указывающим для каждого состояния СМО класс заявки, которая должна быть выбрана на обслуживание из очереди в момент освобождения ОА. В пределах одного класса (если $N_r > 1$) заявки обслуживаются в порядке поступления. После завершения обслуживания заявка возвращается в источник. Таким образом, D включает в себя множество дисциплин (в том числе дисциплину с относительными приоритетами), не допускающими прерываний, в которых решение о том, какая заявка из очереди будет обслуживаться, принимается на основании лишь текущего состояния СМО, без учета предыстории обслуживания. Последовательные времена обслуживания заявок класса "r" суть независимые случайные величины (СВ) с произвольной ФРВ - $B_r(t), r = 1, \dots, R$. Такая СМО используется в качестве модели для оценки временных характеристик таких ресурсов, как каналы связи, контроллеры (шины) и процессоры ввода-вывода, программные процессы, работающие без прерываний.

Целью настоящего исследования является получение аналитических выражений для вычисления ФРВ состояний СМО, средних времен пребывания (ожидания) заявок каждого класса - $v_r(w_r)$ и загрузок ОА - $u_r, r = 1, \dots, R$.

Решение задачи.

Поскольку в рассматриваемой СМО в каждый момент времени имеется лишь одна СВ, имеющая неэкспоненциальное распределение, то, согласно [6, с.47], для анализа ее характеристик можно использовать метод вложенных цепей Маркова.

Состояние СМО представим в виде $(R+1)$ -мерного вектора с целочисленными компонентами:

$K = (j^K, K_1, K_2, \dots, K_{j^K}, \dots, K_R)$ - состояние K ;

$L = (j^L, L_1, L_2, \dots, L_{j^L}, \dots, L_R)$ - состояние L ;

$\emptyset = (0, 0_1, 0_2, \dots, 0_j, \dots, 0_R)$ - «нулевое» состояние (все заявки в источнике).

Здесь:

$K, L, \emptyset \in S^*$, где S^* - конечное множество состояний СМО, содержащее $M = \prod_{r=1}^R (N_r + 1)$ элементов;

$j^K(j^L)$ - номер класса заявки, обслуживаемой в ОА в состоянии $K(L)$, $j^K, j^L \in [1, R]$.

$K_i(L_i) = \begin{cases} 0, 1, \dots, N_i, & \text{если } i \neq j^K(j^L) \\ 1, 2, \dots, N_i, & \text{если } i = j^K(j^L) \end{cases}$ - количество заявок класса "i" в СМО в состоянии $K(L)$.

$d = \{j^K, K \in S^*\}$ - $(M-R)$ -мерный вектор, определяющий дисциплину обслуживания заявок в СМО.

Рассмотрим последовательность $\{t_n + 0\}$, $n = 1, 2, \dots$ моментов времени завершения обслуживания в ОА очередной заявки (сразу после ее ухода в источник), представляющую собой последовательность точек регенерации [8, с.13]. Тогда последовательность Z состояний СМО в эти моменты времени $Z = \{K(t_n + 0)\}$, $n = 1, 2, \dots$ связана в однородную дискретную цепь Маркова с конечным числом состояний S : $S = (S^* \setminus N)$, где $N = \{(1, N_1, N_2, \dots, N_R), (2, N_1, N_2, \dots, N_R), \dots, (R, N_1, N_2, \dots, N_R)\}$ есть множество состояний, когда все заявки находятся в СМО (в источнике пусто).

Процесс обслуживания заявок в СМО полностью описывается вложенной цепью Маркова Z и ФРВ времен перехода СМО из состояния " K " в состояние " L " ($K, L \in S$).

Анализ вложенной цепи Маркова Z .

Определим элементы матрицы вероятностей переходов $P = [p_{KL}]_{M \times M}$ цепи Z . Рассмотрим два случая (I - $K \neq \emptyset$, II - $K = \emptyset$), отличающихся тем, что при $K \neq \emptyset$ переход в состояние L происходит за время обслуживания заявки класса j^K , а при $K = \emptyset$ переход в состояние L происходит за время, когда СМО пуста и, затем, в течение времени обслуживания заявки, пришедшей в пустую СМО первой.

Случай I. $K \neq \emptyset$. Переход $K \rightarrow L$ есть событие, состоящее в том, что за время обслуживания заявки класса j^K в СМО придут заявки, которых нет в состоянии K , но которые есть в состоянии L и не придут заявки, которых нет в состоянии L . Поскольку процессы поступления заявок каждого класса независимы друг от друга, можно записать:

$$p_{KL} = \prod_{i=1}^R p(A_i), \quad (1)$$

Где A_i ($i = 1, \dots, R$) есть событие, состоящее в следующем:

$$A_i = \begin{cases} \text{за время обслуживания заявки класса } j^K \text{ придет ровно } (L_i - K_i) \text{ заявок класса } i, (K_i \leq L_i) \wedge (i \neq j^K) \\ (K_i > L_i) \wedge (i \neq j^K) - \text{невозможное событие} \\ \text{за время обслуживания заявки класса } j^K \text{ придет ровно } (L_i - K_i + 1) \text{ заявок класса } i, ((K_i - 1) \leq L_i) \wedge (i = j^K) \\ (K_i - 1 > L_i) \wedge (i = j^K) - \text{невозможное событие} \end{cases}$$

Рассмотрим ситуацию $(K_i \leq L_i) \wedge (i \neq j^K)$:

$$p(A_i) = \int_0^\infty p(B_i^t) dB_{j^K}(t), \quad (2)$$

где B_i^t - событие, состоящее в том, что к моменту времени t ($t > 0$) придет ровно $(L_i - K_i)$ заявок класса i . Событие B_i^t есть сумма $\binom{N_i - K_i}{L_i - K_i}$ несовместных событий, каждое из которых, в свою очередь, заключается в том, что к моменту времени t $(L_i - K_i)$ каких-либо заявок из числа $(N_i - K_i)$ заявок, присутствующих в источнике класса i , завершит свое пребывание в источнике, а остальные $(N_i - K_i) - (L_i - K_i) = (N_i - L_i)$ заявок класса i не покинут источник.

$$\begin{aligned}
p(B_i^t) &= \binom{N_i - K_i}{L_i - K_i} (1 - e^{-\lambda_i t})^{(L_i - K_i)} \times (e^{-\lambda_i t})^{(N_i - L_i)} = \binom{N_i - K_i}{L_i - K_i} \sum_{m=0}^{L_i - K_i} (-e^{-\lambda_i t})^m \times e^{-(N_i - L_i)\lambda_i t} = \\
&= \binom{N_i - K_i}{L_i - K_i} \sum_{m=0}^{L_i - K_i} \binom{L_i - K_i}{m} (-1)^m e^{-(N_i - L_i + m)\lambda_i t}
\end{aligned} \quad (3)$$

Здесь $\binom{n}{k} = C_n^k$ - число сочетаний из n элементов по k . При выводе (3) использовалась формула бинома Ньютона. Подставляя (3) в (2), получим:

$$\begin{aligned}
p(A_i) &= \int_0^\infty \binom{N_i - K_i}{L_i - K_i} \sum_{m=0}^{L_i - K_i} \binom{L_i - K_i}{m} (-1)^m e^{-(N_i - L_i + m)\lambda_i t} dB_{j^K}(t) = \\
&= \binom{N_i - K_i}{L_i - K_i} \sum_{m=0}^{L_i - K_i} \binom{L_i - K_i}{m} (-1)^m B_{j^K}^* [(N_i - L_i + m)\lambda_i], \quad i \neq j^K.
\end{aligned} \quad (4)$$

В (4) $B_{j^K}^*(.)$ суть преобразование Лапласа-Стильтьеса ФРВ $B_{j^K}(t)$.

Ситуация $((K_i - 1) \leq L_i) \wedge (i = j^K)$ отличается от рассмотренного тем, что заявка класса j^K , обслуживаемая в ОА в состоянии K , не может находиться в СМО в состоянии L . Рассуждая аналогично, получим:

$$p(A_i) = \binom{N_i - K_i + 1}{L_i - K_i + 1} \sum_{m=0}^{L_i - K_i + 1} \binom{L_i - K_i + 1}{m} (-1)^m B_{j^K}^* [(N_i - L_i + m - 1)\lambda_i], \quad i = j^K \quad (5)$$

В остальных случаях $p(A_i) = 0$.

Случай II. $K = \emptyset$ Переход $\emptyset \rightarrow L$ есть сумма R несовместных событий, каждое из которых состоит в том, что в пустую СМО первой придет заявка класса i , т.е. СМО перейдет в состояние $(i, 0_1, 0_2, \dots, 1_i, \dots, 0_R)$ - событие C_i , а затем произойдет переход из этого промежуточного состояния в состояние L

$$P_{\emptyset L} = \sum P(C_i) P_{(i, 0_1, 0_2, \dots, 1_i, \dots, 0_R) L} \quad (6)$$

В свою очередь вероятность события C_i очевидно равна:

$$P(C_i) = N_i \lambda_i / \Lambda, \quad \text{здесь} \quad \Lambda = \sum_{r=1}^R N_r \lambda_r \quad (7)$$

Итак, для случая $\emptyset \rightarrow L$ имеем:

$$P_{\emptyset L} = \Lambda^{-1} \sum_{i=1}^R N_i \lambda_i P_{(i, 0_1, 0_2, \dots, 1_i, \dots, 0_R) L} \quad (8)$$

Здесь $P_{(i,0_1,0_2,\dots,1_i,\dots,0_R)L}$ - частный случай (1).

Окончательно матрица переходных вероятностей P для вложенной цепи Маркова Z определяется следующими соотношениями:

$$\left\{ \begin{array}{l} K \neq \emptyset \quad p_{KL} = \prod_{i=1}^R P(A_i), \text{ где} \\ P(A_i) = \binom{N_i - K_i}{L_i - K_i} \sum_{m=0}^{L_i - K_i} \binom{L_i - K_i}{m} (-1)^m B_{j^K}^* [(N_i - L_i + m)\lambda_i], \text{ при } (K_i \leq L_i) \wedge (i \neq j^K) \\ P(A_i) = \binom{N_i - K_i + 1}{L_i - K_i + 1} \sum_{m=0}^{L_i - K_i + 1} \binom{L_i - K_i + 1}{m} (-1)^m B_{j^K}^* [(N_i - L_i + m - 1)\lambda_i], \text{ при } ((K_i - 1) \leq L_i) \wedge (i = j^K) \\ P(A_i) = 0 \quad \text{- в остальных случаях} \end{array} \right\} \quad (9)$$

$$\left\{ K = \emptyset \quad P_{\emptyset L} = \left(\sum_{r=1}^R N_r \lambda_r \right)^{-1} \sum_{i=1}^R N_i \lambda_i P_{(i,0_1,0_2,\dots,1_i,\dots,0_R)L} \text{ где } P_{(i,0_1,0_2,\dots,1_i,\dots,0_R)L} \text{ - частный случай (1)} \right\} \quad (10)$$

Рассмотренная нами однородная конечная цепь Маркова неприводима и апериодична, следовательно, в силу эргодической теоремы Маркова, эргодична, т.е. существует предел [6]:

$$\lim_{n \rightarrow \infty} \Pr \{ K(t_{n+1} + 0) = L \mid K(t_n + 0) = K \} = \pi_L, \quad L \in S \quad (11)$$

Стационарное распределение вероятностей состояний цепи Z есть решение системы из M линейных алгебраических уравнений $\pi = \pi P$, в которой любое из уравнений заменяется условием нормировки $\sum_{L \in S} \pi_L = 1$. Здесь $\pi = \{\pi_L\}$, $L \in S$. Размерность системы уравнений быстро возрастает с ростом R . Так при $N_r = 1$ и $R=10$, $M = 2^{10} - 10 = 1014$.

Теперь определим ФРВ времени перехода $K \rightarrow L$.

При $K \neq \emptyset$ ФРВ времени перехода $K \rightarrow L$ $\Phi_{KL}(t)$ суть ФРВ времени обслуживания заявки класса $j^K B_{j^K}(t)$, т.е.:

$$\Phi_{KL}(t) = B_{j^K}(t), \quad K, L \in S \quad (12)$$

Рассмотрим случай $K = \emptyset$. Переход $\emptyset \rightarrow L$ есть сумма R несовместных событий, состоящих в том, что первой пришла заявка класса i , а затем, за время ее обслуживания, пришли заявки, которые есть в состоянии L . Длительность перехода $\emptyset \rightarrow L$ складывается из отрезка времени, когда СМО пуста, и времени обслуживания заявки, пришедшей первой. Заметим, что переход $\emptyset \rightarrow L$ происходит только тогда, когда первой в «пустую» СМО приходит заявка, отсутствующая в состоянии L . Следовательно, интенсивность потока заявок, вызывающих переход $\emptyset \rightarrow L$ есть $\Lambda^* = \sum_{i=1}^R (N_i - L_i) \lambda_i$, а вероятность того, что этот переход вызовет заявка класса i равна $(N_i - L_i) \lambda_i / \Lambda^*$.

Обозначим ПРВ времени перехода $\emptyset \rightarrow L$ через $\varphi_{\emptyset L}(t) = d\Phi_{\emptyset L}(t)/dt$. Тогда имеем:

$$\varphi_{\emptyset L}(t) = d\Phi_{\emptyset L}(t)/dt = (\Lambda^*)^{-1} \sum_{i=1}^R (N_i - L_i) \lambda_i \varphi_i(t), \quad (13)$$

$$\text{где: } \varphi_i(t) = \int_0^\infty [N_i \lambda_i e^{-N_i \lambda_i (t-x)}] dB_i(x) = N_i \lambda_i e^{-N_i \lambda_i t} B_i^*(-N_i \lambda_i) \quad (14)$$

есть свертка ПРВ интервала времени от момента перехода СМО в состояние $\emptyset$ и до момента прихода заявки класса i и ПРВ времени обслуживания этой заявки $dB_i(t)/dt$. Процесс обслуживания заявок в данной СМО полностью описывается вложенной цепью Маркова Z , соотношения (9) и (10), и ПРВ времен переходов $K \rightarrow L$, соотношения (12 - 14).

Поскольку стационарные вероятности состояний исходной СМО $P = \{P_K\}$, $K \in S$ отличаются от распределения $\pi = \{\pi_K\}$, $K \in S$ вложенной цепи Маркова, возникает задача перехода от распределения π к распределению P .

Определение стационарных вероятностей состояний исходной СМО.

Рассмотрим поведение исходного процесса $K(t) = (j^K, K_1, K_2, \dots, K_R)$ между моментами $\{t_n\}$, $n=1, 2, \dots$ выхода заявок из ОА. Нетрудно видеть, что на интервале $(t_{n-1} + 0, t_n - 0)$ поведение исходного процесса $K(t)$ таково, что $j^K(t) = \text{const}$, а компоненты $K_r(t)$, $r=1, \dots, R$ могут только увеличиваться. Причем, в силу того, что время пребывания заявок в источниках имеет экспоненциальное распределение, за $\Delta t \rightarrow 0$ лишь одна из компонент $K_r(t)$, $r=1, \dots, R$ может увеличиться на 1 (процесс $K(t) = (j^K, K_1, K_2, \dots, K_R)$ - перейти в одно из смежных состояний). Под смежным к состоянию K состоянием понимается состояние L , в котором одна из компонент L_r , $r=1, \dots, R$ больше соответствующей компоненты K_r , $r=1, \dots, R$ на 1, а все остальные - совпадают. Таким образом, можно провести аналогию с процессом чистого размножения (интенсивности потоков гибели равны нулю [7]). При этом, если в момент $(t_n + 0)$ СМО находилась в состоянии K , то интенсивность перехода ν_K исходного процесса $K(t)$ в смежные состояния будет определяться формулой:

$$\nu_K = \sum_{r=1}^R (N_r - K_r) \lambda_r \quad (15)$$

Обозначим интенсивность выхода из ОА обслуженных заявок в стационарном состоянии через η , а стационарную вероятность того, что СМО находится в состоянии K через P_K , $K \in S$. В соответствии с результатами работы [9] можно записать:

$$\nu_K P_K = \eta \pi_K, \quad K \in S \quad (16)$$

Выражение (16) представляет собой систему уравнений глобального баланса для состояний замкнутой СМО. Смысл уравнений состоит в том, что в стационарном режиме каждое из состояний K , $K \in S$ СМО является равновесным: интенсивности переходов в состояние K из состояния K совпадают.

Из (15) и (16) следует, что для определения вероятностей $P = \{P_K\}$, $K \in S$ достаточно определить значение η . ПРВ интервала времени (t_{n-1}, t_n) при $n \rightarrow \infty$ обозначим $\eta(t)$. Очевидно, что:

$$\eta(t) = \sum_{K \in S} (\pi_K \sum_{L \in S} \varphi_{KL}(t) P_{KL}) \quad (17)$$

Здесь $\sum_{L \in S} \varphi_{KL}(t) P_{KL}$ суть ПРВ времени пребывания СМО в состоянии K , а $\varphi_{KL}(t)$ суть ПРВ времени перехода $K \rightarrow L$, определяемая соотношениями (12-14), а P_{KL} , $K, L \in S$ - переходные вероятности вложенной цепи Маркова Z , определяемые соотношениями (9,10).

Для η имеем:

$$\eta = \left[\int_0^\infty t \eta(t) dt \right]^{-1} \quad (18)$$

Из (16) имеем:

$$P_K = \eta \pi_K / \nu_K, \quad K \in S \quad (19)$$

Таким образом, искомая ФРВ $P = \{P_K\}$, $K \in S$ определяется соотношением (19) и последовательным применением выражений (9-18) для определения входящих в него π_K , ν_K , $K \in S$ и η соответственно. Вероятность P_N того, что в СМО находится ровно N_r , $r = 1, \dots, R$ заявок каждого класса (в источниках пусто) определяется из условия нормировки:

$$P_N = 1 - \sum_{K \in S} P_K \quad (20)$$

Определение математических ожиданий времен пребывания (ожидания) и загрузок ОА для заявок каждого класса в СМО.

Обозначим математическое ожидание времени пребывания (ожидания) заявки класса r в СМО через $\nu_r(w_r)$, а загрузку ОА - u_r . Поскольку в рассматриваемой СМО используется консервативная дисциплина обслуживания (сохраняется выполненная работа), то для данной системы справедлив результат Литтла [10]:

$$m_r = \lambda_r^* \nu_r, \quad r = 1, \dots, R \quad (21)$$

Здесь: m_r - среднее число заявок класса r в СМО; λ_r^* - фактическая интенсивность поступления заявок класса r из источника в СМО; ν_r - среднее время пребывания заявок класса r в СМО.

Нетрудно видеть, что $\lambda_r^* = \lambda_r (N_r - m_r)$, $r = 1, \dots, R$, следовательно, из (21) получим:

$$\nu_r = \frac{m_r}{\lambda_r (N_r - m_r)}, \quad r = 1, \dots, R \quad (22)$$

Определим для всех $\mathbf{r} (\mathbf{r}=1,...,R)$ и для всех $K_r (K_r = 0,1,...,N_r)$ подмножества состояний $S_{K_r}^r \subset S^*$ следующим образом:

$$S_{K_r}^r = \{L : L_r = K_r\}, L \in S^*, K_r \in [0,1,...,N_r], r \in [1,R], \quad (23)$$

Т.е. $S_{K_r}^r$ есть подмножество таких состояний СМО, в которых число заявок класса r равно K_r .

Вероятность $\bar{P}_{K_r}$ того, что в СМО находится K_r заявок класса $\mathbf{r}$ определяется соотношением:

$$\bar{P}_{K_r} = \sum_{L \in S_{K_r}^r} P_L, \quad (24)$$

а среднее число заявок класса $\mathbf{r}$ в СМО -

$$m_r = \sum_{K_r=0}^{N_r} K_r \bar{P}_{K_r}. \quad (25)$$

Подставляя (24) в (25) и (25) в (22), окончательно получим для среднего времени пребывания заявок в СМО:

$$v_r = \left[\sum_{K_r=0}^{N_r} (K_r \sum_{L \in S_{K_r}^r} P_L) \right] / \left[\lambda_r (N_r - \sum_{K_r=0}^{N_r} (K_r \sum_{L \in S_{K_r}^r} P_L) \right], r \in [1,R] \quad (26)$$

Для среднего времени ожидания W_r имеем очевидное соотношение:

$$w_r = v_r - b_r, r = 1,...,R \quad (27)$$

Загрузка ОА u_r заявками класса $\mathbf{r}$ и суммарная нагрузка ОА u определяются следующими формулами:

$$u_r = \lambda_r^* b_r = \lambda_r \left[N_r - \sum_{K_r=0}^{N_r} (K_r \sum_{L \in S_{K_r}^r} P_L) \right] b_r \quad (28)$$

$$u = \sum_{r=1}^R u_r \quad (29)$$

Заключение.

Доказано, что получение искомым аналитических выражений для вычисления ФРВ состояний СМО, средних времен пребывания (ожидания) заявок каждого класса - $v_r(w_r)$ и нагрузок ОА - $u_r, r=1,...,R$. для замкнутой СМО типа $Mr|G|1||N$ с относительными приоритетами сводится к следующей последовательности относительно простых шагов .

1. Построение вложенной цепи Маркова, связанной с моментами завершения обслуживания заявок в ОА (точками регенерации), определение матрицы переходных вероятностей и вероятностей состояний этой цепи - соотношения (1-14).
2. Определение распределения вероятностей состояний СМО в стационарном режиме на основании системы уравнений глобального баланса для состояний замкнутой СМО – соотношения (15- 20).
3. Определение аналитических выражений для средних времен пребывания, ожидания и загрузок ОА для заявок различных классов - соотношения (21–29).
4. Возможности полученной аналитической модели ограничивается возможностью обращения матрицы переходных вероятностей вложенной цепи Маркова.

Список литературы

1. Гнеденко Б.В., Коваленко И.Н. Введение в теорию массового обслуживания. М.: Изд-во ЛКИ, 2007. 400 с.
2. Вишневский В.М. Теоретические основы проектирования компьютерных сетей. М.: Техносфера, 2003. 512 с.
3. Клейнрок Л. Теория массового обслуживания: пер. с англ. М.: Машиностроение, 1979. 432 с.
4. Клейнрок Л. Вычислительные системы с очередями: пер. с англ. М.: Мир, 1979. 600 с.
5. Нестеров Ю.Г. Декомпозиционный метод анализа замкнутых сетей массового обслуживания // Наука и образование. МГТУ им. Н.Э. Баумана. Электрон. журн. 2014. № 2. Режим доступа: <http://technomag.bmstu.ru/doc/700018.html> (дата обращения 28.02.2014). DOI: [10.7463/0214.0700018](https://doi.org/10.7463/0214.0700018)
6. Кёниг Д., Штойян Д. Методы теории массового обслуживания: пер. с нем. М.: Радио и связь, 1981. 128 с.
7. Вентцель Е.С., Овчаров Л.А. Теория случайных процессов и ее инженерные приложения. М.: Высшая школа, 2000. 383 с.
8. Джейсуол Н. Очереди с приоритетами: пер. с англ. М.: Мир, 1973. 280 с.
9. König D., Rolfsky T., Smidt V., Stoyan D. Stochastic processes with imbedded marked point processes (PMP) and their application in queuing theory // Math. Operationsforschung und Statistik. Ser. Optimization. 1978. Vol. 9. P. 125-142.
10. Little J.D.C. A proof for the queueing formula $L = \lambda W$ // Operations Research. 1961. Vol. 9, no. 3. P. 383-387. DOI: [10.1287/opre.9.3.383](https://doi.org/10.1287/opre.9.3.383)
11. Szep A. Iterative Method for Solving M/G/1/N-type Loops with Priority Queues. Режим доступа: http://www.inf.u-szeged.hu/actacybernetica/edb/vol07n3/pdf/Szep_1986_ActaCybernetica.pdf (дата обращения 28.02.2014).

12. Fatnes J.N. Flow-times in an M/G/1 Queue under a Combined Preemptive/ Non-preemptive Priority Discipline: Master of Science in Physics and Mathematics. Norwegian University of Science and Technology, Department of Mathematical Sciences, 2010. Режим доступа: <http://www.diva-portal.org/smash/get/diva2:348953/FULLTEXT01.pdf> (дата обращения 28.02.2014).
13. Madan K.C. A Non-Preemptive Priority Queueing System with a Single Server Serving Two Queues M/G/1 and M/D/1 with Optional Server Vacations Based on Exhaustive Service of the Priority Units // Applied Mathematics. 2011. Vol. 2, no. 6. P. 791-799. DOI: [10.4236/am.2011.26106](https://doi.org/10.4236/am.2011.26106)
14. Atar R., Biswas A., Kaspi H. Fluid limits of G/G/1+G queues under the non-preemptive earliest-deadline-rst discipline. Preprint. Technion - Israel Institute of Technology, Haifa, Israel, 2014. 28 p. Режим доступа: <http://webee.technion.ac.il/people/atar/ata-bis-kas.pdf> (дата обращения 28.02.2014).

Feature analysis for closed queuing system with non-preemptive priorities

03, March 2014

DOI: 10.7463/0314.0702664

Yu.G. Nesterov

Bauman Moscow State Technical University, 105005, Moscow, Russian Federation

ugn@bmstu.ru

The object of research in this article is to obtain the analytical expressions for the probability distribution functions (PDF) of states of the closed system of mass service (SMS) like "model of the repairman" $M|G|1 || N$ in a stationary mode. The SMS has a source of requests of final capacity with exponentially distributed time of residence of different class requests in the source, one queue with discipline of requests service i.e. the non-preemptive priorities, one serving device (SD) and arbitrary PDF of a residence time of each class requests.

The method includes the following steps.

The first step, while servicing the non-uniform population of requests, distinguishes a sequence of time points, so-called points of regeneration in which a process behaves as Markov's. It is shown that such points are time points to complete a service of requests in SD. Further, for these points an embedded Markov's chain is designed, and the states space of this chain is defined. Analytical expressions to calculate the elements of transition probability matrix in such chain are obtained. These expressions reflect a relationship between these probabilities and the source parameters, the residence time PDF parameters, the number of each class requests, and a discipline of their service. Further, the solution for PDF of this chain states is found.

The second step establishes connection between the PDF of SMS states in a stationary mode and the previously received PDF of states of the embedded Markov's chain. Such connection is defined by a system of the equations of global balance for the SMS states in a stationary mode and states of the embedded Markov's chain. The physical sense of such balance is that each of SMS state in a stationary mode is in equilibrium, i.e. the intensities of transitions from the state and in the state are equal. The space of states of initial SMS in a stationary mode is defined. It is shown that in intervals of request service time in SD the process behaves as a process of net reproduction when intensity of death – transition of SMS to

"lower" states – is equal to zero. The paper presents the obtained analytical expressions to calculate the stationary PDF of given SMS depending on the PDF of the embedded Markov' chain, the discipline of service, and the PDF of residence times to serve different class requests in SD.

The third step provides expressions to calculate the average values of residence and waiting times of each class requests in SMS, as well as expressions to define the SD loading by requests of each class and the general loading of SD. The formulas development is based on conservatism (work preservation), fair for discipline of service with non-preemptive priorities, and also using Little's result and expressions for average number of requests of each class in SMS which, in turn, are defined by the PDF of SMS states in a stationary mode.

Publications with keywords: [queue](#), [facility](#), [service time](#), [waiting time](#), [residence time](#), [closed queuing system](#), [service discipline](#), [utilization](#), [transition matrix](#), [embedded Markov chain](#)

Publications with words: [queue](#), [facility](#), [service time](#), [waiting time](#), [residence time](#), [closed queuing system](#), [service discipline](#), [utilization](#), [transition matrix](#), [embedded Markov chain](#)

References

1. Gnedenko B.V., Kovalenko I.N. *Vvedenie v teoriyu massovogo obsluzhivaniya* [Introduction to queuing theory]. Moscow, LKI Publ., 2007. 400 p. (in Russian).
2. Vishnevskiy V.M. *Teoreticheskie osnovy proektirovaniya komp'yuternykh setey* [Theoretical bases of designing of computer networks]. Moscow, Tekhnosfera, 2003. 512 p. (in Russian).
3. Kleinrock L. *Queueing Systems. Vol. 1. Theory*. New York, John Wiley and Sons, 1975. (Russ. ed.: Kleinrock L. *Teoriia massovogo obsluzhivaniia*. Moscow, Mashinostroenie, 1979. 432 p.).
4. Kleinrock L. *Queueing Systems. Vol. 2. Computer Applications*. New York, John Wiley and Sons, 1976. (Kleinrock L. *Vychislitel'nye sistemy s ocheredyami*. Moscow, Mir, 1979. 600 p.).
5. Nesterov Yu.G. [Decomposition method for analysis of closed queuing networks]. *Nauka i obrazovanie MGTU im. N.E. Bauman - Science and Education of the Bauman MSTU*, 2014, no. 2. DOI: [10.7463/0214.0700018](https://doi.org/10.7463/0214.0700018) (in Russian).
6. Konig D., Stoyan D. *Metody teorii massovogo obsluzhivaniya* [Methods of queuing theory]. Transl. from German. Moscow, Radio i svyaz', 1981. 128 p.
7. Venttsel' E.S., Ovcharov L.A. *Teoriya sluchaynykh protsessov i ee inzhenernye prilozheniya* [Theory of random processes and its engineering applications]. Moscow, Vysshaya shkola, 2000. 383 p. (in Russian).
8. Jaiswal N.K. *Priority Queues*. Academic Press, New York, 1968. (Russ. ed.: Jaiswal N. *Ocheredi s prioritetaami*. Moscow, Mir, 1973. 280 p.).

9. König D., Rolsky T., Smidt V., Stoyan D. Stochastic processes with imbedded marked point processes (PMP) and their application in queueing theory. *Math. Operationsforschung und Statistik. Ser. Optimization*, 1978, vol. 9, pp. 125-142.
10. Little J.D.C. A proof for the queueing formula $L = \lambda W$. *Operations Research*, 1961, vol. 9, no. 3, pp. 383-387. DOI: [10.1287/opre.9.3.383](https://doi.org/10.1287/opre.9.3.383)
11. Szep A. *Iterative Method for Solving M/G/I/N-type Loops with Priority Queues*. Available at: http://www.inf.u-szeged.hu/actacybernetica/edb/vol07n3/pdf/Szep_1986_ActaCybernetica.pdf , accessed 28.02.2014.
12. Fatnes J.N. *Flow-times in an M/G/1 Queue under a Combined Preemptive/ Non-preemptive Priority Discipline*. Master of Science in Physics and Mathematics. Norwegian University of Science and Technology, Department of Mathematical Sciences, 2010. Available at: <http://www.diva-portal.org/smash/get/diva2:348953/FULLTEXT01.pdf> , accessed 28.02.2014.
13. Madan K.C. A Non-Preemptive Priority Queueing System with a Single Server Serving Two Queues M/G/1 and M/D/1 with Optional Server Vacations Based on Exhaustive Service of the Priority Units. *Applied Mathematics*, 2011, vol. 2, no. 6, pp. 791-799. DOI: [10.4236/am.2011.26106](https://doi.org/10.4236/am.2011.26106)
14. Atar R., Biswas A., Kaspi H. *Fluid limits of G/G/1+G queues under the non-preemptive earliest-deadline-rst discipline*. Preprint. Technion - Israel Institute of Technology, Haifa, Israel, 2014. 28 p. Available at: <http://webee.technion.ac.il/people/atar/ata-bis-kas.pdf> , accessed 28.02.2014.