Автоматика и телемеханика, № 9, 2024

Стохастические системы

© 2024 г. **Ю.В. МАЛИНКОВСКИЙ, д-р физ.-мат. наук (malinkovsky@gsu.by),**
С.Ю. ЕВМЕНЕНКО (stas.evmenenko@yandex.ru)
(Гомельский государственный университет им. Ф. Скорины)

ИНВАРИАНТНОСТЬ СТАЦИОНАРНОГО РАСПРЕДЕЛЕНИЯ ОТКРЫТОЙ СЕТИ ОБСЛУЖИВАНИЯ С ЭКСПОНЕНЦИАЛЬНЫМ ОГРАНИЧЕНИЕМ НА ВРЕМЯ ПРЕБЫВАНИЯ

Рассматривается открытая сеть обслуживания с однолинейными узлами и экспоненциальным ограничением на время пребывания запросов в узлах. Запросы, время пребывания которых в узле закончилось, мгновенно и независимо от других запросов начинают перемещаться по матрице, отличной от матрицы маршрутизации обслуженных запросов. В сеть поступает простейший поток запросов. Устанавливается инвариантность стационарного распределения по отношению к функциональной форме распределений длительностей обслуживания при фиксированных первых моментах.

Ключевые слова: теория вероятностей, теория массового обслуживания, условие эргодичности, открытая сеть массового обслуживания, сеть с экспоненциальным ограничением на время пребывания запросов в узлах, инвариантность стационарного распределения.

DOI: 10.31857/S0005231024090057, **EDN:** ZQRYHX

1. Введение

Важную роль при исследовании сетей массового обслуживания играет доказательство инвариантности стационарного распределения вероятностей состояний по отношению к форме распределения длительностей обслуживания запросов в узлах. Это связано с тем, что на практике распределение продолжительности обслуживания обычно отличается от показательного.

В ряде работ, например в [1, 2], инвариантность установлена в предположении, что распределения длительностей обслуживания имеют рациональные преобразования Лапласа–Стилтьеса. В [3] была доказана инвариатность стационарного распределения для открытых и замкнутых сетей обслуживания с обходами узлов запросами. Результаты по инвариантности стационарного распределения сетей при более общих предположениях получены, например, в [4–7]. В [8, 9] показаны примеры применения аппарата сетей обслуживания к современным инфокоммуникационным системам и сетям. В [10, 11] была обоснована необходимость исследования сетей обслуживания, в которых запрос может покинуть узел, не дожидаясь окончания обслуживания. Однако в указанных работах считалось, что продолжительности обслуживания запросов имеют показательное распределение.

В настоящей работе рассматривается сеть обслуживания с экспоненциальным ограничением на время пребывания запросов в узлах, в которой продолжительности обслуживания имеют произвольный закон распределения, а дисциплина обслуживания запросов в узлах имеет следующую специфику. Запрос, поступающий в некоторый узел, обладает абсолютным приоритетом по отношению ко всем остальным запросам в узле, а вытесненный с прибора запрос при повторном его попадании на прибор дообслуживается оставшееся время. Доказывается, что стационарное распределение вероятностей состояний указанных сетей не зависит от функционального вида законов распределения длительностей обслуживания запросов в узлах при фиксированных первых моментах.

2. Марковский случай

В сеть массового обслуживания, состоящую из N однолинейных узлов (систем), поступает стационарный пуассоновский поток запросов с интенсивностью λ. Каждый поступающий запрос независимо от других запросов с вероятностью p_{0i} направляется в i-й узел ($i = \overline{1, N}, \sum_{i=1}^{N} p_{0i} = 1$). Число мест для ожидания в узле неограниченно.

Состояние сети будем описывать вектором $\mathbf{n}(t) = (n_1(t), \dots, n_N(t))$, где $n_i(t)$ – число запросов в i-м узле в момент времени t.

Время пребывания запроса в i-м узле является случайной величиной, условное распределение которой (если в i-м узле находится n_i запросов) показательное с параметром $\frac{\nu_i}{n_i}$. Другими словами, условная вероятность того, что пребывание каждого запроса в i-м узле закончится в промежутке $[t, t+h)$, если в момент t в узле находилось n_i запросов, равна $\frac{\nu_i}{n_i} h + o(h)$ при $h \to 0$, а условная вероятность завершения пребывания хотя бы одного из этих запросов равна $\nu_i h + o(h)$. Запрос, обслуженный в i-м узле, мгновенно и независимо от других запросов с вероятностью p_{ij} направляется в j-й узел сети, а с вероятностью p_{i0} покидает сеть ($\sum_{j=0}^{N} p_{ij} = 1$). Запрос, время пребывания которого в i-м узле закончилось, мгновенно и независимо от других запросов с вероятностью r_{ij} направляется в j-й узел сети, а с вероятностью r_{i0} покидает сеть ($\sum_{j=0}^{N} r_{ij} = 1$). Для удобства введем узел 0, который отождествим с внешностью сети. Введем также следующие две стохастические квадратные матрицы порядка $N+1$:

$$P = \begin{bmatrix} 0 & p_{01} & p_{02} & \cdots & p_{0N} \\ p_{10} & p_{11} & p_{12} & \cdots & p_{1N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{N0} & p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}, \qquad R = \begin{bmatrix} 0 & p_{01} & p_{02} & \cdots & p_{0N} \\ r_{10} & r_{11} & r_{12} & \cdots & r_{1N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{N0} & r_{N1} & r_{N2} & \cdots & r_{NN} \end{bmatrix},$$

которые назовем матрицами маршрутизации соответственно обслуженных и «потерянных» запросов. Итак, $p_{00} = r_{00} = 0$, $r_{0j} = p_{0j}$ для остальных j. Очевидно, матрица $S = (s_{ij}, i, j = \overline{0, N})$, где для $i \neq 0$

$$s_{ij} = \frac{\mu_i p_{ij} + \nu_i r_{ij}}{\mu_i + \nu_i} = \frac{\mu_i}{\mu_i + \nu_i} p_{ij} + \frac{\nu_i}{\mu_i + \nu_i} r_{ij},$$

а $s_{0j} = p_{oj}$, также является стохастической и по смыслу управляет движением запросов по узлам $0, 1, \ldots, N$ без учета того, за счет чего (окончания обслуживания или окончания времени пребывания) запрос покидает узел. Будем называть эту матрицу матрицей маршрутизации.

Будем предполагать, что матрица s_{ij} неприводима. Уравнение

$$\varepsilon_j = p_{0j} + \sum_{i=1}^{N} \varepsilon_i s_{ij} \quad (j = \overline{1, N}) \tag{1}$$

будем называть уравнением трафика. И в силу неприводимости матрицы s_{ij} уравнение (1) будет иметь единственное решение (ε_j), для которого $\varepsilon_j > 0$, $(j = \overline{1, N})$.

Пусть длительности обслуживания запросов в узлах имеют показательное распределение, т.е. для i-го узла

$$B_i(t) = 1 - \exp\left[-\mu_i t\right] \quad (t > 0).$$

Тогда $\mathbf{n}(t)$ – однородный марковский процесс с непрерывным временем и не более чем счетным фазовым пространством состояний.

В [10] для системы с ограниченным временем пребывания установлено, что при выполнении условия $\frac{\lambda\varepsilon_i}{\mu_i + \nu_i} < 1$, $i = \overline{1, N}$ цепь Маркова, описывающая число запросов в сети в момент времени t, эргодична, ее стационарное распределение имеет вид $p(\mathbf{n}) = p_1(n_1) \ldots p_N(n_N)$ с множителями

$$p_i(n_i) = \left(\frac{\lambda\varepsilon_i}{\mu_i + \nu_i}\right)^{n_i} \left(1 - \frac{\lambda\varepsilon_i}{\mu_i + \nu_i}\right),$$

а уравнение глобального равновесия имеет вид

$$\begin{aligned} p(\mathbf{n}) \left(\lambda + \sum_{i=1}^{N} (\mu_i(1 - p_{ii}) + \nu_i(1 - r_{ii})) I_{n_i \neq 0}\right) = \\ = \sum_{i=1}^{N} p(\mathbf{n} - e_i)\lambda p_{0i} + \sum_{i=1}^{N} p(\mathbf{n} + e_i)(\mu_i p_{i0} + \nu_i r_{i0}) + \\ + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} p(\mathbf{n} + e_j - e_i)(\mu_j p_{ji} + \nu_j r_{ji}). \end{aligned} \tag{2}$$

3. Немарковский случай

Будем предполагать, что времена обслуживания запросов в узлах не зависят от процесса поступления и друг от друга и имеют произвольную функцию распределения $B_i(t)$,

$$(\mu_i)^{-1} = \int\limits_0^\infty [1 - B_i(t)]dt, \tag{3}$$

где μ_i – скорость обслуживания запросов занятым i-м узлом. Если запрос поступает в узел, свободный от запросов, он сразу начинает обслуживаться. Если запрос поступает в узел, в котором уже есть запрос, то он выбивает запрос, находящийся на приборе, и сразу же начинает обслуживаться, а выбитый с прибора запрос становится в начало очереди (дисциплина LCFS Preemtive Resume). В этом случае процесс $\mathbf{n}(t)$, вообще говоря, не является марковским. Тем не менее для него справедлива

Теорема 1. Цепь Маркова, описывающая число запросов в сети в момент времени t, эргодична, если выполнено условие $\frac{\lambda\varepsilon_i}{\mu_i+\nu_i} < 1$, $i = \overline{1, N}$, а ее стационарное распределение имеет вид $p(\mathbf{n}) = p_1(n_1)\dots p_N(n_N)$ с множителями

$$p_i(n_i) = \left(\frac{\lambda\varepsilon_i}{\mu_i + \nu_i}\right)^{n_i} \left(1 - \frac{\lambda\varepsilon_i}{\mu_i + \nu_i}\right).$$

Доказательство. Пусть $\tau_{ik}(t)$ – остаточное время обслуживания запроса в i-м узле с момента t до момента окончания времени обслуживания, а $\tau_i(t) = (\tau_{i1}(t), \tau_{i2}(t), \dots, \tau_{in_i}(t))$ – вектор, описывающий, остаточное время пребывания запросов в i-м узле; где k-номер позиции, на которой находится запрос от “хвоста” к прибору. Поскольку, вообще говоря, $\mathbf{n}(t)$ не является марковским процессом, рассмотрим марковский процесс $\zeta(t) = (\mathbf{n}(t), \tau(t))$, добавляя к $\mathbf{n}(t)$ непрерывную компоненту $\tau(t) = (\tau_1(t); \dots; \tau_N(t))$. Пусть выполнено условие

$$\frac{\lambda\varepsilon_i}{\mu_i + \nu_i} < 1, \quad i = \overline{1, N},$$

т.е. в случае, когда $\mathbf{n}(t)$ – марковский процесс, существует стационарное эргодическое распределение $\mathbf{n}(t)$, а, следовательно, в общем случае и процесса $\tau(t)$, так как $\tau(t)$ получается из $\mathbf{n}(t)$ добавлением непрерывных компонент. Доказательство этого факта может быть проведено либо по схеме, предложенной в [12], либо с помощью предельной теоремы Смита для регенерирующих процессов [13]. Введем обозначение

$$F(\mathbf{n}, x) = F(\mathbf{n}, x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}, \dots, x_{N1}, \dots, x_{Nn_N}) =$$
$$= \lim_{t\to\infty} P\{n(t) = \mathbf{n}, \tau_{i1}(t) < x_{i1}, \dots, \tau_{in_i}(t) < x_{in_i}, i = \overline{1, N}\}.$$

Несложно доказать, что при $\mu^{-1} < \infty$ существуют конечные пределы

$$F(\mathbf{n}, x) = \lim_{t\to\infty} F(\mathbf{n}, t, x).$$

Введем обозначения: $[\widetilde{x}_i]$ – вектор, все элементы которого совпадают с элементами вектора $x_1, \ldots, x_N$, а на месте i-го элемента находится элемент $\widetilde{x}_i$, и $[\widetilde{x}_i, \widetilde{x}_j]$ – вектор, все элементы которого совпадают с элементами вектора $x_1, \ldots, x_N$, а на месте i-го и j-го элементов находятся элементы $\widetilde{x}_i$ и $\widetilde{x}_j$ соответственно.

Для $F(\mathbf{n}, x)$ справедлива следующая система дифференциально-разностных уравнений:

$$\begin{aligned}
&\sum_{i=1}^{N} \lambda p_{0i} F(\mathbf{n}, x) + \sum_{i=1}^{N} F(\mathbf{n}, x) B_i(x_{i,n_i})\nu_i = \\
&= \sum_{i=1}^{N} F(\mathbf{n} - e_i, [x_{i,1}, \ldots, x_{i,n_i-1}]) I_{n_i\neq 0} B_i(x_{i,n_i})\lambda p_{0i} + \\
&+ \sum_{i=1}^{N} \frac{\partial F(\mathbf{n}, [x_{i,1}, \ldots, x_{i,n_i-1}, 0])}{\partial x_{i,n_i}} B_i(x_{i,n_i}) p_{ii} + \\
&+ \sum_{i=1}^{N} F(\mathbf{n}, x) B_i(x_{i,n_i})\nu_i r_{ii} + \\
&+ \sum_{i=1}^{N} \frac{\partial F(\mathbf{n} + e_i, [x_{i,1}, \ldots, x_{i,n_i}, 0])}{\partial x_{i,n_i+1}} p_{i0} + \\
&+ \sum_{i=1}^{N} F(\mathbf{n} + e_i, [x_{i,1}, \ldots, x_{i,n_i}, x_{i,n_i+1}])\nu_i r_{i0} + \\
&+ \sum_{i=1}^{N} \left(\frac{\partial F(\mathbf{n}, x)}{\partial x_{i,n_i}} - \frac{\partial F(\mathbf{n}, [x_{i,1}, \ldots, x_{i,n_i-1}, 0])}{\partial x_{i,n_i}} \right) + \\
&+ \sum_{j=1}^{N} \sum_{i=1, i\neq j}^{N} \frac{\partial F(\mathbf{n} + e_j - e_i, [x_{j,1}, \ldots, x_{j,n_j}, 0])}{\partial x_{j,n_j+1}} B_i(x_{i,n_i}) p_{ji} + \\
&+ \sum_{j=1}^{N} \sum_{i=1, i\neq j}^{N} F(\mathbf{n} + e_j - e_i, [x_{j,1}, \ldots, x_{j,n_j}, x_{j,n_j+1}]) B_i(x_{i,n_i})\nu_j r_{ji}.
\end{aligned} \tag{4}$$

Перепишем уравнение следующим образом:

(5)
$$\sum_{i=1}^{N} \lambda p_{0i} F(\mathbf{n}, x) = \sum_{i=1}^{N} \frac{\partial F(\mathbf{n}+e_i, [x_{i,1}, \dots, x_{i,n_i}, 0])}{\partial x_{i,n_i+1}} p_{i0} + \\ + \sum_{i=1}^{N} F(\mathbf{n}+e_i, [x_{i,1}, \dots, x_{i,n_i}, x_{i,n_i+1}]) \nu_i r_{i0};$$

(6)
$$\sum_{i=1}^{N} \left(\frac{\partial F(\mathbf{n}, [x_{i,1}, \dots, x_{i,n_i-1}, 0])}{\partial x_{i,n_i}} - \frac{\partial F(\mathbf{n}, x)}{\partial x_{i,n_i}} \right) - \\ - \sum_{i=1}^{N} \frac{\partial F(\mathbf{n}, [x_{i,1}, \dots, x_{i,n_i-1}, 0])}{\partial x_{i,n_i}} B_i(x_{i,n_i}) p_{ii} + \\ + \sum_{i=1}^{N} F(\mathbf{n}, x) B_i(x_{i,n_i}) \nu_i - \sum_{i=1}^{N} F(\mathbf{n}, x) B_i(x_{i,n_i}) \nu_i r_{ii} = \\ = \sum_{i=1}^{N} F(\mathbf{n}-e_i, [x_{i,1}, \dots, x_{i,n_i-1}]) I_{n_i \neq 0} B_i(x_{i,n_i}) \lambda p_{0i} + \\ + \sum_{j=1}^{N} \sum_{i=1, i\neq j}^{N} \frac{\partial F(\mathbf{n}+e_j-e_i, [x_{j,1}, \dots, x_{j,n_j}, 0])}{\partial x_{j,n_j+1}} B_i(x_{i,n_i}) p_{ji} + \\ + \sum_{j=1}^{N} \sum_{i=1, i\neq j}^{N} F(\mathbf{n}+e_j-e_i, [x_{j,1}, \dots, x_{j,n_j}, x_{j,n_j+1}]) B_i(x_{i,n_i}) \nu_j r_{ji}.$$

Нетрудно убедиться, что неотрицательным абсолютно непрерывным по x решением уравнений (5), (6), а, следовательно, и уравнения (4) является

(7)
$$F(\mathbf{n}, x) = p(\mathbf{n}) \prod_{i=1}^{N} \prod_{k=1}^{n_i} \mu_i \int_0^{x_{ik}} [1 - B_i(u)] du,$$

где $p(\mathbf{n})$ – стационарная вероятность состояния n в процессе $\mathbf{n}(t)$ в марковском случае. Действительно, подставляя (7) в (5), поделив обе части полученного уравнения на $F(\mathbf{n}, x)$ и умножив на $p(\mathbf{n})$, получим:

$$p(\mathbf{n})\lambda = \sum_{i=1}^{N} p(\mathbf{n}+e_i)(\mu_i p_{i0} + \nu_i r_{i0}).$$

Затем подставив (7) в (6), поделив обе части полученного соотношения на $F(\mathbf{n},x)B_i(x_{i,n_i})$ и умножив на $p(\mathbf{n})\int_0^{x_{i,n_i}}[1-B_i(u)]du$, получим:

$$
p(\mathbf{n})\left(\sum_{i=1}^{N}(\mu_i(1-p_{ii})+\nu_i(1-r_{ii}))I_{n_i\neq 0}\right)=
$$
$$
=\sum_{i=1}^{N}p(\mathbf{n}-e_i)\lambda p_{0i}+\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}p(\mathbf{n}+e_j-e_i)(\mu_j p_{ji}+\nu_j r_{ji}). \tag{8}
$$

Сложив оба полученных уравнения, получим уравнение глобального равновесия для марковского случая (2), и с учетом, что $F(\mathbf{n},+\infty)=p(n)$, теорема 1 доказана.

4. Заключение

В статье для сети массового обслуживания с однолинейными узлами и экспоненциальным ограничением на время пребывания запросов в узлах доказана нечувствительность стационарного распределения по отношению к распределениям длительностей обслуживания запросов при фиксированных первых моментах этих распределений.

Полученные результаты могут применяться при проектировании новых и модернизации уже существующих информационно-вычислительных сетей и сетей передачи данных.

СПИСОК ЛИТЕРАТУРЫ

1. *Baskett F.E., Chandy K.M., Muntz R.R., Palacios F.G.* Open, closed and mixed networks of queues with different classes of customers // J. Assoc. Comput. Mach. 1975. V. 22. No. 2. P. 248–260.
2. *Towsley D.* Queueing network models with state-dependent routin // J. Assoc. Comput. Mach. 1980. V. 27. No. 2. P. 323–337.
3. *Малинковский Ю.В.* Инвариантность стационарного распределения состояний модифицированных сетей Джексона и Гордона–Ньюэлла // АиТ. 1998. № 9. С. 29–35.
4. *Ивницкий В.А.* Об условии инвариантности стационарных вероятностей для сетей массового обслуживания // Теория вероятностей и ее применения. 1982. Т. 27. № 1. С. 188–192.
5. *Barbour A.D.* Networks of queues and the method of stages // Adv. Appl. Probab. 1976. V. 8. No. 3. P. 584–591.
6. *Chandy K.M., Howard J.H., Jr., Towsley D.F.* Product-form and local balance in queueing networks // J. Assoc. Comput. Mach. 1977. V. 24. No. 2. P. 250–263.
7. *Samelson C.L., Bulgren W.G.* A note on product-form solution for queueing networks with Poisson arrivals and general service-time distributions with finite means // J. Assoc. Comput. Mach. 1982. V. 29. No. 3. P. 830–840.

8. *Скоба А.Н., Состина Е.В.* Применение аппарата сетей массового обслуживания для аналитико-численного моделирования работы информационной системы без учета влияния блокировок // Инженерный вестник Дона. 2015. Вып. 3.
9. *Кузнецов Н.А., Семенихин К.В.* Анализ и оптимизация управляемой модели замкнутой сети массового обслуживания // АиТ. 2020. № 3. С. 67–85.
10. *Малинковский Ю.В.* Сети Джексона с однолинейными узлами и ограниченным временем пребывания или ожидания // АиТ. 2015. № 4. С. 67–79.
11. *Малинковский Ю.В.* Стационарное распределение вероятностей состояний G-сетей с ограниченным временем пребывания // АиТ. 2017. № 10. С. 155–167.
12. *Коваленко И.Н.* Об условии независимости вероятностей состояний системы от вида закона распределения времени обслуживания // Проблемы передачи информации. 1962. Вып. 11. С. 147–151.
13. *Гнеденко Б.В., Коваленко И.Н.* Введение в теорию массового обслуживания // М.: Наука, 1966.

Статья представлена к публикации членом редколлегии В.М. Вишневским.

Поступила в редакцию 30.03.2024

После доработки 28.06.2024

Принята к публикации 25.07.2024