Ссылка на статью:
// Ученые записки УлГУ. Сер. Математика и информационные технологии. УлГУ.Электрон. журн. 2022, № 1, с. 56-64.

Поступила: 19.04.2022
Окончательный вариант: 19.04.2022

УДК 519.872.3

Математическая модель многоканальной СМО с динамическим приоритетом

Савинов Ю.Г.*, Подгорнов М.Д.

*uras@aport.ru
УлГУ, Ульяновск, Россия

В работе развивается семимартингальный (траекторный) подход к математическому описанию и моделированию систем массового обслуживания (СМО) с приоритетами в обслуживании. Рассмотрена модель многоканальной СМО с динамическим приоритетом. Показан переход от математической модели к итерационным формулам, по которым проводится имитационное моделирование.

Ключевые слова*: система массового обслуживания, динамический приоритет, семимартингальное описание, точечный процесс, компенсатор.*

Введение

В работе [1] была описана одноканальная модель СМО с динамическим приоритетом. Под динамическим приоритетом понимается ситуация, когда приоритет заявок различных классов/типов заявок меняется со временем. Управление приоритетом в [1] происходило с использованием процесса «телеграфного типа», скачки которого определялись моментами ухода обслуженных заявок из СМО. В данной статье приводится обобщение модели [1] для многоканального случая, а для управления приоритетом использован альтернативный подход, основанный на регулировании размеров очередей.

В настоящее время распространены два основных способа описания и моделирования СМО: марковский (см., например, работы [2-27]) и траекторный (см., например, работы [28-37]). Одноканальные СМО с приоритетами достаточно хорошо исследуются марковскими методами (см., например, [5]). К сожалению, сложность анализа резко возрастает при переходе к многоканальным системам с приоритетами, особенно для СМО с различной средней длительностью обслуживания (см., например, [24], [25]). Еще сложнее анализ многоканальных систем в случае относительных приоритетов, что связано с непомерным разрастанием пространства состояний [22].

Основным преимуществом траекторного описания является легкость перехода от математической модели к итерационным формулам, по которым проводится имитационное моделирование, причем сложность модели практически не растет с ростом числа каналов, в отличие от моделей в марковском описании. Поэтому в данной работе, с учетом того, что аналитическое исследование рассматриваемой в работе СМО затруднено, выбрано семимартингальное описание модели в терминах точечных (считающих) процессов и их компенсаторов [29].

Для расчета многоканальных немарковских СМО также можно использовать аппроксимацию непоказательных распределений распределениями фазового типа [23], но сложность таких аппроксимаций также быстро растет с увеличением числа каналов в СМО.

В работе показано, что при семимартингальном описании этой проблемы не возникает, и сложность как математической, так и компьютерной модели практически не растет с ростом числа каналов в СМО.

1. Постановка задачи

Рассмотрим многоканальную СМО, в которую поступают заявки двух типов: первый тип – важные, но поступающие редко заявки, второй тип – менее важные, но часто поступающие заявки, которые при этом нельзя терять. Заявки обоих типов поступают независимо друг от друга и образуют простейшие потоки с интенсивностями $\lambda_1 > 0$, $\lambda_2 > 0$ (см. рис. 1).

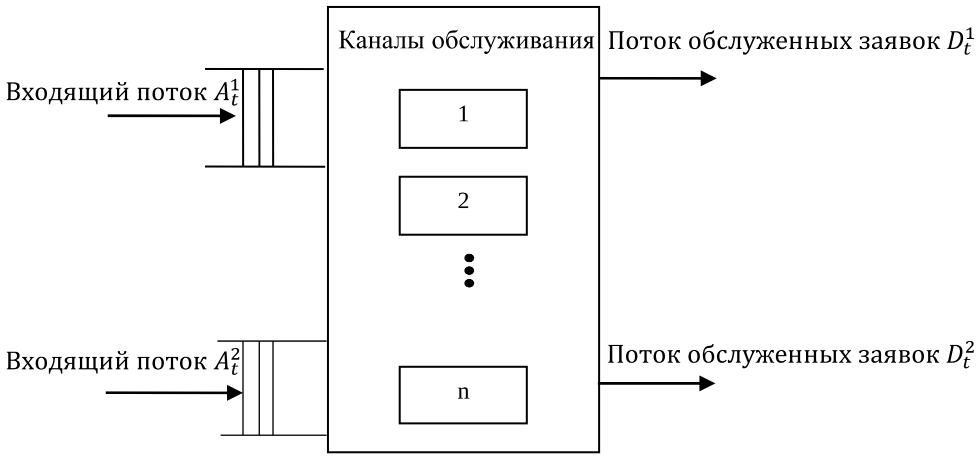

Рис. 1. Схема СМО

Обслуживают заявки n операторов, имеющих одинаковую квалификацию со средним временем обслуживания $\mu_1 > 0$ для первого типа заявок и $\mu_2 > 0$ для второго. Для каждого типа заявок формируется отдельная неограниченная очередь.

Общая очередь в данном случае, очевидно, не эффективна, поскольку в таком случае важные, но редкие заявки первого типа будут долго ждать начала обслуживания, что может быть неприемлемо. Другие стратегии разделения ресурса (в данном случае n операторов), такие как стратегия абсолютного приоритета или стратегия подвижной границы [26],

[37] также не эффективны, так как допускают или потери заявок второго типа или большую часть времени операторы, зарезервированные под заявки первого типа, будут простаивать.

Поэтому предложена следующая стратегия с динамическим приоритетом. Заявки первого типа имеют относительный приоритет (ожидают окончания обслуживания текущей заявки и после этого встают на внеочередное обслуживание). После обслуживания заявки первого типа приоритет переходит к заявкам второго типа, если есть заявки второго типа в очереди для заявок второго типа. Если заявок второго типа в очереди нет, то приоритет остается у заявок первого типа. Аналогично, после выхода из СМО обслуженной заявки второго типа приоритет переходит к заявкам первого типа, если есть заявки первого типа в соответствующей очереди. Если заявок первого типа в очереди нет, то приоритет остается у заявок второго типа. То есть при наличии заявок обоих типов заявки проходят на обслуживание «через одного», а при отсутствии заявок одного из типов (заявки первого типа редки по условию) приоритет (очередность) не передается. Таким способом достигается приемлемое среднее время ожидания начала обслуживания для редких, но важных заявок первого типа при отсутствии необслуженных заявок второго типа и минимальных простоях операторов.

2. Математическая модель

1. Опишем сначала работу по обработке заявок первого типа.

Введем считающие процессы: $A^1 = (A^1_t)_{t\geq 0}$ - число заявок первого типа, поступивших в систему за время $t \geq 0$, $A^1_0 = 0$; $D^1 = (D^1_t)_{t\geq 0}$- число обслуженных заявок первого типа за время $t \geq 0$, $D^1_0 = 0$.

Точечные процессы A^1, D^1 определяются своими компенсаторами $\tilde{A}^1 = (\tilde{A}^1_t)_{t\geq 0}$ и $\widetilde{D}^1 = (\widetilde{D}^1_t)_{t\geq 0}$ в соответствии с разложением Дуба-Мейера для субмартингалов [29]:

$$A^1_t = \tilde{A}^1_t + m^{A^1}_t, \tag{1}$$

$$D^1_t = \widetilde{D}^1_t + m^{D^1}_t, \tag{2}$$

где $\tilde{A}^1$ и $\widetilde{D}^1$ – неубывающие предсказуемые процессы m^{A^1}и m^{D^1}– мартингалы.

Для рассматриваемой в данной работе системы $A^1 = (A^1_t)_{t\geq 0}$ – пуассоновский процесс с компенсатором:

$$\widetilde{A^1_t} = \lambda_1 \cdot t,\ \lambda_1 > 0. \tag{3}$$

Компенсатор для процесса $D^1 = (D^1_t)_{t\geq 0}$ определяется следующим соотношением:

$$\widetilde{D^1_t} = \int_0^t \mu_1\,(\xi^1_s - q^1_s)ds. \tag{4}$$

Здесь ξ^1_t – число заявок первого типа в СМО в момент времени $t \geq 0$, $\xi^1_0 = 0$ (в очереди или на обслуживании), q^1_t – число заявок первого типа в очереди в момент времени $t \geq 0$, $q^1_0 = 0$. Соответственно $(\xi^1_t - q^1_t)$ это число заявок первого типа, находящихся непосредственно на обслуживании в момент времени $t \geq 0$.

Кроме этого, для ξ^1_t в момент времени $t \geq 0$ можно написать следующее основное балансовое соотношение:

$$\xi_t^1 = \xi_0^1 + A_t^1 - D_t^1. \quad (5)$$

2. Аналогично опишем работу по обработке заявок второго типа.

Введем считающие процессы: $A^2 = (A_t^2)_{t\geq 0}$ - число заявок второго типа, поступивших в систему за время $t \geq 0$, $A_0^2 = 0$; $D^2 = (D_t^2)_{t\geq 0}$- число обслуженных заявок второго типа за время $t \geq 0$, $D_0^2 = 0$.

Точечные процессы A^2, D^2 также определяются своими компенсаторами $\tilde{A}^2 = (\tilde{A}_t^2)_{t\geq 0}$ и $\widetilde{D}^2 = (\widetilde{D}_t^2)_{t\geq 0}$:

$$A_t^2 = \tilde{A}_t^2 + m_t^{A^2}, \quad (6)$$

$$D_t^2 = \widetilde{D}_t^2 + m_t^{D^2}, \quad (7)$$

где $\tilde{A}^2$ и $\widetilde{D}^2$ – неубывающие предсказуемые процессы m^{A^2}и m^{D^2}– мартингалы.

Аналогично $A^2 = (A_t^2)_{t\geq 0}$ – пуассоновский процесс с компенсатором:

$$\widetilde{A_t^2} = \lambda_2 \cdot t,\ \lambda_2 > 0. \quad (8)$$

Компенсатор для процесса $D^2 = (D_t^2)_{t\geq 0}$ определяется следующим соотношением:

$$\widetilde{D_t^2} = \int_0^t \mu_2\,(\xi_s^2 - q_s^2)ds. \quad (9)$$

Здесь ξ_t^2 – число заявок второго типа в СМО в момент времени $t \geq 0$, $\xi_0^2 = 0$ (в очереди или на обслуживании), q_t^2 – число заявок второго типа в очереди в момент времени $t \geq 0$, $q_0^2 = 0$.

Для ξ_t^2 в момент времени $t \geq 0$ также можно написать основное балансовое соотношение:

$$\xi_t^2 = \xi_0^2 + A_t^2 - D_t^2. \quad (10)$$

3. Опишем управление приоритетом через регулирование размеров очередей.

Для размера очереди из заявок первого типа q_t^1 в момент времени $t \geq 0$ можно написать следующее балансовое уравнение:

$$dq_t^1 = I(\xi_t^1 + \xi_t^2 \geq n)dA_t^1 - I(q_t^2 = 0,\ q_t^1 > 0)dD_t^1 - I(\ q_t^1 > 0)dD_t^2, \quad (11)$$

где $I(\cdot)$ –индикаторная функция, $q_0^1 = 0$. Логика построения уравнения (11) следующая. Во-первых, размер очереди заявок первого типа q_t^1 может увеличиться на одну заявку, если поступит заявка первого типа ($dA_t^1 = 1$) в момент, когда все n операторов заняты ($\xi_t^1 + \xi_t^2 \geq n$). Во-вторых, размер очереди заявок первого типа q_t^1 может уменьшиться на одну заявку, если закончится обслуживание заявки первого типа ($dD_t^1 = 1$), заявок второго типа в очереди нет ($q_t^2 = 0$), а заявки первого типа в очереди есть ($q_t^1 > 0$). В-третьих, размер очереди заявок первого типа q_t^1 может уменьшиться на одну заявку, если закончится обслуживание заявки второго типа ($dD_t^2 = 1$) и есть заявки первого типа в очереди ($q_t^1 > 0$).

Аналогичное уравнение верно и для размера очереди из заявок второго типа q_t^2:

$$dq_t^2 = I(\xi_t^1 + \xi_t^2 \geq n)dA_t^2 - I(q_t^1 = 0,\ q_t^2 > 0)dD_t^2 - I(\ q_t^2 > 0)dD_t^1,\ q_0^2 = 0\ . \quad (12)$$

Логика построения уравнения (12) полностью совпадает с предыдущей формулой. Во-первых, размер очереди заявок второго типа q_t^2 может увеличиться на одну заявку, если поступит заявка второго типа ($dA_t^2 = 1$) в момент, когда все n операторов заняты ($\xi_t^1 +$

$\xi_t^2 \geq n$). Во-вторых, размер очереди заявок второго типа q_t^2 может уменьшиться на одну заявку, если закончится обслуживание заявки второго типа ($dD_t^2 = 1$), заявок первого типа в очереди нет ($q_t^1 = 0$), а заявки второго типа во второй очереди есть ($q_t^2 > 0$). В-третьих, размер очереди заявок второго типа q_t^2 может уменьшиться на одну заявку, если закончится обслуживание заявки первого типа ($dD_t^1 = 1$) и есть заявки второго типа в очереди ($q_t^2 > 0$).

3. Итерационные формулы для компьютерного моделирования

Выведем формулы, необходимые для имитационного моделирования СМО. На стохастическом базисе $B = (\Omega, \mathcal{F}, F = (\mathcal{F}_t)_{t\geq 0}, P)$ из формул (1)-(12) можно получить следующие инфинитезимальные соотношения:

$$P\{A_{t+\Delta}^1 - A_t^1 = 1|\mathcal{F}_t\} = \lambda_1 \cdot \Delta + o(\Delta), \quad (13)$$

$$P\{A_{t+\Delta}^2 - A_t^2 = 1|\mathcal{F}_t\} = \lambda_2 \cdot \Delta + o(\Delta), \quad (14)$$

$$P\{D_{t+\Delta}^1 - D_t^1 = 1|\mathcal{F}_t\} = \mu_1 \cdot (\xi_s^1 - q_s^1) \cdot \Delta + o(\Delta), \quad (15)$$

$$P\{D_{t+\Delta}^2 - D_t^2 = 1|\mathcal{F}_t\} = \mu_2 \cdot (\xi_s^2 - q_s^2) \cdot \Delta + o(\Delta), \quad (16)$$

Формулы (13)-(16) позволяют, основываясь на понятии геометрической вероятности, провести имитационное моделирование. А именно, введя дискретизацию (шаг по времени) Δ из условия $\lambda_i \cdot \Delta \ll 1$, $\mu_i \cdot (\xi_s^i - q_s^i) \cdot \Delta \ll 1, i = 1,2$ получим следующие итерационные формулы (для вычисления значений процессов в момент времени $t + \Delta$ через значения процессов в момент t):

$$A_{t+\Delta}^1 = A_t^1 + \delta(\lambda_1), \quad (17)$$

$$A_{t+\Delta}^2 = A_t^2 + \delta(\lambda_2), \quad (18)$$

$$D_{t+\Delta}^1 = D_t^1 + \delta(\mu_1 \cdot (\xi_t^1 - q_t^1)), \quad (19)$$

$$D_{t+\Delta}^2 = D_t^2 + \delta(\mu_2 \cdot (\xi_t^2 - q_t^2)), \quad (20)$$

$$q_{t+\Delta}^1 = q_t^1 + I(\xi_t^1 + \xi_t^2 \geq n)\Delta A_t^1 - I(q_t^2 = 0,\ q_t^1 > 0)\Delta D_t^1 - I(\ q_t^1 > 0)\Delta D_t^2, \quad (21)$$

$$q_{t+\Delta}^2 = q_t^2 + I(\xi_t^1 + \xi_t^2 \geq n)\Delta A_t^2 - I(q_t^1 = 0,\ q_t^2 > 0)\Delta D_t^2 - I(\ q_t^2 > 0)\Delta D_t^1, \quad (22)$$

где

$$\delta(\gamma) = \begin{cases} 1, \text{с вероятностью } \gamma \cdot \Delta, \\ 0, \text{с вероятностью } 1 - \gamma \cdot \Delta. \end{cases}$$

Здесь $\Delta A_t^1 = A_{t+\Delta}^1 - A_t^1, \Delta D_t^1 = D_{t+\Delta}^1 - D_t^1, \Delta A_t^2 = A_{t+\Delta}^2 - A_t^2, \Delta D_t^2 = D_{t+\Delta}^2 - D_t^2$.

4. Результаты компьютерного моделирования

Практическая реализация СМО осуществлена с помощью язык программирования высокого уровня C# в среде разработки Visual Studio 2019 (см. рис. 2, 3). При параметрах T = 10, λ_1 = 1, λ_2 = 10, μ_1 = 2, μ_2 = 5 и количестве каналов, равном 3, система полностью стабильна, что можно понять по графику среднего количества заявок первого и второго типа в СМО (см. рис. 2).

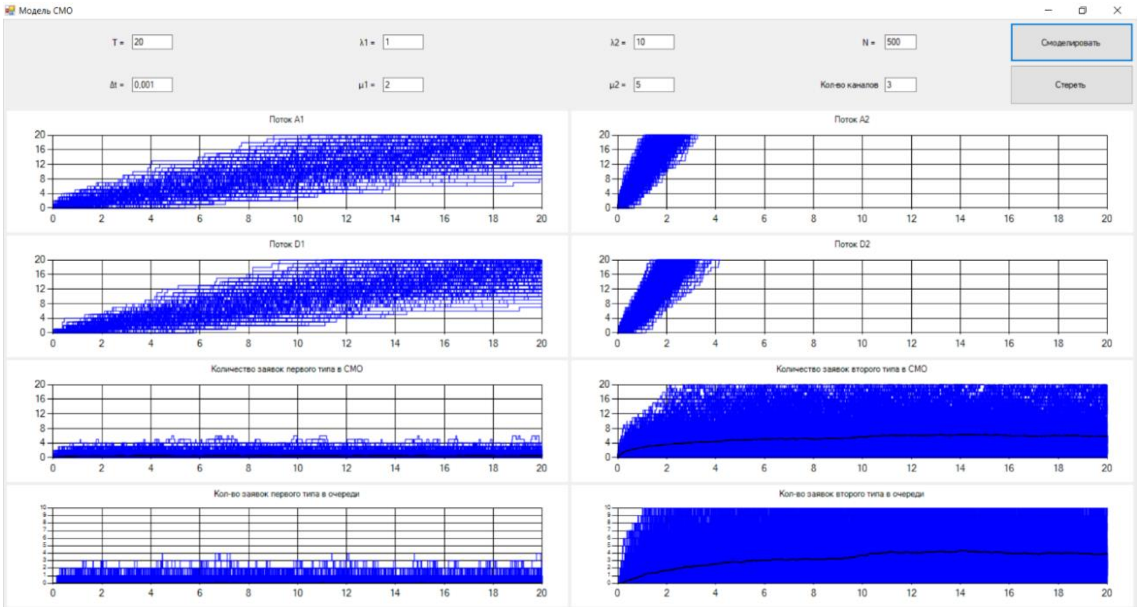

Рис. 2. Диалоговое окно программы. Практическая реализация в среде Visual Studio 2019

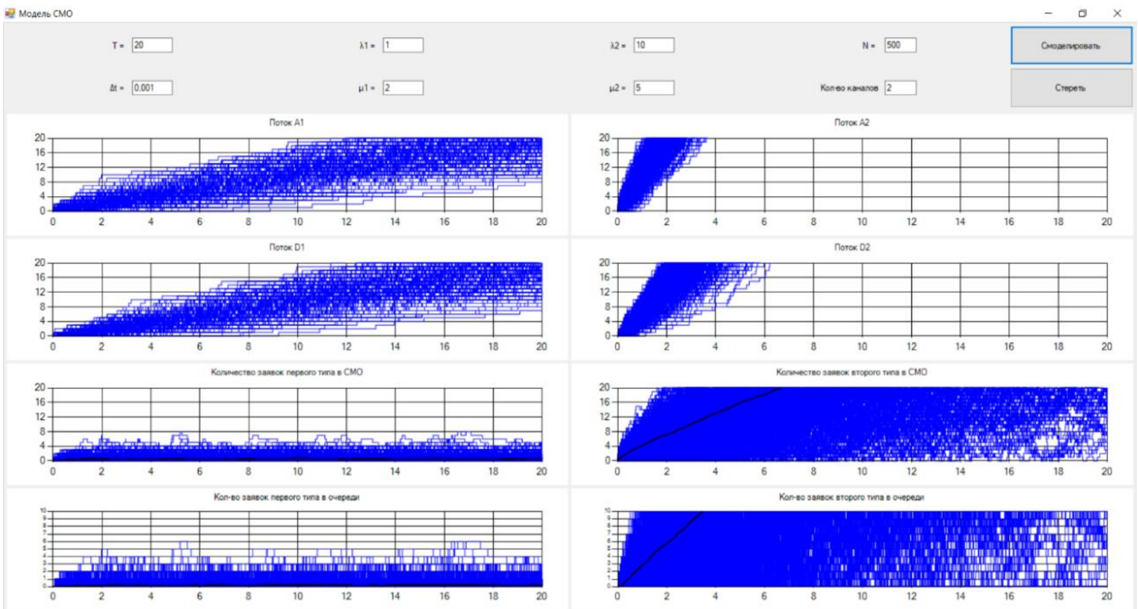

Рис. 3. Диалоговое окно программы при изменении количества каналов

При уменьшении количества обслуживающих каналов до двух картина меняется (см. рис. 3). Очередь заявок второго типа не стабилизируется, т.к. система не успевает справляться с такой нагрузкой. Однако, заявки первого типа, которые являются более редкими, но важными, обслуживаются своевременно, средняя длина их очереди стабильная и низкая. Это наглядно отражает преимущество стратегии динамического приоритета.

Список литературы

1. Столяров И.А., Савинов Ю.Г. Семимартингальная модель СМО с динамическим приоритетом // *Прикладная математика и информатика: современные исследования в области естественных и технических наук: материалы III научно-практической всероссийской конференции (школы-семинара) молодых ученых: 24-25 апреля 2017 г. Тольятти: Издатель Качалин Александр Васильевич*, 2017. с. 553–557.
2. Гнеденко Б.В., Коваленко И.Н. Введение в теорию массового обслуживания. М.:URSS, 2013. 400 с.
3. Риордан Д. *Вероятностные системы обслуживания*. Пер. с англ. М.: Связь, 1966. 184 с.
4. Алиев Т.И. *Основы моделирования дискретных систем*. СПб: СПбГУ ИТМО, 2009. 363 с.
5. Джейсуол Н. К. *Очереди с приоритетами*. Пер. с англ. М.: Мир, 1973. 279 с.
6. Бронштейн О. И., Духовный И. М. *Модели приоритетного обслуживания в информационно-вычислительных системах.* М.: Наука, 1976. 220 с.
7. Лысенкова В. Т. *Исследование многолинейных систем массового обслуживания с ограниченным накопителем и приоритетами*. Автореф. дис. … канд. техн. Наук. Институт проблем передачи информации. М., 1973.
8. Haugen R.B., Skogan E. Queueing systems with stochastic time out // *IEEE Transactions on Communications*. 1980, v.28, p.1984–1989.
9. Мова В. В., Пономаренко Л. А., Калиновский А. М. *Организация приоритетного обслуживания в АСУ.* Киев: Техника, 1977. 160 с.
10. Рыжиков Ю. И. Комплекс программ для расчета систем массового обслуживания повышенной сложности // *Программирование.* 1978, № 4, с. 87–91.
11. Palm C. Research on telephone traffic carried by full availability groups // *Tele*. 1957, no.1, p. 107.
12. Саати Т. Л. *Элементы теории массового обслуживания и ее приложения*. Пер. с англ. М.: Сов. радио, 1965. 510 с.
13. Aguir S., Karaesmen F., Aksin O.Ż., Dallery Y. On the interaction between retrials and sizing of call centers // *European J. Oper. Res.* 2008, v. 191, no. 2, p. 398–408.
14. Canales, M., Hernández-Solana, Á., Gállego, J.R. et al. Adaptive resource sharing strategies for UMTS multiservice mobiles // Telecommun. Syst. 2005, v. 28, p. 151–167. https://doi.org/10.1007/s11235-004-5014-0
15. Gerasimenko M., Moltchanov D., Andreev S., Koucheryavy Y., Himayat N., Yeh S.-P., Talwar S. Adaptive resource management strategy in practical multi-radio heterogeneous networks // *IEEE Access. Institute of Electrical and Electronics Engineers Inc.* 2017, v. 5, p. 219–235.

16. Antonioli R.P., Rodrigues E.B., Maciel T.F. et al. Adaptive resource allocation framework for user satisfaction maximization in multi-service wireless networks // *Telecommun. Syst.* 2018, v. 68, p. 259–275. https://doi.org/10.1007/s11235-017-0391-3
17. Росляков А.В., Цыганков Н.И. Анализ моделей распределенных центров обслуживания вызовов // *Электросвязь*. 2005, №8, с. 22–25.
18. Зарипова Э.Р. Математическая модель центра обслуживания вызовов с двумя типами абонентов // *Discrete and Continuous Models and Applied Computational Science*. 2010. №4, с.76–82.
19. Ваняшин С.В. Исследование и разработка моделей мультисервисного центра обслуживания вызовов: дис. ... канд. тех. наук: 05.12.13. М.: 2006. 157 с.
20. Томашевский В. Л. Многоканальные приоритетные системы массового обслуживания: автореф. дис. … канд. техн. наук. М.: МГУ, 1986. 14 с.
21. Хомоненко А. Д. Вероятностный анализ приоритетного обслуживания с прерываниями в многопроцессорных системах // *Автоматика и вычислительная техника*. 1990, № 2, с. 55–61.
22. Рыжиков Ю. И. Средние времена ожидания и пребывания в многоканальных приоритетных системах // *Информационно-управляющие системы*. 2006, № 6, с. 43–49.
23. Рыжиков Ю.И., Уланов А.В. Применение гиперэкспоненциальной аппроксимации в задачах расчета немарковских систем массового обслуживания // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2016, № 3(36), с. 60–65.
24. Gail H. R., Hantler S. L., Taylor B. A. Analysis of a non-preemptive priority multiserver queue // *Advances in applied prob.* 1988, v. 20, i.4.
25. Miller D. R. Steadystate algorithmic analysis of M/M/c two priority queues with heterogeneous rates // *Applied probability – computer science: the interface.* 1982, v. 2. Boston: Birkhauser. P. 207–222.
26. Крылов В.В., Самохвалова С.С. *Теория телетрафика и ее приложения.* СПб.: БХВ-Петербург, 2005. 288 с.
27. Бутов А.А., Савинов Ю.Г. *Теория массового обслуживания*: учебно-методическое пособие. Ульяновск: УлГУ, 2007.
28. Бутов А.А., Галимов Л.А. Оптимальное управление интенсивностью входящего потока многоканальной СМО с роутером при эпизодически наблюдаемой длине очереди на приборах // *Современные проблемы науки и образования*. 2015, № 2, с. 758.
29. Бутов А.А., Раводин К.О. *Теория случайных процессов*: учебно-методическое пособие. Ульяновск: УлГУ, 2009.
30. Бутов А.А., Галимов Л.А. Стохастическая имитационная модель оценки резерва произошедших, но не заявленных страховых убытков в терминах СМО // *Фундаментальные исследования*. 2016, № 8–2, с. 234–238.

31. Бутов А.А., Галимов Л.А. Оптимальное управление распределением заявок в многоканальной системе массового обслуживания с входящим пуассоновским потоком поступления заявок и экспоненциальным временем обслуживания // *Естественные и технические науки*. 2014, № 9–10 (77), с. 244–247.
32. Савинов Ю.Г., Исмаилова М.В., Рослов М.Э. Траекторные методы моделирования многофазных СМО // *Ученые записки УлГУ. Серия: Математика и информационные технологии.* 2019, № 1, с. 85–91.
33. Савинов Ю.Г., Тихоненко А.А., Пронин В.И., Щукин А.Н. Семимартингальная модель СМО с произвольным временем ожидания "нетерпеливых" заявок // *Ученые записки УлГУ. Сер. Математика и информационные технологии*. УлГУ. Электрон. журн. 2019, № 2, с. 81–88.
34. Савинов Ю.Г., Сафиуллов И.Д., Дуняшина М.С. Модель обслуживания парка оборудования с возможной задержкой в начале ремонта // *Ученые записки УлГУ. Сер. Математика и информационные технологии.* УлГУ. Электрон. журн. 2020, № 1, с. 103–108.
35. Савинов Ю.Г., Щукин А.Н., Подгорнов М.Д. Математическая модель мультисервисного кол-центра с многоэтапным обслуживанием и дообслуживанием неприоритетных заявок // *Ученые записки УлГУ. Сер. Математика и информационные технологии*. УлГУ. Электрон. журн. 2021, № 1, с. 109–117.
36. Савинов Ю.Г., Пронин В. И., Курицин А. Е. Математическая модель центра обслуживания вызовов со случайной задержкой при многоэтапном обслуживании // *Ученые записки УлГУ. Сер. Математика и информационные технологии. УлГУ.* Электрон. журн. 2021, № 1, с.102–108.
37. Савинов Ю.Г., Рослов М.Э., Куманина Я.А. Стратегия адаптивного разделения ресурсов в мультисервисных сетях// *Ученые записки УлГУ. Сер. Математика и информационные технологии*. УлГУ. Электрон. журн. 2020, № 1, с. 95–102.

Mathematical model of a multi-channel queuing system with dynamic priority

Savinov, Y.G. [*]***, Podgornov, M.D.***

[*] uras@aport.ru

The paper develops a semi-martingale (trajectory) approach to the mathematical description and modeling of queuing systems (QS) with service priorities. The model of multichannel QS with dynamic priority is considered. The transition from a mathematical model to iterative formulas, which are used for simulation, is shown.

Keywords: *queuing system, dynamic priority, semi-martingale description, point process, compensator*