УДК 004.9
DOI 10.17513/snt.40836

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И РЕШЕНИЕ ЗАДАЧИ НАХОЖДЕНИЯ ХАРАКТЕРИСТИК СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ ЗАЯВОК СО СЛУЧАЙНЫМ ВРЕМЕНЕМ ЖИЗНИ

Шайдуллина Н. К., Нуриев Н. К., Печеный Е. А.

Федеральное государственное бюджетное образовательное учреждение высшего образования «Казанский национальный исследовательский технологический университет», Казань, Российская Федерация, e-mail: nshaydullina@yandex.ru

В работе рассмотрены одноканальные системы массового обслуживания заявок, время жизни которых ограничено случайной величиной. Целью исследования является получение аналитического решения задачи нахождения характеристик функционирования одноканальной системы массового обслуживания с накопителем и случайным ограничением на время жизни заявок для проверки адекватности имитационной модели. Приведены примеры реальных систем со случайным временем жизни заявок, моделируемых как системы массового обслуживания. Описана математическая модель задачи нахождения функции распределения фактического времени ожидания обслуживания для стационарного состояния одноканальной системы обработки простейшего потока заявок, концепция которой разработана Б. В. Гнеденко и И. Н. Коваленко. Приведена разновидность исходной модели для системы обслуживания заявок со случайным ограничением на общее время пребывания и стандартной дисциплиной обслуживания заявок в порядке их поступления. В качестве примера рассмотрена система с простейшим входным потоком требований, экспоненциально распределенным временем обслуживания и равномерно распределенным временем жизни заявок. Для нахождения искомой характеристики построено и решено интегральное уравнение. С помощью полученной функции распределения фактического времени ожидания обслуживания найдены доли заявок, получивших полное и неполное обслуживание, совсем не попавших на обслуживание. Проведено сравнение результатов работы аналитической и имитационной моделей, говорящее о высокой точности последней.

Ключевые слова: система массового обслуживания, очередь, случайное время жизни заявок, распределение фактического времени ожидания обслуживания, имитационная модель

MATHEMATICAL MODELING AND SOLUTION TO THE PROBLEM OF FINDING THE CHARACTERISTICS OF A QUEUE SERVICE SYSTEM WITH A RANDOM LIFETIME

Shaydullina N. K., Nuriev N. K., Pechenyy E. A.

Federal State Budgetary Educational Institution of Higher Education "Kazan National Research Technological University", Kazan, Russian Federation, e-mail: nshaydullina@yandex.ru

This paper considers single-channel queuing systems with a randomly bounded lifetime. The aim of the study is to obtain an analytical solution to the problem of finding the operating characteristics of a single-channel queuing system with an accumulator and a random constraint on the lifetime of requests in order to verify the adequacy of the simulation model. Examples of real systems with a random lifetime of requests, modeled as queuing systems, are given. A mathematical model is described for the problem of finding the distribution function of the actual waiting time for service for the steady state of a single-channel system for processing a simple request flow, the concept of which was developed by B. V. Gnedenko and I. N. Kovalenko. A variation of the original model is presented for a request servicing system with a random constraint on the total sojourn time and a standard discipline of servicing requests in the order they are received. As an example, a system with a simple input flow of requests, exponentially distributed service time, and uniformly distributed lifetime of requests is considered. To find the desired characteristic, an integral equation is constructed and solved. Using the resulting distribution function of actual service wait times, we determined the proportions of requests that received full and partial service, as well as those that were not serviced at all. A comparison of the results of the analytical and simulation models was conducted, demonstrating the high accuracy of the latter.

Keywords: queueing system, queue, random lifetime of requests, distribution of actual waiting time for service, simulation model

Введение

Системы массового обслуживания с ограниченным временем жизни заявок встречаются во многих областях деятельности человека: информационные технологии, финансы, телемедицина, техника, маркетинг и др. [1–3]. Характерной чертой таких систем является риск утраты заявок в процессе обслуживания, что приводит к неэффективному использованию ресурсов [4].

Ограничение на срок жизни требования может определяться случайным образом, согласно некоторому вероятностному распределению [5–7]. Случайное время пребывания заявки в системе характерно для многих важных процессов: оказание экстренной помощи тяжелым больным, затяжные процессы со сроком давности в юриспруденции, спасение людей в чрезвычайных ситуациях, обработка объектов в ограниченной зоне действия обслуживающего устройства [8].

Получение эксплуатационных характеристик таких СМО необходимо не только для оценки эффективности их функционирования, но и для прогнозирования поведения систем при различных сценариях нагрузки в процессе предпроектных исследований.

Цель исследования – получение аналитического решения задачи нахождения характеристик функционирования одноканальной системы массового обслуживания с накопителем и случайным ограничением на время жизни заявок для проверки адекватности имитационной модели.

Материалы и методы исследования

Системы массового обслуживания с ограниченным временем жизни заявок требуют особого подхода к изучению и специализированного математического аппарата, так как введение временных границ нарушает свойство отсутствия последействия, характерное для классических марковских моделей. Это происходит из-за того, что вероятность перехода заявки из состояния ожидания в состояние обслуживания начинает зависеть от времени, которое она уже провела в накопителе. Таким образом, пространство состояний перестает быть дискретным, что влечет необходимость введения функций распределения дополнительных случайных величин. Это позволяет сохранить аналитическую замкнутость и избежать расширения пространства состояний системы.

Одними из первых публикаций, посвященных моделированию СМО с ограниченным временем жизни заявок, являются работы советских математиков Б. В. Гнеденко и И. Н. Коваленко [9, 10], а также венгерского ученого Л. Такача [11]. Ими проработаны концептуальные вопросы моделирования одноканальных систем с простейшим входным потоком заявок и предложен соответствующий математический аппарат. Фокус исследований сосредоточен на нахождении функции распределения фактического времени ожидания обслуживания, которая является аналитическим базисом для определения ряда производных характеристик СМО. В частности, на ее основе находятся функция распределения реального времени пребывания заявок в системе, вероятность полного обслуживания, вероятность отказа и др. Таким образом, построение функции распределения фактического времени ожидания обслуживания обеспечивает полное вероятностное описание рассматриваемой СМО.

Результаты исследования и их обсуждение

Рассмотрим одноканальную СМО с накопителем, время жизни заявок в которой ограничено и является случайной величиной, входной поток требований простейший.

Для построения модели вводятся следующие характеристики:

λ – интенсивность простейшего входного потока заявок;

$H(x)$ – функция распределения случайной величины – времени обслуживания;

$B(x)$ – распределение вероятностей того, что время ожидания заявки в очереди равно некоторой случайной величине x;

$G_y(x)$ – распределение вероятностей того, что после пребывания заявки в очереди в течение отрезка времени y, заявка попадет на обслуживание, где будет находиться в течение отрезка времени, не большего x;

$\xi(t)$ – случайная величина – время ожидания начала обслуживания заявки, поступившей в систему в момент времени t;

$F(t, x) = P(\xi(t) < x)$ – функция распределения фактического времени ожидания обслуживания.

Пусть заявка пришла в систему в момент времени t. За промежуток времени $t + \Delta t$ могут произойти следующие несовместные события, которые обеспечат выполнение неравенства $\xi(t + \Delta t) < x$:

− за указанный промежуток времени не поступит ни одна заявка;

− за указанный промежуток времени поступает одна заявка и уходит из очереди из-за истечения срока жизни;

− за указанный промежуток времени поступает одна заявка, ожидает в очереди в течение некоторого времени, начинает обслуживание, но не успевает его завершить из-за истечения срока жизни;

− за указанный промежуток времени поступает одна заявка, ожидает в очереди в течение некоторого времени, начинает обслуживание и завершает его до истечения срока жизни.

Рассмотренные несовместные события образуют полную группу, поэтому вероятность того, что неравенство $\xi(t + \Delta t) < x$ верно, равно сумме вероятностей описанных событий:

$$F(x,t+\Delta t) = (1-\lambda\Delta t)F(x+\Delta t,t) + \\ +\lambda\Delta t\int_0^{x+\Delta t}\{B(y)+[1-B(y)]G_y(x-y+\Delta t)+ \\ +[1-B(y)][1-G_y(x-y+\Delta t)]H(x-y+\Delta t)\}d_yF(y,t). \quad (1)$$

При делении уравнения (1) на Δt, $\Delta t \to 0$ и учете того, что в стационарном состоянии, существование которого показано в работе [12], верно равенство

$$\frac{\partial F(\omega,t)}{\partial t} = 0, \quad (2)$$

получается уравнение для нахождения функции распределения ожидания обслуживания:

$$\frac{\partial F(x)}{\partial x} - \lambda\int_0^x (1-B(y))(1-G_x(x-y))(1-H(x-y))dF(y) = 0. \quad (3)$$

Функция F(x) терпит разрыв в точке 0 и абсолютно непрерывна при $x > 0$ [10]. Если обозначить через $f(x)$ производную справа от функции $F(x)$, то формула (3) преобразуется к виду

$$f(x) - \lambda\int_0^x (1-B(y))(1-G_x(x-y))(1-H(x-y))f(y)dy = \\ = \lambda F(0)(1-G_0(x))(1-H(x)), \quad (4)$$

где $F(0)$ находится из условия нормировки

$$F(0) + \int_0^T f(x)dx = 1, \quad (5)$$

здесь T – максимальное время жизни требования.

Наиболее значимыми эксплуатационными характеристиками функционирования любых СМО являются доли обслуженных и необслуженных заявок. В случае, когда ограничено время жизни заявок, появляется еще и множество заявок, которые начали обслуживание, но не завершили его.

Для того, чтобы получить долю заявок, совсем не попавших на обслуживание, нужно применить формулу

$$P_{\text{ОТК}} = \int_0^T B(y)dF(y). \quad (6)$$

Для того, чтобы найти долю заявок, не получивших полное обслуживание, нужно применить формулу

$$\text{P}_{\text{недо}} = \int_0^2 (1-B(y))\int_0^\infty G_y(z)dH(z)dF(y). \quad (7)$$

Уравнение (4) универсально в том смысле, что обеспечивает возможность моделирования разных способов наложения ограничения на время жизни заявок: отсутствие ограничения как такового, случайное ограничение на время ожидания обслуживания в очереди, случайное ограничение на время пребывания заявок в системе.

Важно отметить, что функция $G_y(x)$ предопределяется только в системах со сложной дисциплиной обслуживания. В случае применения стандартной дисциплины FIFO процесс функционирования системы обработки заявок целиком и полностью обусловлен законами распределения входного потока, потока обслуживания и времени жизни заявок.

В рассматриваемом авторами случае ограничение времени жизни заявок является общим, то есть накладывается на время пребывания заявок в системе. При этом истечение времени жизни требования может состояться как в ходе ожидания, так и в процессе обслуживания. Для такого характера ограничения функция B(x) определяет распределение времени пребывания заявок в системе, уравнение (4) преобразуется к виду

$$f(x)-\lambda\big(1-B(x)\big)\int_0^x\big(1-H(x-y)\big)f(y)\,dy=\lambda F(0)\big(1-B(x)\big)\big(1-H(x)\big), \quad (8)$$

а функция $G_y(x)$ находится как

$$G_y(x)=\big(B(x+y)-B(y)\big)/\big(1-B(y)\big). \quad (9)$$

Уравнение (8) относится к линейным уравнениям Вольтерра второго рода. Основными методами решения таких уравнений являются метод преобразований Лапласа или метод последовательных приближений. В отдельных случаях оно может быть сведено к дифференциальному уравнению дифференцированием по верхнему пределу.

Оценим трудоемкость аналитического решения на конкретном примере.

Пусть средняя интенсивность простейшего входного потока заявок $\lambda = 6$, средняя интенсивность обслуживания $\mu = 10$, ограничение времени жизни заявки распределено равномерно на интервале [0, 2]. Тогда уравнение (8) принимает вид

$$f(x)-6e^{-10x}\left(1-\frac{x}{2}\right)\int_0^x e^{10y}f(y)\,dy=6F(0)e^{-10x}\left(1-\frac{x}{2}\right). \quad (10)$$

Решим уравнение, учитывая нормировку (3) при условии $x <= 2$.

Введем обозначение

$$F(x)=\int_0^x e^{10y}f(y)\,dy. \quad (11)$$

Тогда

$$f(x)=e^{-10x}F'(x). \quad (12)$$

Подставив выражение (12) в уравнение (10), получим

$$e^{-10x}F'(x)-6e^{-10x}\left(1-\frac{x}{2}\right)F(x)=6F(0)e^{-10x}\left(1-\frac{x}{2}\right)$$

или

$$F'(x)-(6-3x)F(x)=F(0)(6-3x). \quad (13)$$

Это линейное дифференциальное уравнение первого порядка. Для решения введем интегрирующий множитель:

$$\mu(x)=e^{\int-(6-3\tilde{o})dx}=e^{-6x+1.5x^2}.$$

Решением уравнения (11) является функция

$$F(x)=F(0)(1-e^{6x-1.5x^2}). \quad (14)$$

Тогда

$$f(x)=e^{-10x}F'(x)=F(0)(6-3x)e^{-4x-1.5x^2}. \quad (15)$$

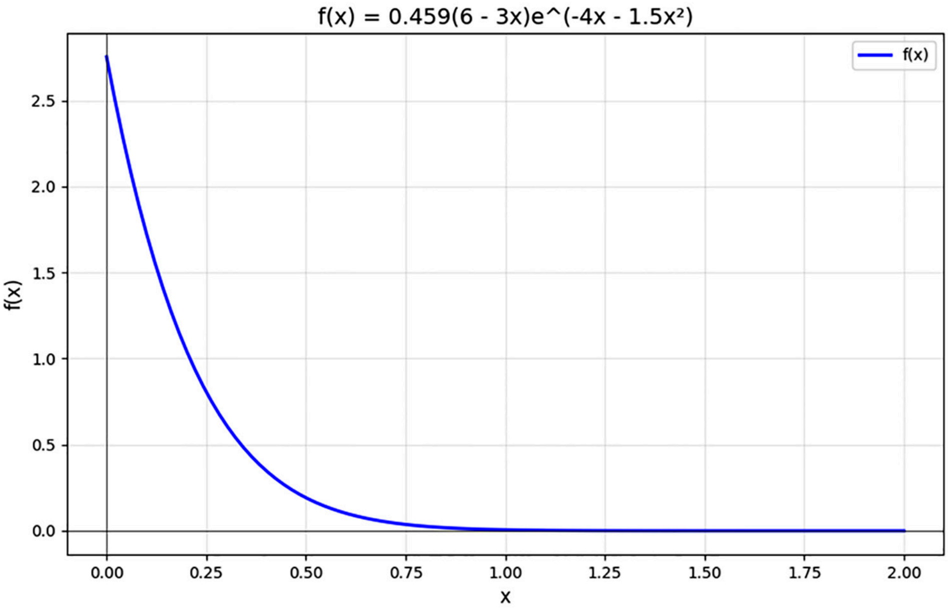
Для нахождения константы $F(0)$ воспользуемся условием нормировки (5):

$$F(0)+F(0)\int_0^2(6-3x)e^{-4x-1.5x^2}dx=1,$$

откуда

$$F(0) \approx 0.456, f(x) \approx 0.456(6-3x)e^{-4x-1.5x^2}.$$

График функции f(x) представлен на рисунке.



Плотность распределения фактического времени ожидания обслуживания
Примечание: составлен авторами по результатам данного исследования

$F(0) \approx 0{,}456$ означает, что примерно 45,6 % поступивших заявок попали на обслуживание без ожидания в очереди.

Получим долю заявок, совсем не попавших на обслуживание по формуле (6):

$$P_{\text{ОТК}} = 0{,}456\int_0^2 \left(\frac{x}{2}\right)(6-3x)e^{-4x-1{,}5x^2}dx \approx 0{,}0483. \tag{16}$$

Вычислим долю заявок, не получивших полного обслуживания, по формуле (7), используя соотношение (9):

$$\text{P}_{\text{недо}} = 10\int_{0{,}0001}^{2}\left(\int_0^\infty \frac{z}{2-y}e^{-10z}dz\right)\left(1-\frac{y}{2}\right)(6-3y)e^{-4y-1{,}5y^2}dy \approx 0{,}0589. \tag{17}$$

Тогда доля заявок, получивших полное обслуживание, находится так:

$$\text{P}_{\text{полн}} = 1 - \text{P}_{\text{ОТК}} - \text{P}_{\text{недо}} = 0{,}8927.$$

Таким образом, при данных входных параметрах почти 90 % пришедших в систему заявок будет обслужено полностью.

Приведенный пример показывает, что даже в простом варианте – с экспоненциальным обслуживанием и равномерно распределенным временем жизни заявок – модель сложна для практического применения. А для вариантов с другими функциями распределения потоков или групповым входным потоком почти непригодна. В этих случаях наиболее целесообразным решением является применение имитационного моделирования [13, 14].

На основе представленной математической модели была проведена верификация имитационной модели, построенной авторами для оценки характеристик СМО с ограниченным временем жизни заявок [15, 16]. Результаты работы имитационного алгоритма с описанными входными параметрами таковы:

$$\text{P}_{\text{полн}} = 0{,}9,\quad \text{P}_{\text{ОТК}} = 0{,}0454,\quad \text{P}_{\text{недо}} = 0{,}052.$$

Полученные данные свидетельствуют о высокой точности имитационной модели.

Выводы

1. Проведен анализ одноканальной системы массового обслуживания заявок со случайным ограничением на общее время пребывания и стандартной дисциплиной обслуживания заявок в порядке их поступления.

2. На основе концептуальной модели Б. В. Гнеденко, И. Н. Коваленко для равномерно распределенного времени жизни заявок сделана постановка задачи и получено аналитическое решение.

3. Путем сравнения с аналитическим решением проведена верификация построенной авторами имитационной модели для изучения поведения СМО с ограниченным временем жизни заявок.

4. Доказана практически (до сотых долей) точность совпадения результатов имитационного и аналитического моделирования, которая позволяет утверждать, что разработанная имитационная модель может быть использована для других функций распределения потоков заявок без поиска аналитического решения.

5. Показано, что в практических целях только имитационное моделирование позволяет разрешать задачи нахождения эксплуатационных характеристик функционирования СМО с ограниченным временем жизни заявок для любых законов распределения случайных величин и любого характера входного потока требований.

6. Обосновано, что разработанная имитационная модель является уникальной и представляет ценность для решения прикладных задач проектирования и эксплуатации СМО: позволяет модифицировать структуру системы; дает возможность анализировать поведение всей системы и отдельных ее частей на этапе проектирования; реализовать поддержку эффективных управленческих решений с привлечением искусственного интеллекта.

Список литературы

1. Лохвицкий В. А., Гончаренко В. А., Левчик Э. С. Модель масштабируемого микросервиса на основе системы массового обслуживания с «охлаждением» // Интеллектуальные технологии на транспорте. 2022. № 1 (29). С. 39–44. DOI: 10.24412/2413-2527-2022-129-39-44. EDN: ZALQRA.

2. Gontis V. Discrete q-Exponential Limit Order Cancellation Time Distribution. // Fractal and Fractional. 2023. № 7 (8). P. 581. DOI: 10.3390/fractalfract7080581.

3. Маколкина М. А., Шарлаева М. В. Исследование средней задержки в сетях связи, предоставляющих телемедицинские услуги // Труды учебных заведений связи. 2024. Т. 10. № 3. С. 59–65. DOI: 10.31854/1813-324X-2024-10-3-59-65. EDN: DRRDAM.

4. Jafarnejad Ghomi E., Rahmani A. M., Qader N. N. Applying queue theory for modeling of cloud computing: A systematic review // Concurrency Computat Pract Exper. 2019. № 31. P. e5186. DOI: 10.1002/cpe.5186.

5. Науменко В. В., Маталыцкий М. А., Монько В. Д. Об имитационном моделировании НМ-сетей с некоторыми особенностями // Вестник Гродненского государственного университета имени Янки Купалы. Серия 2. Математика. Физика. Информатика, вычислительная техника и управление. 2016. Т. 6. № 1. С. 149–156. EDN: VIUTAV.

6. Xiaohu Wu., Francesco De Pellegrini, Giuliano Casale Delay and Price Differentiation in Cloud Computing: A Service Model, Supporting Architectures, and Performance // ACM Trans. Model. Perform. Eval. Comput. Syst. 2023. Vol. 8. Is. 3. Article 6. 40 p. DOI: 10.1145/3592852.

7. Кумков С. И., Игнатенкова Л. А. Интервальный подход к оцениванию стабильности характеристик стандартного образца // Вычислительные технологии. 2017. Т. 22. № 2. С. 85–98. EDN: YPLUFJ.

8. Рыжиков Ю. И. Многоканальные системы обслуживания с марковским нетерпением // Информационные технологии и математическое моделирование (ИТММ-2018): материалы XVII Международной конференции имени А. Ф. Терпугова (г. Томск, 10–15 сентября 2018 г.). Томск: Научно-технической литературы, 2018. С. 125–131. EDN: YMYTJJ.

9. Саати Т. Л. Элементы теории массового обслуживания и ее приложения. М.: Издательство «Советское радио», 1971. 520 с.

10. Коваленко И. Н. Некоторые задачи массового обслуживания с ограничением // Теория вероятностей и ее применение. 1961. Т. 6. № 2. С. 222–228. URL: https://www.mathnet.ru/rus/tvp/v6/i2/p222 (дата обращения: 11.05.2026).

11. Такасs L. Investigation of waiting – time problems by reduction to Markov processes. Acta Math., Acad. Scient. Hung. 1955. Vol. 6. P. 101–129. URL: https://zbmath.org/0067.10903 (дата обращения: 11.05.2026).

12. Афанасьева Л. Г. О существовании предельного распределения в системах массового обслуживания с ограниченным временем пребывания // Теория вероятностей и ее применения. 1965. Т. 10. № 3. С. 570–578. URL: https://www.mathnet.ru/rus/tvp/v10/i3/p570 (дата обращения: 11.05.2026).

13. Шайдуллина Н. К., Печеный Е. А., Нуриев Н. К. Проблема эффективного администрирования системы массового обслуживания с ограниченным временем жизни заявок // Современные наукоемкие технологии. 2022. № 7. С. 69–73. DOI: 10.17513/snt.39235.

14. Шайдуллина Н. К., Печеный Е. А., Нуриев Н. К. Моделирование процесса администрирования системы массового обслуживания с ограниченным временем жизни заявок // Современные наукоемкие технологии. 2023. № 11. С. 81–86. DOI: 10.17513/snt.39824.

15. Шайдуллина Н. К., Печеный Е. А., Нуриев Н. К. Численный анализ поведения смешанного потока заявок с ограниченным временем жизни // Современные наукоемкие технологии. 2024. № 10. С. 94–99. URL: https://top-technologies.ru/ru/article/view?id=40177 (дата обращения: 11.05.2026). DOI: 10.17513/snt.40177.

16. Шайдуллина Н. К. Численное решение задачи эффективного обслуживания группы заявок с ограниченным временем жизни // Актуальные проблемы науки и образования в условиях современных вызовов: материалы XXXIX Международной научно-практической конференции (г. Москва, 03 апреля 2025 г.). М.: ООО «Издательство Академическая среда», 2025. С. 408–413. URL: https://www.elibrary.ru/qptmfo (дата обращения: 11.05.2026).

Конфликт интересов: Авторы заявляют об отсутствии конфликта интересов.

Conflict of interest: The authors declare that there is no conflict of interest.

Финансирование: Авторы заявляют об отсутствии внешнего финансирования.

Financing: The research was performed without external funding.