ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
2014 Управление, вычислительная техника и информатика № 2 (27)

УДК 519.872

Ю.И. Рыжиков, А.В. Уланов

РАСЧЁТ ГИПЕРЭКСПОНЕНЦИАЛЬНОЙ СИСТЕМЫ ОБСЛУЖИВАНИЯ $M/H_2/n\text{-}H_2$ С ЗАЯВКАМИ, НЕТЕРПЕЛИВЫМИ В ОЧЕРЕДИ

Представлен метод расчёта многоканальной системы массового обслуживания с простейшим входящим потоком и гиперэкспоненциальными второго порядка (H_2) распределениями времён обслуживания и предельного ожидания в очереди. Получена диаграмма переходов между микросостояниями системы $M/H_2/n\text{-}H_2$, на основе которой составлена программа, реализующая вычисление стационарного распределения числа заявок в системе и вероятности преждевременного ухода заявки из очереди. Результаты верифицированы с помощью имитационных моделей.

Ключевые слова: системы массового обслуживания; нетерпеливые заявки; ограниченное ожидание; численные методы; гиперэкспоненциальное распределение.

Одной из актуальных задач теории очередей является расчёт систем массового обслуживания (СМО) с «нетерпеливыми» заявками, уходящими из очереди при превышении некоторого допустимого (в общем случае случайного) времени ожидания (времени терпения). Обстоятельный обзор современного состояния вопроса приводится, например, в [1]. Приходится констатировать, однако, что подавляющее большинство известных методов либо относится к чисто марковским системам $M/M/n\text{-}M$ [2] (в дополнении к нотации Кендалла через дефис указан тип распределения времени терпения), либо ограничено весьма стеснительными условиями – одноканальное обслуживание и большая загрузка [1].

Из работ по анализу *многоканальных немарковских* систем стоит выделить статью [3], в которой рассматривается диаграмма переходов системы с марковским обслуживанием и H_2-распределением терпения $M/M/n\text{-}H_2$, а также работу [4], в которой предложен метод расчета системы с MAP-входящим потоком, матрично-фазовым (Ph) распределением обслуживания и марковским терпением.

Ниже представлен метод расчета многоканальной системы с простейшим входящим потоком и H_2-распределениями обслуживания и терпения $M/H_2/n\text{-}H_2$.

1. Метод фиктивных фаз

Традиционной технологией анализа многоканальных немарковских систем является метод фиктивных фаз. Расчёт при этом проходит следующие этапы:

– аппроксимация непоказательных распределений распределениями фазового типа;

– построение диаграммы переходов;

– формирование в соответствии с диаграммой матриц интенсивностей переходов;

– составление уравнений баланса переходов, их решение и расчёт стационарных характеристик СМО.

Остановимся подробнее на каждом этапе.

Аппроксимируем непоказательные распределения обслуживания и терпения по методу моментов H_2-распределением. Оно относится к фазовым и позволяет выровнять три начальных момента любого (кроме Эрланга второго порядка) исходного. М.Дж. Кендалл и А. Стьюарт утверждают: «Практически аппроксимация такого рода оказывается очень хорошей, даже если совпадают только три или четыре момента» [5. С. 127]. Многочисленные вычислительные эксперименты [6] в области теории очередей также подтверждают, что трех моментов для проведения расчётов вполне достаточно.

Далее, объединив представленную в [3] диаграмму для $M/M/n\text{-}H_2$ с диаграммой системы без нетерпения $M/H_2/n$, удалось получить схему переходов между состояниями системы $M/H_2/n\text{-}H_2$. Остановимся подробнее на технологии ее построения.

Напомним, что H_2-распределение представляет собой две параллельные фазы с экспоненциальной задержкой в каждой из них. Пусть H_2-обслуживание имеет параметры $\{y_i, \mu_j\}$, где $\{y_i\}$, $i = 1, 2,$ – вероятности выбора фаз, $\{\mu_j\}$, $j = 1, 2,$ – интенсивности показательного распределения времени обслуживания в них. Аналогично зададим параметры H_2-распределения терпения – $\{u_i, \gamma_j\}$. Интенсивность простейшего входящего потока обозначим λ, число каналов – n. Далее рассмотрим состояния системы $M/H_2/n\text{-}H_2$ без очереди. Поскольку в этом случае отсутствуют уходы по нетерпению, диаграмма переходов представлена на рис. 1.

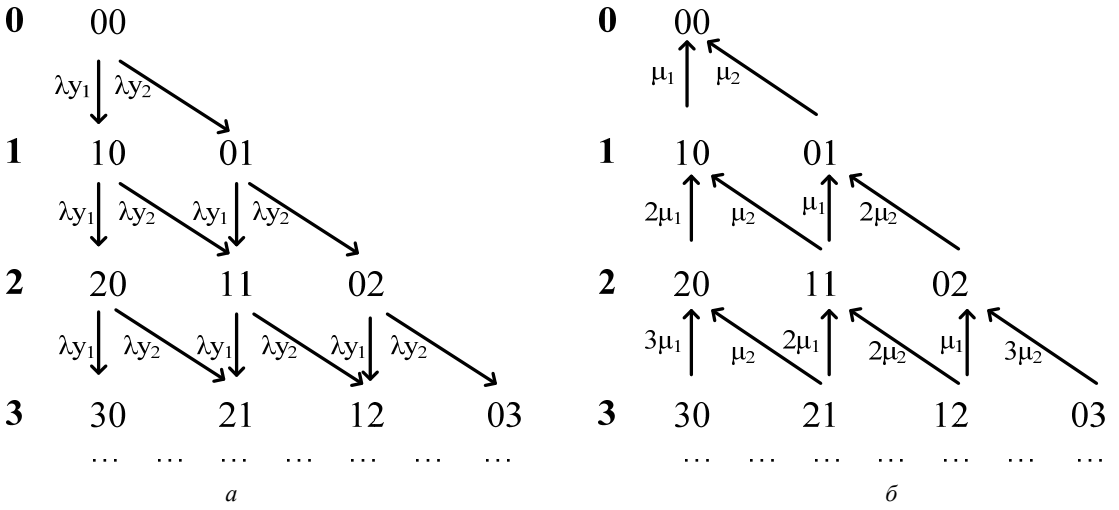

Рис. 1. Диаграммы переходов в «недогруженной» системе $M/H_2/n\text{-}H_2$

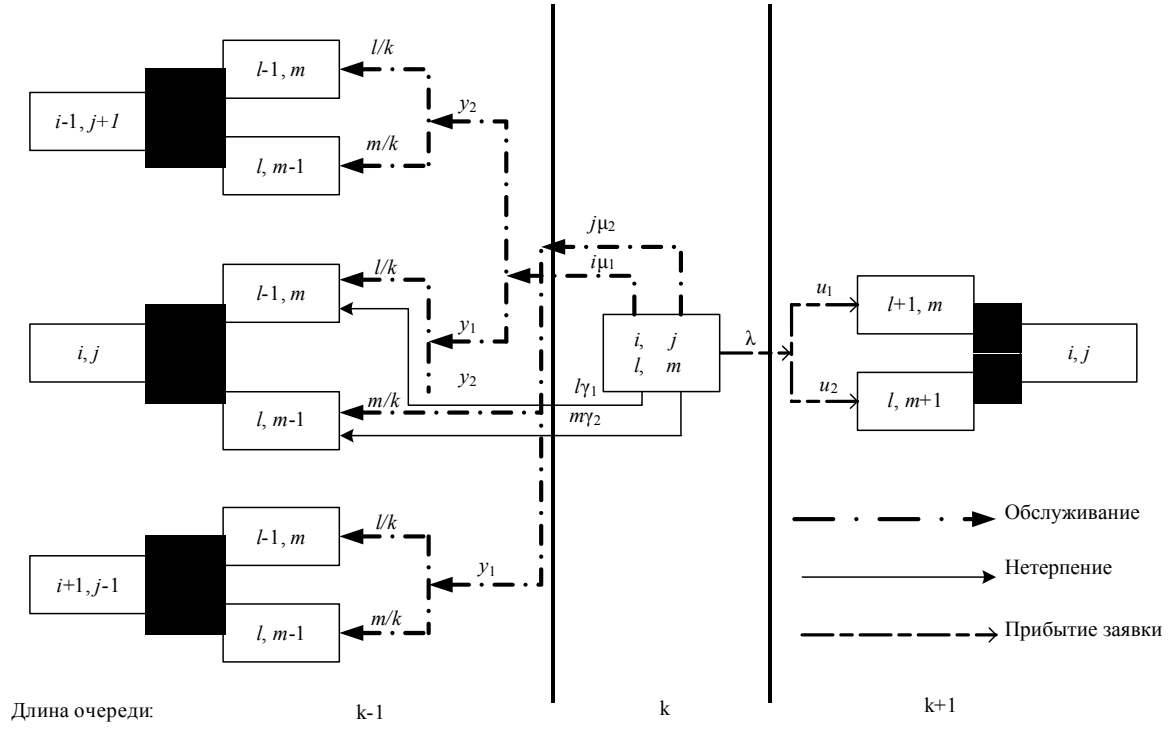

Рис. 2. Схема переходов в «загруженной» системе $M/H_2/n\text{-}H_2$

На рис. 1, *а* показана диаграмма переходов по прибытию заявок, на рис. 1, *б* – по завершению обслуживания. Полужирные столбцы слева от диаграмм – номера ярусов, они соответствуют количеству заявок в системе. Двухразрядные индексы (00, 10, 01, 20 и т.д.) называются ключами микросостояний и указывают расстановку заявок по фазам (типам) H_2-распределения обслуживания. Например, ключ «21» говорит о том, что в двух каналах заявки получают обслуживание первого типа (с показательной интенсивностью μ_1), а в одном – второго типа (с интенсивностью μ_2). У стрелок нанесены интенсивности соответствующих переходов между смежными микросостояниями.

Заявка, пришедшая в занятую систему, становится в очередь. Для нее необходимо зафиксировать фазу H_2-распределения терпения. При этом каждое микросостояние, начиная с (n+1)-го яруса, расщепляется для указания расстановки заявок по фазам H_2-терпения. Например, на n-м ярусе (рис. 1) количество микросостояний равно (n+1), все каналы обслуживания заняты и очереди нет. Прибывшая заявка застаёт систему в одном из микросостояний и в зависимости от типа её терпения попадает в одну из двух фаз. Поскольку терпящие в очереди заявки не влияют на процесс обслуживания, на (n+1)-м ярусе количество микросостояний удваивается, а к их ключам добавляется ключ (l, m), содержащий информацию о расстановке заявок по фазам H_2-терпения.

Изобразить диаграмму переходов с сохранением наглядности не представляется возможным, поэтому представим всевозможные уходы из одного микросостояния с ключом (i, j; l, m), где (i, j) – расстановка заявок по фазам обслуживания, $i+j=n$; (l, m) – по фазам терпения, $l+m=k$, где k – число заявок в очереди (см. рис. 2).

Поясним логику построения схемы. Заявка *прибывает* в систему с интенсивностью λ, застает все каналы занятыми и становится в очередь. В зависимости от типа ее терпения (с вероятностью u_1 она будет первого типа, в вероятностью u_2 – второго) выбирается следующее микросостояние. При уходе заявки *по обслуживанию* для начала выбирается, какой тип обслуживания требуется головной заявке очереди (с вероятностью y_1 – первый тип, в вероятностью y_2 – второй) и далее – какой тип терпения имела обслуженная заявка (вероятности пропорциональны доле заявок соответствующих типов терпения в очереди). При уходе *нетерпеливой* заявки уменьшается количество заявок соответствующего типа терпения. Таким образом, при переходе из одного микросостояния в другое коэффициенты у стрелок последовательно перемножаются.

2. Матрицы переходов и уравнения баланса

На основе диаграммы на рис. 1 и схемы переходов на рис. 2 могут быть получены интенсивности переходов в микросостояния (i, j, l, m), необходимые для записи уравнений баланса вероятностей состояний системы.

Обозначим через S_j множество всех микросостояний системы, когда в системе находится ровно j заявок (что соответствует j-му ярусу), а через σ_j – количество элементов в S_j. Определим матрицы интенсивностей переходов:

$A_j[\sigma_j \times \sigma_{j+1}]$ – в S_j (по прибытию заявок),

$B_j[\sigma_j \times \sigma_{j-1}]$ – в S_{j-1} (по завершению обслуживания или уходу нетерпеливых),

$D_j[\sigma_j \times \sigma_j]$ – ухода из микросостояний j-го яруса (диагональная матрица).

В квадратных скобках здесь и далее указывается размер матриц. Элемент (i, k) любой из этих матриц представляет интенсивность перехода из i-го состояния j-го яруса в k-е состояние смежного (по переходам рассматриваемого типа) яруса.

Поскольку уход нетерпеливых заявок автоматически уменьшает вероятность длинной очереди, количество ярусов диаграммы можно ограничить сравнительно небольшим числом R. Если в результате расчета окажется, что вероятность наличия в системе R заявок велика, необходимо повторить расчет с увеличенным R.

Введём векторы-строки $\pi_j = \{\pi_{j,1}, \pi_{j,2}, ..., \pi_{j,\sigma_j}\}$ вероятностей нахождения СМО в микросостояниях j-го яруса. Теперь можно записать векторно-матричные уравнения баланса переходов между состояниями:

$$\begin{aligned} &\pi_0 D_0 = \pi_1 B_1; \\ &\pi_j D_j = \pi_{j-1} A_{j-1} + \pi_{j+1} B_{j+1}, \quad j = \overline{1, R}. \end{aligned} \tag{1}$$

Система уравнений (1), дополненная условием нормировки, даже при ограничении очереди характеризуется чрезвычайно высокой размерностью, и стандартные методы решения систем линейных алгебраических уравнений применительно к ней оказываются неэффективными. Мы развили применительно к ситуации с нетерпеливыми заявками предложенный в [7] и детально проработанный в [8] итерационный метод решения системы (1).

Вероятность ухода заявки по нетерпению находилась как

$$P_{imp} = \lambda^{-1} \sum_{k=1}^{\infty} \sum_{j=0}^{k} \left(j\gamma_1 + (k-j)\gamma_2 \right) \pi'_{k,j}, \tag{2}$$

где $\pi'_{k,j}$ – вероятность того, что в очереди k заявок, из которых j – первого типа терпения, находится суммированием вероятностей соответствующих микросостояний.

Ценность итерационного метода напрямую зависит от возможности автоматического построения матриц интенсивностей переходов на основе диаграмм. В нашем случае алгоритмы построения матриц относительно очевидны, что следует из рис. 1, 2. Однако на практике программная реализация метода расчета системы $M/H_2/n\text{-}H_2$ на языке Фортран-90 оказалась нетривиальной и весьма поучительной. Исходными данными для программы являются число каналов и наборы моментов распределений обслуживания и терпения для их H_2-аппроксимации.

Ниже представлены результаты расчета системы $M/G/n\text{-}G$ при аппроксимации различных исходных распределений обслуживания и терпения гиперэкспонентами второго порядка.

3. Численные результаты

Были рассчитаны стационарное распределение заявок в системе $M/H_2/n\text{-}H_2$ и вероятность потери заявки по нетерпению – формула (2) – при следующих исходных данных:

– число каналов $n = 3$;

– среднее время обслуживания $b_1 = 1$;

– среднее время терпения $g_1 = 1{,}5$;

– коэффициент загрузки $\rho = 0{,}5$;

– предельное число заявок в системе $R = 11$.

Необходимая для расчёта интенсивность входящего потока находится по формуле $\lambda = \rho n / b_1$. Заметим, что коэффициент загрузки не учитывает уход нетерпеливых заявок, поэтому он теряет определяющую роль и условие $\rho < 1$ для существования стационарного режима в данном случае не обязательно. По трём начальным моментам рассчитывались параметры гиперэкспоненты для аппроксимации следующих распределений терпения и обслуживания: детерминированного D (коэффициент вариации $v = 0$), Эрланга третьего порядка E_3 ($v = 0{,}577$), показательного M ($v = 1$). Также был выполнен расчет при H_2-распределении с вещественными параметрами ($v = 1{,}5$ и $v = 2{,}0$).

В табл. 1 приведены результаты расчёта стационарного распределения $\{p_j\}$, $j = \overline{0, 11}$, числа заявок в системе $M/D/3\text{-}E_3$ численным методом (числ.) и с помощью имитационной модели (ИМ), описание которой имеется в [9]. Заметим, что при аппроксимации D- и E_3-распределений гиперэкспоненциальным параметры последнего принимают комплексные значения, однако на конечный результат это не повлияло.

Относительное отклонение рассчитанных численным методом характеристик от полученных на имитационной модели составляет 10%, что принято считать вполне приемлемым [10]. Заметим, что из-за несовершенства датчиков равномерно распределённых случайных чисел имитационная модель не может считаться идеальным эталоном. К тому же численный метод предполагает сохранение ограниченного количества начальных моментов исходных распределений при аппроксимации.

Таблица 1

Вероятности $\{p_j\}$ состояний системы $M/D/3\text{-}E_3$

j	ИМ	Числ.	j	ИМ	Числ.	j	ИМ	Числ.
0	3,912e–2	4,120e–2	4	1,807e–1	1,833e–1	8	1,534e–2	1,420e–2
1	1,206e–1	1,235e–1	5	1,227e–1	1,195e–1	9	5,926e–3	5,987e–3
2	1,935e–1	1,896e–1	6	7,060e–2	6,966e–2	10	1,919e–3	1,878e–3
3	2,147e–1	2,247e–1	7	3,486e–2	3,329e–2	11	9,029e–4	9,164e–4

Рассмотрим вероятности потери «нетерпеливой» заявки (формула (2)), которая является интегральной характеристикой СМО, при других распределениях обслуживания и терпения (табл. 2) и различных коэффициентах загрузки $\rho=\lambda b_1/n$. В скобках указаны коэффициенты вариации.

Таблица 2

Вероятность ухода заявки по нетерпению при различных коэффициентах вариации и загрузке

Распред. обслуж.	ρ	Распределение терпения							
		D (0)		E_3 (0,577)		M (1)		H_2 (2)	
		ИМ	Числ.	ИМ	Числ.	ИМ	Числ.	ИМ	Числ.
D (0)	0,2	6,84e–3	7,55e–3	8,45e–4	5,22e–4	3,59e–3	3,47e–3	7,27e–3	6,96e–3
	0,7	1,10e–1	1,23e–1	3,95e–2	4,18e–2	8,05e–2	8,04e–2	1,23e–1	1,22e–1
	1,4	3,54e–1	3,61e–1	2,59e–1	2,72e–1	3,09e–1	3,10e–1	3,58e–1	3,56e–1
E_3 (0,578)	0,2	2,43e–2	2,46e–2	9,53–3	9,21e–3	4,00–3	3,98e–3	7,64e–3	7,54e–3
	0,7	1,48e–1	1,56e–1	5,17e–2	5,42e–2	8,96e–2	8,95e–2	1,30e–1	1,29e–1
	1,4	3,70e–1	3,79e–1	2,63e–1	2,74e–1	3,14e–1	3,15e–1	3,66e–1	3,62e–1
M (1)	0,2	3,97e–1	4,04e–1	2,06e–3	1,93e–3	4,79e–3	4,79e–3	8,40e–3	8,35e–3
	0,7	1,95e–1	1,99e–1	6,90e–2	7,18e–2	1,02–1	1,02e–1	1,39–1	1,38e–1
	1,4	3,70e–1	3,79e–1	2,69e–1	2,77e–1	3,20e–1	3,20e–1	3,74e–1	3,71e–1
H_2 (1,5)	0,2	7,40e–2	7,25e–2	3,40e–3	3,31e–3	5,88e–3	5,86e–3	9,23e–3	9,22e–3
	0,7	2,41e–1	2,42e–1	8,91e–2	9,15e–2	1,16–1	1,16e–1	1,48e–1	1,47e–1
	1,4	4,27e–1	4,32e–1	2,79e–1	2,80e–1	3,27e–1	3,25e–1	3,81e–1	3,78e–1

Из табл. 2 следует, что при фиксированном распределении терпения с ростом коэффициента вариации обслуживания вероятность ухода заявки из очереди возрастает. Также с ростом коэффициента загрузки прослеживается снижение влияния вида распределений обслуживания и терпения на вероятность ухода из очереди.

Поскольку в реальных организационно-технических системах уходы нетерпеливых заявок, как правило, приводят к негативным последствиям, интересной является задача исследования дробления производительности (т.е. замены одного канала обслуживания несколькими с той же суммарной производительностью) на вероятность потери заявки. С этой целью была рассчитана системы $M/D/5\text{-}E_3$ при коэффициенте загрузки $\rho = 0{,}9$. Оказалось, что при замене одного канала двумя исследуемая вероятность уменьшилась в 1,7 раза, тремя – в 2,4 раза, четырьмя – в 3 раза.

Заключение

Предложена диаграмма переходов для системы $M/H_2/n\text{-}H_2$, на основе диаграммы составлена Фортран-программа генерации матриц интенсивностей переходов и расчёта стационарного распределения числа заявок и вероятности ухода нетерпеливой заявки. Хорошее согласие результатов, полученных численным методом и с помощью имитационной модели, подтверждает правильность теоретических выкладок и реализующих их программ. Показаны возможности H_2-аппроксимации различных исходных распределений обслуживания и терпения; недопустимость показательной аппроксима-

ции времени обслуживания и терпения в том случае, когда коэффициенты вариации распределений заметно отличаются от единицы и коэффициент загрузки системы невысок; нетривиальный эффект снижения доли нетерпений от дробления производительности. В настоящее время ведется разработка метода расчета временных характеристик.

Предложенный метод может быть применен при проектировании колл-центров, служб быстрого реагирования, систем обработки устаревающей информации и т.д.

ЛИТЕРАТУРА

1. *Hoshi K., Iijima S., Takahashi Y., Komatsu N.* TrafficPerfomance for a Time-Out Scheme Communication System // Proc. of International Conference ICUMT 2009. St. Petersburg, 2009. P. 1–6.
2. *Бочаров П.П., Печинкин А.В.* Теория массового обслуживания : учеб. М. : РУДН им. П. Лумумбы, 1995. 529 с.
3. *Roubos A., Jouini O.* Call Centers with Hyperexponential Patience Modeling // International Journal of Production Economics. 2013. V. 141. P. 307–315.
4. *Дудин С.А., Дудина О.С.* Модель функционирования колл-центра как система MAP/PH/N/R-N с нетерпеливыми запросами // Проблемы передачи информации. 2011. № 47. С. 68–83.
5. *Кендалл М. Дж., Стьюарт А.* Теория распределений : пер. с англ. М. : Наука, 1966. 587 с.
6. *Рыжиков Ю.И., Уланов А.В.* Опыт расчёта сложных систем массового обслуживания // Информационно-управляющие системы. 2009. № 2. С. 56–62.
7. *Takahashi Y., Takami Y.A.* Numerical Method for the Steady-State Probabilities of a GI/G/c Queuing System in General Class // Journal of the Operations Research Society of Japan. 1976. V. 19, No. 2. P. 147–155.
8. *Рыжиков Ю.И., Хомоненко А.Д.* Итеративный метод расчета многоканальных систем с произвольным распределением времени обслуживания // Проблемы управления и теории информации. 1980. № 3. С. 203–213.
9. *Рыжиков Ю.И., Уланов А.В.* Имитационное моделирование систем с «нетерпеливыми» заявками // Имитационное моделирование. Теория и практика : тр. VI Всерос. конф. Казань, 2013. С. 339–342.
10. *Башарин Г.П., Бочаров П.П., Коган Я.А.* Анализ очередей в вычислительных системах. М. : Наука ; Физматгиз, 1989. 336 с.

Рыжиков Юрий Иванович, д-р техн. наук, профессор. E-mail: ryzhbox@yandex.ru
Военно-космическая академия им. А.Ф. Можайского,
Санкт-Петербургский институт информатики и автоматизации РАН
Уланов Александр Викторович. E-mail: ulanov246@rambler.ru
Военно-космическая академия им. А.Ф. Можайского

Поступила в редакцию 12 марта 2014 г.

Ryzhikov Yury I., Ulanov Alexader V. (Mozhaisky Military Space Academy, St. Petersburg Institute of Informatics and Automation of RAS, St. Petersburg, Russian Federation).

Calculation of the M/H_2/n-H_2 queuing system with impatient customers.

Keywords: queuing systems; impatient customers; restricted waiting time; hyperexponential distribution; numerical methods.

The paper deals with queuing systems with impatient customers. The known methods for calculating these systems refers to either the purely Markov model M/M/n-M (in addition to the Kendall notation the hyphen indicates the type of a patience distribution law) or such methods have very restricted conditions, namely, the single-channel service and great loading, the exponential distribution service or patience. We propose a method for calculating the multi-channel QS with the second order hyperexponential distributions both service time and patience M/H_2/n-H_2.

The method relies on the method of fictitious phases. The calculation consists of the following stages:

- approximation by phase type distributions;
- construction of the diagram of transitions;
- forming the matrixes of transition intensities in accordance with the diagram;
- deriving the balance equations and their solutions and calculating the QS stationary characteristics.

The second order hyperexponential distribution (H_2) and three initial moments of the origin distribution have been used. Then we have combined the proposed Roubos & Jouini diagram for M/M/n-H_2 queue with the well known diagram of system without impatient customers M/H_2/n. The balance equations were solved by the iterative Takahashi & Takami's method.

The probability of impatience queue is defined as

$$P_{imp} = \lambda^{-1} \sum_{k=1}^{\infty} \sum_{j=0}^{k} \left(j\gamma_1 + (k-j)\gamma_2 \right) \pi'_{k,j},$$

where λ is the intensity of the incoming flow; γ_1 and γ_2 are the intensities for the exponential phase of the H_2-patience; $\pi'_{k,j}$ is the probability that the queue length is k, of which j are the first type of patience.

It is shown the influence of the variation coefficients of service and patience. Also, the dependence of the probability of removing for impatience queue on the load coefficient is investigated. We concluded using Markov models for the calculation of systems, in which the service and patience distributions are markedly different from the exponential ones, gives incorrect results. The effect of fragmentation performance and its influence on the probability of impatience is made known.

The results can be applied in the design of call centers, emergency services, systems of obsolescent information processing, etc.

REFERENCES

1. Hoshi K., Iijima S., Takahashi Y., Komatsu N. Traffic Performance for a Time-Out Scheme Communication System. *Proc. of International Conference* ICUMT. St. Petersburg, 2009, pp. 1-6.
2. Bocharov P.P., Pechinkin A.V. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: Russian University of Peoples' Friendship Publ., 1995. 529 p.
3. Roubos A., Jouini O. Call centers with hyperexponential patience modeling. *International Journal of Production Economics*, 2013, vol. 141, pp. 307-315. DOI: 10.1016/j.ijpe.2012.08.011
4. Dudin S.A., Dudina O.S. Call center operation model as a MAP/PH/N/R-N system with impatient customers. *Problemy peredachi informatsii – Problems of Information Transmission*, 2011, vol. 47, no. 4, pp. 68-83. (In Russian).
5. Kendall M., Stuart A. *The advanced theory of statistics. Distribution theory*. Translated from English. Mosow: Nauka Publ., 1966. 587 p.
6. Ryzhikov Yu.I., Ulanov A.V. Experience of the Complex Queuing Systems Calculation. *Informatsionno-upravlyayushchie sistemy – Information and Control System*, 2009, vol. 39, no. 2, pp. 56-62. (in Russian).
7. Takahashi Y., Takami Y.A. Numerical Method for the Steady-State Probabilities of a GI/G/c Queuing System in General Class. *Journal of the Operations Research Society of Japan*, 1976, vol. 19, no. 2, pp. 147-155.
8. Ryzhikov Yu.I., Khomonenko A.D. Iterativnyy metod rascheta mnogokanal'nykh sistem s proizvol'nym raspredeleniem vremeni obsluzhivaniya [Iterative methods of multichannel queuing system calculation with arbitrary service time distribution]. *Problemy upravleniya i teorii informatsii*, 1980, vol. 3, pp. 203-213.
9. Ryzhikov Yu.I., Ulanov A.V. [Simulation of queuing systems with impatient customers]. *Imitatsionnoe modelirovanie. Teoriya i praktika : tr. VI Vseros. konf* [Proc. of 6th Russian conference "Simulation. Theory and Practice"]. Kazan, 2013, pp. 339-342. (In Russian).
10. Basharin G.P., Bocharov P.P., Kogan Ya.A. *Analiz ocheredey v vychislitel'nykh sistemakh* [Queueing analysis in computing systems]. Moscow: Nauka Publ., 1989. 336 p.