

УДК 519.872

РАСЧЕТ СИСТЕМ ОБСЛУЖИВАНИЯ С ГРУППОВЫМ ПОСТУПЛЕНИЕМ ЗАЯВОК

Ю. И. Рыжиков,

доктор техн. наук, профессор

Военно-космическая академия им. А. Ф. Можайского

Предлагается способ расчета распределения времени ожидания заявок в одно- и n -канальной системе при простейшем потоке пачек заявок случайного объема с ограниченным размахом, а также распределения числа заявок в системе. Точность расчетов иллюстрируется сопоставлением с результатами имитационного моделирования.

A method is proposed to compute the number-in-the-system and waiting time distributions for a one- and multichannel queuing systems with a Poissonian flow of the random (finite volume) demands. The results are compared with imitation ones.

Введение

Работа многих систем обслуживания связана с обслуживанием неординарного потока заявок, поступающих пачками случайного, вообще говоря, объема. Такие ситуации порождаются прежде всего расщеплением первичных заявок, например при передаче данных в сетях с коммутацией пакетов, при заявках на комплекты изделий. Другими примерами могут служить прибытие групп пассажиров (семейных, экскурсионных, командированных) в пункт отправления, медицинская помощь жертвам катастроф и террористических актов, прорыв группой самолетов или ракет зоны ПВО и т. п.

Групповое поступление заявок существенно ухудшает показатели системы в сравнении с ординарным потоком той же средней интенсивности. Для оценки работы системы и последующего принятия организационных мер (управления потоком, увеличения быстродействия обслуживающих устройств или их числа) необходимо уметь рассчитывать следующие распределения:

- времени ожидания начала обслуживания пачки;
- дополнительных задержек внутри пачки и средней задержки заявок пачки;
- задержки пачки в целом;
- числа заявок, находящихся в системе.

Поставленная задача уже довольно давно обсуждается в литературе [8–12, 14]. Однако:

- все эти результаты относятся только к *одноканальным* системам;

- как правило, они имеют частный характер [14], а в ряде случаев [11] ошибочны;

- верификация результатов и данные о численной реализации отсутствуют.

В данной статье последовательно рассматриваются подходы к расчету систем обслуживания групповых заявок, решающие упомянутые проблемы. Объем пачек предполагается ограниченным, а входящий поток пачек — простейшим.

Имитационное моделирование обслуживания групповых заявок

Имитационное моделирование (ИМ) в рассматриваемых ситуациях в связи с отсутствием альтернативных (и даже просто апробированных) методов расчета является практически единственным способом получения эталонных результатов. В инструментальных системах ИМ, например в GPSS World [1, 4, 7], предусмотрены ситуации с расщеплением и последующей сборкой заявок, означающей окончание обслуживания пачки. Однако ответа на все поставленные вопросы любая система со встроенным интерпретатором не дает. В GPSS World нельзя, к примеру, непосредственно получить моменты распределения времени ожидания пачки порядка выше второго, распределение времени пребывания в системе в зависимости от номера заявки в пачке, распределение числа заявок в системе. Встроенный в упомянутую систему язык PLUS все же является усеченным подмножеством универсальных алгоритмических языков, ориентированных на численные приложения, и по

возможностям и удобству заметно уступает Фортрану (см. раздел по оценке GPSS World [4]). Методика и приемы построения имитационных моделей на Фортране обсуждались в работах [4, 5]. Для решения вышеперечисленных задач имитационная модель должна иметь следующую специфику.

- По прибытии заявки случайным образом (в соответствии с заданным распределением) формируется объем пачки. Соответственно определяются количества d — принимаемых в каналы заявок и l — направляемых в очередь. Для первых свободные каналы определяются в цикле просмотра моментов освобождения с выходом из него после выявления d каналов. Свободные каналы занимают, для них формируются моменты освобождения; идет подсчет количества принятых заявок.

- Для заявок, направляемых в очередь, определяется возможность их приема (очередь ограничена). Для всех принимаемых запоминается момент их прибытия. Кроме того, для головной заявки пачки в ее паспорте фиксируется объем пачки, а для последующих — ее номер в пачке. В случае приема на обслуживание хотя бы одной заявки пачки в паспортах остальных запоминается время начала обслуживания пачки.

- При завершении обслуживания и наличии заявок в очереди прежде всего выясняется статус головной заявки очереди. Если она — первая в своей пачке, то для остальных заявок пачки фиксируется момент начала ее обслуживания. Далее накапливаются степени истекшего времени ожидания. Если заявка — не первая, копятся степени дополнительной задержки по отношению к первой заявке пачки. Накопление идет отдельно по номерам заявок в пачке.

- Для построения распределения числа заявок в очереди фиксируются моменты его изменения, связанные с постановкой в очередь или началом обслуживания первой заявки пачки (последнее необходимо для расчета распределения ожидания начала обслуживания). Соответственно, имеется счетчик текущего числа пачек, и при его изменении к соответствующим накопительным ячейкам добавляется интервал неизменности.

- Время ожидания начала обслуживания пачки определяется как разность между временем выбора на обслуживание первой заявки пачки и моментом ее прибытия в систему. Суммирование интервалов и количества реализаций идет в отдельных (по номерам заявок в пачке) счетчиках, причем количество реализаций увеличивается и при немедленном приеме заявки на обслуживание.

Верификация модели проводилась сопоставлением результатов ее работы с полученным аналитическим расчетом системы $M^X/G/1$ (см. далее).

Распределение числа заявок, прибывших за время обслуживания

Рассмотрим пуассоновский поток пачек требований, каждая из которых имеет одно и то же рас-

пределение $\{f_i\}$ числа заявок. Тогда распределение $\{f_i^{n*}\}$ суммарного числа заявок в n пачках будет n -кратной сверткой распределения $\{f_i^{1*}\} = \{f_i\}$.

Для дальнейшего нам необходимы вероятности прибытия ровно i заявок за случайный интервал времени между смежными обслуживаниями, имеющий распределение $B(t)$:

$$h_i = \int_0^\infty e^{-\lambda t} \sum_{n=0}^\infty \frac{(\lambda t)^n}{n!} f_i^{n*} dB(t) = \sum_{n=0}^\infty f_i^{n*} \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^n}{n!} dB(t). \quad (1)$$

Интегралы следует вычислить предварительно согласно рекомендациям разд. 3.6.3 работы [3], а свертки $\{f_i\}$ (для конечного размаха) последовательно получать численно и выполнять подсуммирование отдельно для каждого i . Нужно иметь в виду происходящее на каждом шаге свертки удлинение массива вероятностей $\{f_i^{n*}\}$.

В табл. 1 приводится сопоставление расчета по вышеописанной схеме и моделирования (1 млн испытаний) числа заявок обобщенного пуассоновского потока.

Объем пачки предполагался равновероятным в диапазоне 1÷6, интервал времени — равномерно распределенным на интервале $[0, 10]$.

Расчет одноканальной системы

Относительно стандартной системы $M/G/1$ известно (см., например, работу [3]), что стационарные вероятности наличия в системе ровно k заявок вычисляются как

$$p_0 = 1 - \lambda b_1; \quad (2)$$

$$p_k = \left(p_{k-1} - p_0 q_{k-1} - \sum_{j=1}^{k-1} p_j q_{k-j} \right) / q_0, \quad k = 1, 2, \dots,$$

где b_1 — среднее время обслуживания заявки, а

$$q_i = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^i}{i!} dB(t)$$

есть вероятность прибытия ровно i заявок за случайное время обслуживания с распределением $B(t)$. Преобразование Лапласа—Стилтьеса (ПЛС) распределения времени ожидания получается согласно известной формуле Полячека—Хинчина (ФПХ)

$$\omega(s) = \frac{p_0}{1 - \frac{\lambda}{s} [1 - \beta(s)]}, \quad (3)$$

где $\beta(s)$ есть ПЛС от плотности $B(t)$.

■ Таблица 1. Обобщенный пуассоновский поток

<i>n</i>	Расчет	Модель	<i>n</i>	Расчет	Модель
0	.43143e+00	.43237e+00	20	.16770e-02	.17110e-02
1	.49974e-01	.49654e-01	21	.12536e-02	.12640e-02
2	.54396e-01	.53787e-01	22	.91291e-03	.88300e-03
3	.59148e-01	.59052e-01	23	.64986e-03	.67300e-03
4	.64250e-01	.64103e-01	24	.45644e-03	.50100e-03
5	.69724e-01	.69234e-01	25	.32116e-03	.32400e-03
6	.75594e-01	.75437e-01	26	.22857e-03	.24000e-03
7	.31910e-01	.32270e-01	27	.15886e-03	.18600e-03
8	.29800e-01	.30080e-01	28	.10808e-03	.12900e-03
9	.27174e-01	.27388e-01	29	.72346e-04	.91000e-04
10	.23976e-01	.24203e-01	30	.48001e-04	.55000e-04
11	.20146e-01	.20224e-01	31	.31759e-04	.46000e-04
12	.15622e-01	.15626e-01	32	.20907e-04	.30000e-04
13	.10335e-01	.10292e-01	33	.13506e-04	.19000e-04
14	.86349e-02	.86930e-02	34	.86336e-05	.12000e-04
15	.70107e-02	.68830e-02	35	.55360e-05	.80000e-05
16	.55038e-02	.53480e-02	36	.36252e-05	.60000e-05
17	.41612e-02	.40850e-02	37	.24694e-05	.30000e-05
18	.30353e-02	.29600e-02	38	.17744e-05	.10000e-05
19	.21854e-02	.21270e-02	39	.13579e-05	.00000e+00

Избавившись от знаменателя в правой части формулы (3), разложим входящие в нее преобразования Лапласа по степеням *s* и приравняем коэффициенты при одинаковых степенях в правой и левой частях. Из этих равенств следуют формула для первого момента распределения длительности ожидания пачки и рекуррентные формулы для высших моментов:

$$w_1 = \frac{\lambda b_2}{2(1-\lambda b_1)};$$

$$w_k = \frac{\lambda}{1-\lambda b_1} \sum_{j=0}^{k-1} \frac{k!}{j!(k+1-j)!} b_{k+1-j} w_j, \quad k = 2, 3, \dots, \quad (4)$$

при этом $w_0 = 1$.

В одноканальной системе заявки пачки обслуживаются строго последовательно, а ПЛС времени обработки каждой выражается через ПЛС трудоемкости отдельной заявки $\beta(s)$ формулой

$$\gamma(s) = \sum_{m=1}^M f_m [\beta(s)]^m = F(\beta(s)), \quad (5)$$

где $F(\cdot)$ — производящая функция объема пачки. Моменты $\{g_j\}$ этой трудоемкости можно получить многократным численным дифференцированием

таблицы $\gamma(s)$ в окрестности нуля с последующей сменой знака у нечетных производных или численной сверткой в моментах на основе символического равенства $c^k = (a+b)^k$ с переносом показателей степеней в индексы. Обе эти технологии при разумном выборе шага построения и длины таблицы дают практически совпадающие результаты и могут применяться при решении других задач. Тогда моменты распределения времени ожидания начала обслуживания пачки могут быть найдены по формулам (4) с заменой $\{b_j\}$ на $\{g_j\}$. Вероятность свободного состояния системы здесь вычисляется как

$$p_0 = 1 - \lambda \bar{f} b_1, \quad (6)$$

где \bar{f} — средний объем пачки. Последующие вероятности распределения числа пачек вычисляются по формулам, аналогичным (2), с вычислением $\{q_i\}$ для определяемого (5) распределения трудоемкости пачки.

Распределение суммарного количества заявок в системе с групповым потоком следует рассчитывать согласно системе (2) с заменой первой формулы на (6) и вероятностей $\{q_i\}$ на вычисляемые согласно (1) вероятности $\{h_i\}$.

Приведем сравнительные результаты моделирования (1 млн пачек равновероятного объема от 1 до 6) и расчета системы с равномерной на [0, 10]

■ Таблица 2. Распределение числа пачек в очереди по $M/G/1$

j	Расчет	Модель	j	Расчет	Модель
0	.40341e+00	.40224e+00	10	.85699e-02	.88636e-02
1	.16340e+00	.16312e+00	11	.61429e-02	.63665e-02
2	.12174e+00	.12099e+00	12	.44032e-02	.44870e-02
3	.88167e-01	.88313e-01	13	.31563e-02	.31766e-02
4	.63288e-01	.63167e-01	14	.22624e-02	.21522e-02
5	.45342e-01	.45830e-01	15	.16217e-02	.16070e-02
6	.32483e-01	.32488e-01	16	.11625e-02	.11304e-02
7	.23276e-01	.23682e-01	17	.83326e-03	.78494e-03
8	.16681e-01	.17140e-01	18	.59728e-03	.60351e-03
9	.11953e-01	.12123e-01	19	–	.50317e-03

■ Таблица 3. Моменты распределения ожидания для пачки

Способ расчета	Порядок момента		
	1	2	3
По ФПХ	.46594e+02	.51851e+04	.86156e+06
Через МФАСТ	.46594e+02	.51851e+04	.86156e+06
В модели	.46938e+02	.52437e+04	.86378e+06
В модели через МФАСТ	.46943e+02	.52475e+04	.86595e+06

длительностью обслуживания заявки для коэффициента загрузки 0,8 (табл. 2).

Кроме того, были получены моменты длительности ожидания для пачки (табл. 3).

Здесь ФПХ подразумевает расчет высших моментов по рекуррентным формулам (4), а МФАСТ — через факториальные моменты $\{m_{[k]}\}$ длины очереди пачек согласно формуле Брюмелля [3]

$$w_k = m_{[k]} / \lambda^k.$$

■ Таблица 4. Распределение числа заявок в системе $M^X/G/1$

j	Расчет	Модель	j	Расчет	Модель
0	.20025e+00	.19985e+00	21	.11847e-01	.11988e-01
1	.49231e-01	.49081e-01	22	.10809e-01	.10988e-01
2	.52554e-01	.52130e-01	23	.98625e-02	.99888e-02
3	.54630e-01	.54487e-01	24	.89987e-02	.91151e-02
4	.55031e-01	.54938e-01	25	.82106e-02	.83720e-02
5	.53278e-01	.52955e-01	26	.74914e-02	.76478e-02
6	.48852e-01	.48722e-01	27	.68353e-02	.69639e-02
7	.41197e-01	.41141e-01	28	.62366e-02	.63042e-02
8	.38521e-01	.38470e-01	29	.56904e-02	.57371e-02
9	.35595e-01	.35358e-01	30	.51920e-02	.52330e-02
10	.32594e-01	.32395e-01	31	.47372e-02	.47071e-02
11	.29685e-01	.29406e-01	32	.43223e-02	.43543e-02
12	.27006e-01	.26686e-01	33	.39438e-02	.39469e-02
13	.24627e-01	.24475e-01	34	.35984e-02	.36217e-02
14	.22507e-01	.22442e-01	35	.32832e-02	.33337e-02
15	.20541e-01	.20569e-01	36	.29956e-02	.30517e-02
16	.18737e-01	.18869e-01	37	.27333e-02	.28815e-02
17	.17093e-01	.17107e-01	38	.24939e-02	.26155e-02
18	.15595e-01	.15729e-01	39	.22755e-02	.24088e-02
19	.14231e-01	.14347e-01	40	.20762e-02	.22400e-02
20	.12985e-01	.13017e-01	41	.18944e-02	.20520e-02

Распределение числа заявок в системе (табл. 4), как отмечалось выше, можно получить по алгоритму для стандартной системы $M/G/1$ после замены $\{q_j\}$ на их аналоги для обобщенного пуассоновского потока из табл. 1.

Обращает на себя внимание вызванное группировкой заявок в пачки сильнейшее затягивание «хвостов» распределения.

Поскольку *дополнительная* задержка i -й заявки пачки имеет ПЛС $\varphi_i(s) = \beta^{i-1}(s)$, для пачки в целом ПЛС задержки

$$\varphi_e(s) = \sum_{m=1}^M f_m \beta^{m-1}(s).$$

Для средней задержки произвольной заявки пачки имеем

$$\bar{\varphi}(s) = \sum_{m=1}^M \frac{f_m}{m} \sum_{i=1}^m \beta^{i-1}(s) = \frac{1}{1-\beta(s)} \sum_{m=1}^M \frac{f_m}{m} [1-\beta^m(s)].$$

Распределение полного времени пребывания заявки в системе получается сверткой распределений ожидания пачки, дополнительной задержки в пачке и чистой длительности обслуживания.

Многоканальная система

Прежде всего отметим, что наиболее удобным и универсальным методом расчета многоканальных систем с произвольным распределением обслуживания является аппроксимация последнего гиперэкспоненциальным H_2 . Такая аппроксимация позволяет сохранить три момента исходного распределения, что можно считать необходимым и достаточным.

Ниже обсуждается модель с ординарным потоком заявок, на основе которой далее предлагается метод для «групповой» задачи.

Диаграммы переходов и расчет $M/H_2/n$

Пусть вероятность того, что длительность обслуживания превышает t (она же ДФР — дополнительная функция распределения):

$$\bar{B}(t) = \sum_{i=1}^2 y_i e^{-\mu_i t}.$$

Тогда задачу можно рассматривать как процесс обслуживания заявок двух типов, причем тип заявки назначается с вероятностями $\{y_i\}$ в момент выбора ее на обслуживание. Характеризуя состояние системы полным числом заявок в ней (номер яруса) и распределением обслуживаемых заявок по типам, получаем диаграммы переходов по прибытию заявок и завершению обслуживания — для трехканальной системы (рис. 1 и 2 соответственно).

На рис. 2 при $j > n$ поток обслуживания заявок i -го типа равен $q_i \mu_i$, где q_i — содержимое i -й позиции кода микросостояния. При наличии очереди завершения обслуживания в зависимости от типа

выбранной из очереди заявки с вероятностями $\{y_i\}$ приводит в одно из двух микросостояний вышележащего яруса.

На основе диаграммы переходов могут быть получены матрицы интенсивностей переходов:

A_j — с j -го яруса на $(j+1)$ -й по прибытию заявки,

B_j — с j -го на $(j-1)$ -й по завершению обслуживания,

D_j — ухода из микросостояний j -го яруса (диагональные).

В данной модели нет переходов в пределах одного яруса, так что фигурирующие в общем случае

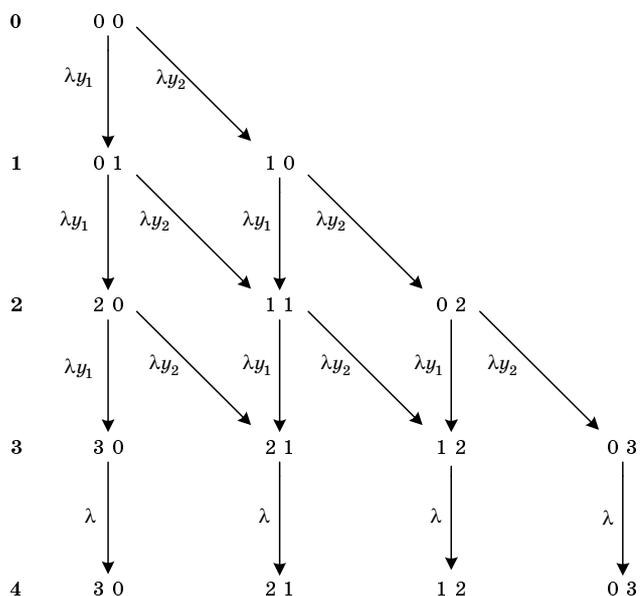


Рис. 1. Переходы по прибытию заявки в системе $M/H_2/3$

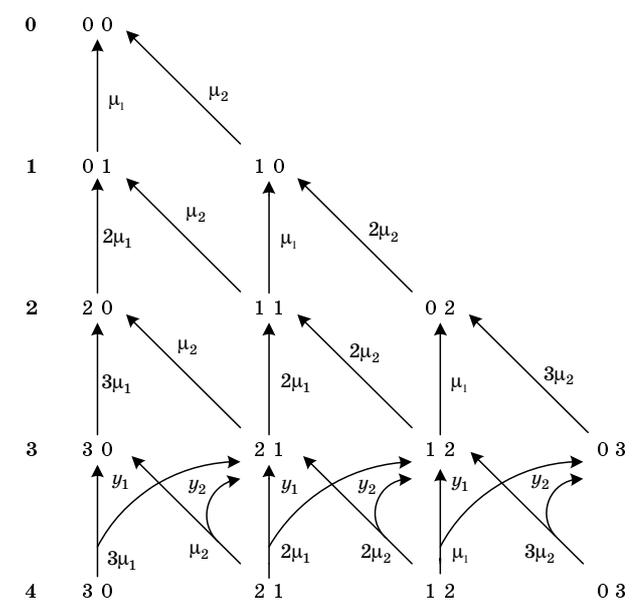


Рис. 2. Переходы по обслуживанию заявки в системе $M/H_2/3$

[3] матрицы $\{C_j\}$ здесь отсутствуют. Микросостояния яруса-источника соответствуют строкам, а яруса-приемника – столбцам матриц. Любой ненулевой элемент матрицы равен метке на представляющей переход стрелке. Элементы матриц $\{B_j\}$ для четвертого и последующих ярусов получаются умножением исходной («корневой») интенсивности на y_j , соответствующий состоянию-приемнику. Поскольку для $j \geq n$ матрицы $A_j = \lambda I$, $D_j = D_n$, а для $j \geq n + 1$ $B_j = B_{n+1}$, запоминать нужно сравнительно небольшое число матриц.

С помощью этих матриц можно записать векторно-матричные уравнения баланса переходов между ярусами. Решение упомянутых уравнений можно получить методом матрично-геометрической прогрессии (см. [3]) или восходящим к Такахаши и Таками итерационным методом [2, 13]. Последний далее развивается нами применительно к потоку групповых заявок.

Распределение количества пачек и моменты распределения ожидания

Распределение количества находящихся в системе $M^X/H_2/n$ пачек заявок заманчиво рассчитывать аналогично одноканальному случаю, т. е. применяя H_2 -аппроксимацию к распределению трудоемкости пачки. Далее моменты распределения ожидания пачки могут быть найдены по распределению числа пачек в очереди. Имитационные эксперименты подтвердили хорошее согласие полученных результатов с непосредственным определением ожидания головной заявки пачки.

Этот подход неявно предполагает, что все заявки каждой пачки последовательно обслуживаются в одном и том же канале. Тем самым исключается разделение пачки между несколькими каналами в недозгруженной системе и снижается реальная производительность последней. Можно предполагать, что возникающая погрешность будет возрастать при увеличении числа каналов и уменьшении номинального коэффициента загрузки системы:

$$\rho = \lambda \bar{f} b_1 / n. \quad (7)$$

Упомянувшиеся ожидаемые тенденции иллюстрируются сопоставлением результатов расчета (Р) и имитации (И) в табл. 5.

Таким образом, описанный подход вполне может найти практическое применение, но в ограниченном диапазоне условий.

Дополнительные задержки в пачках

Эти задержки могут быть рассчитаны аналогично одноканальному случаю с заменой распределения времени на распределение интервалов между выбираемыми из очереди заявками. ДФР этого распределения можно [6, с. 253] аппроксимировать формулой

$$\bar{B}_n(t) = \left[\bar{B}^*(t) \right]^{n-1} \bar{B}(t), \quad (8)$$

которая имеет прозрачный вероятностный смысл: для одного из каналов (только что завершившего обслуживание) берется полное распределение длительности обслуживания, а для прочих — помеченное звездочкой остаточное (случайная модификация полного).

Возможность применения формулы (8) была проверена на имитационной модели, предварительно откалиброванной по задаче с показательным распределением длительности обслуживания (в этом случае остаточное распределение совпадает с исходным и $\bar{B}_n(t) = e^{-n\mu t}$). Рабочее моделирование выполнялось для распределений длительности обслуживания Эрланга 3-го порядка (E_3), коэффициент вариации $v = 0,577$, и гиперэкспоненциального H_2 , $v = 1,36$, при одинаковой средней длительности $b_1 = 5$. Результаты расчета моментов $B_n(t)$ на основании формулы Саати (С) и посредством модели (М) представлены в табл. 6.

Анализ таблицы с учетом общеизвестного факта быстрого роста относительных погрешностей моделирования при увеличении порядка вычисляемых моментов указывает на допустимость практического использования формулы (8).

Еще одним вариантом описания потока обслуживаний в полностью загруженной системе является суммирование потоков от n каналов по методике, описанной в работе [3, разд. 3.6].

Наконец, для решения задачи можно воспользоваться распределением Вейбулла. Здесь ДФР имеет вид $B(t) = \exp(-t^k / T)$, а вероятность того, что обслуживание не завершится ни в одном из n каналов:

$$\bar{B}_n(t) = \exp(-nt^k / T), \quad (9)$$

т. е. описывается тем же распределением с заменой T на T/n . Подставляя это значение в формулы для моментов распределения Вейбулла, убеждаем-

■ Таблица 5. Моменты распределения ожидания пачек

Моменты	$\rho = 0,8$				$\rho = 0,6$			
	$n = 3$		$n = 5$		$n = 3$		$n = 5$	
	И	Р	И	Р	И	Р	И	Р
w_1	.145e+2	.128e+2	.787e+1	.667e+1	.477e+1	.362e+1	.244e+1	.149e+1
w_2	.550e+3	.483e+3	.170e+3	.152e+3	.838e+2	.674e+2	.265e+2	.173e+2
w_3	.307e+5	.268e+5	.531e+4	.510e+4	.217e+4	.180e+4	.428e+3	.284e+3

■ Таблица 6. Моменты интервалов между обслуживаниями в n -канальной системе

Тип $B(t)$	Порядок момента	Число каналов n					
		2		3		4	
		M	C	M	C	M	C
E_3	1	2.50	2.50	1.62	1.67	1.22	1.25
	2	9.48	9.46	4.28	4.46	2.50	2.60
	3	4.64e1	4.57e1	1.47e1	1.57e1	6.83	7.21
	4	2.75e2	2.64e2	6.08e1	6.66e1	2.30e1	2.45e1
	5	1.90e3	1.77e3	2.90e2	3.28e2	9.13e1	9.72e1
	6	1.49e4	1.36e4	1.55e3	1.83e3	4.20e2	4.36e2
H_2	1	2.62	2.45	1.68	1.59	1.27	1.16
	2	1.70e1	1.58e1	6.48	6.57	3.62	3.48
	3	2.05e2	1.75e2	4.36e1	4.54e1	1.75e1	1.74e1
	4	3.95e3	2.94e3	4.41e2	4.55e2	1.25e2	1.25e2
	5	1.09e5	6.94e4	5.88e3	6.18e3	1.17e3	1.19e3
	6	3.80e6	2.13e6	9.37e4	1.09e5	1.30e4	1.43e4

ся, что его моменты порядка m получаются делением исходных на n^m/k . Итак, здесь надлежит:

— найти параметры аппроксимации распределением Вейбулла остаточного распределения длительности обслуживания по двум моментам;

— вышеупомянутым пересчетом (с заменой n на $n-1$) получить три момента распределения (9);

— аппроксимировать это распределение гиперэкспоненциальным;

— найти такую же аппроксимацию для исходного распределения;

— выполнить суммирование двух потоков с найденными распределениями интервалов между заявками.

Распределение числа заявок в системе $M^X/H_2/n$

Запишем условия баланса для векторов-строк $\{\gamma_j\}$ вероятностей микросостояний системы с учетом прибытия заявок пачками случайного объема не свыше M :

$$\gamma_j \mathbf{D}_j = \lambda \sum_{m=1}^{\widehat{M}} f_m \gamma_{j-m} \prod_{i=0}^{m-1} \mathbf{A}_{j-m+i} + \gamma_{j+1} \mathbf{B}_{j+1}. \quad (10)$$

Здесь все матрицы $\{\mathbf{A}_j\}$ разделены на интенсивность λ потока пачек. Каждое слагаемое с множителем f_m соответствует пачке из m заявок, так что для попадания на j -й ярус исходный должен иметь номер $j-m$. Сомножители произведения с учетом вышеупомянутого деления имеют смысл вероятностей переходов по прибытии заявки между смежными ярусами, а все произведение — вероятности перехода между микросостояниями ярусов $j-m$ и j . Предельный индекс суммирования

$\widehat{M} = \min\{j, M\}$, так что для нулевого яруса правая часть (10) сводится к $\gamma_1 \mathbf{B}_1$.

Выразим векторы $\{\gamma_j\}$ через векторы $\{t_j\}$ условных вероятностей микросостояний яруса и суммарные вероятности $\{p_j\}$ наличия в системе ровно j заявок: $\gamma_j = p_j t_j$, и введем отношения вероятностей смежных ярусов

$$z_j = p_{j-1} / p_j, \quad x_j = p_{j+1} / p_j. \quad (11)$$

Теперь уравнения (10) можно переписать в виде

$$t_j \mathbf{D}_j = \lambda \sum_{m=1}^{\widehat{M}} f_m t_{j-m} \left(\prod_{i=0}^{m-1} z_{j-i} \right) \left(\prod_{i=0}^{m-1} \mathbf{A}_{j-m+i} \right) + x_j t_{j+1} \mathbf{B}_{j+1}. \quad (12)$$

Таким образом:

$$t_j = z_j \beta'_j + x_j \beta''_j, \quad (13)$$

где

$$\beta'_j = \lambda \left[\sum_{m=1}^{\widehat{M}} f_m t_{j-m}^{(k-1)} \left(\prod_{i=1}^{m-1} z_{j-i} \right) \left(\prod_{i=0}^{m-1} \mathbf{A}_{j-m+i} \right) \right] \mathbf{D}_j^{-1}; \quad (14)$$

$$\beta''_j = t_{j+1}^{(k)} \mathbf{B}_{j+1} \mathbf{D}_j^{-1}. \quad (15)$$

Верхними индексами (в скобках) указаны номера последовательных приближений, в каждом из которых делается прогонка по всем ярусам. Смысл перехода к векторам условных вероятностей в том, что только для них удастся задать хорошие начальные приближения, причем лишь для больших j . Последнее обстоятельство определяет и направление прогонки — от больших значений индексов j к нулю. Второе преимущество перехода

к условным векторам вероятностей — возможность замыкания расчетной схемы на основе допущения

$$\mathbf{t}_{j_{\max}}^{(k)} = \mathbf{t}_{j_{\max}-1}^{(k-1)}.$$

Переход к условным вероятностям связан с введением для каждого яруса дополнительных неизвестных $\{x_j\}$ и $\{z_j\}$. Соответственно нужны дополнительные уравнения для их определения. В качестве первого выберем баланс переходов между j -м и $(j+1)$ -м ярусами. Интенсивность переходов сверху должна учитывать $\widehat{M} = \min\{M-1, j\}$ ярусов выше j -го; она составит

$$\Lambda_j = \lambda \left[p_j + p_{j-1}(1-f_1) + p_{j-2}(1-f_1-f_2) + \dots + p_{j-\widehat{M}+1} \left(1 - \sum_{i=1}^{\widehat{M}-1} f_i \right) \right] = p_j \lambda \left[1 + z_j \sum_{i=1}^{\widehat{M}} \left(1 - \sum_{m=1}^i f_m \right) \left(\prod_{m=1}^{i-1} z_{j-m} \right) \right].$$

Интенсивность переходов снизу есть $p_{j+1} \mathbf{t}_{j+1}^{(k)} \mathbf{B}_{j+1} \mathbf{1}_j$. Приравнявая эти количества и разделив обе части равенства на p_j , получаем уравнение

$$\lambda + \lambda z_j \left[\sum_{i=1}^{\widehat{M}} \left(1 - \sum_{m=1}^i f_m \right) \left(\prod_{m=1}^{i-1} z_{j-m} \right) \right] = x_j \mathbf{t}_{j+1}^{(k)} \mathbf{B}_{j+1} \mathbf{1}_j. \quad (16)$$

Второе дополнительное уравнение $(z_j \beta'_j + x_j \beta''_j) \mathbf{1}_j = 1$ дает условие нормировки компонент \mathbf{t}_j . Из него следует, что

$$x_j = (1 - z_j \beta'_j \mathbf{1}_j) / (\beta''_j \mathbf{1}_j). \quad (17)$$

В этих формулах $\mathbf{1}_j$ есть вектор-столбец из единиц с числом элементов, равным количеству микросостояний на j -м ярусе. Умножение на него справа соответствует подсчету суммы компонент левого сомножителя. Подставляя (17) в (16), убеждаемся, что

$$z_j = \frac{\mathbf{B}_{j+1} \mathbf{1}_j / (\beta''_j \mathbf{1}_j) - \lambda}{\lambda \left[\sum_{i=1}^{\widehat{M}} \left(1 - \sum_{m=1}^i f_m \right) \left(\prod_{m=1}^{i-1} z_{j-m} \right) \right] + \mathbf{B}_{j+1} \mathbf{1}_j (\beta'_j \mathbf{1}_j) / (\beta''_j \mathbf{1}_j)}. \quad (18)$$

Выведенные соотношения для $j = j_{\max}, j_{\max}-1, \dots, 1$ применяются в следующей очередности:

- 1) получить β'_j и β''_j согласно (14) и (15);
- 2) найти z_j по формуле (18);
- 3) вычислить x_j согласно (17);
- 4) получить очередное приближение $\mathbf{t}_j^{(k)}$ по формуле (13).

Поскольку на нулевом ярусе имеется всего одно микросостояние, $\mathbf{t}_0 \equiv 1$ и в пересчете не нуждается.

Начальные приближения

Для расчета начальных приближений воспользуемся установленной экспериментально стабилизацией векторов $\{\mathbf{t}_j\}$ и отношений смежных вероятностей при $j \rightarrow \infty$. В ряде работ по теории очереди для предельного x предлагается формула

$$x = \rho^{2/(v_A^2 + v_B^2)}, \quad (19)$$

где ρ — коэффициент загрузки системы; v_A^2 и v_B^2 — квадраты коэффициентов вариации распределений интервалов между прибывающими заявками и длительности обслуживания соответственно. В нашем случае коэффициент загрузки вычисляется согласно (7), а интервалы между заявками для второй и последующих заявок пачки равны нулю. Таким образом, для простейшего потока пачек интенсивности λ и пачки объема m средний интервал между заявками

$$a_1(m) = \frac{1}{m} \frac{1}{\lambda} + \frac{m-1}{m} \cdot 0 = 1/(m\lambda),$$

а второй момент $a_2(m) = 2/(m\lambda)^2$. Подставляя в выражение для квадрата коэффициента вариации $v_A^2 = a_2/a_1^2 - 1$ моменты интервалов между заявками, усредненные по распределению объема пачки, находим

$$v_A^2 = 2 / \left(\sum_{m=1}^M f_m / m \right) - 1. \quad (20)$$

Для определения предельного вектора \mathbf{t} воспользуемся отмеченной выше стабилизацией при $j > n$ матриц интенсивностей переходов и самих векторов $\{\mathbf{t}_j\}$. Тогда индексы при векторах, матрицах, $\{x_j\}$ и $\{z_j\}$ можно опустить, а все матрицы \mathbf{A} , упоминаемые в уравнении (12), считать единичными и из рассмотрения исключить. Теперь уравнение (12) запишется в виде

$$\mathbf{tD} = \lambda \left(\sum_{m=1}^M f_m z^m \right) \mathbf{t} + x \mathbf{tB}, \quad (21)$$

где $z = 1/x$. Расписав это векторно-матричное уравнение покомпонентно и заменив одно из уравнений на условие нормировки компонент вектора \mathbf{t} , получаем систему линейных алгебраических уравнений для начальных приближений к $\{\mathbf{t}_j\}$, $j = \overline{n, j_{\max}}$. Для $j = \overline{1, n-1}$ компоненты начальных векторов $\{\mathbf{t}_j\}$ приходится принять равновероятными.

Начальные значения $\{z_j\}$ для $j > n$ следует считать как $1/x$, где x определяется согласно (19) с подстановкой v_A^2 из (20). Начальные $\{z_j\}$ для первых ярусов диаграммы можно получить с помощью соображений, использованных при выводе (16). Проводя горизонтальные разрезы между ярусами j и $j-1$, убеждаемся, что

$$p_j^{(0)} = \frac{\lambda}{\mu_j} [p_{j-1} + (1-f_1)p_{j-2} + \dots + (1-f_1 - \dots - f_{M-1})p_{j-M+1}].$$

Набор учитываемых вероятностей для каждого разреза ограничивается требованием неотрицательности индексов. Усредненные интенсивности обслуживания $\bar{\mu}_j = j/b_1$. Вероятность $p_0^{(0)}$ предполагается равной единице. Далее вычисляются начальные приближения к отношениям вероятностей

$$z_j^{(0)} = p_{j-1}^{(0)} / p_j^{(0)}, \quad j = \overline{1, n}.$$

Условие прекращения итераций

Итерации можно считать законченными, когда стабилизируются значения $\{z_j\}$, т. е. будет выполнено условие $\max_j \{ |z_j^{(k)} - z_j^{(k-1)}| \} \leq \epsilon$. Требования по точности естественно задать на уровне практически значимых вероятностей: $\epsilon \approx 10^{-6}$.

Расчет итоговых вероятностей состояний

Условие баланса заявок для n -канальной системы в данном случае имеет вид

$$\sum_{j=0}^{n-1} (n-j)p_j = n - \lambda \bar{f} b_1. \quad (22)$$

Очевиден вероятностный смысл условия: ожидаемое число свободных каналов равно недогрузке системы. Входящие в (22) вероятности последовательно выражаются через p_0 : $p_1 = p_0/z_1$, $p_2 = p_1/z_2 = p_0/(z_1 z_2)$ и т. д. Следовательно:

$$p_0 \left[n + \sum_{j=1}^{n-1} (n-j) / \prod_{i=1}^j z_i \right] = n - \lambda \bar{f} b_1,$$

откуда и определяется p_0 . Последующие вероятности $p_j = p_{j-1}/z_j$, $j = 1, 2, \dots$

Результаты расчета

Расчет по описанной методике выполнялся для трехканальной системы с равномерным на $[0, 10]$ распределением обслуживания и равновероятным (от 1 до 6) объемом пачек заявок. Интенсивность потока пачек выбиралась для получения коэффи-

циента загрузки $\rho = 0,8$. Опорное отношение вероятностей $z_\infty = 1/x_\infty$ составило 1,111. Для $j_{\max} = 39$ стабилизация $\{z_j\}$ (табл. 7) с точностью 10^{-9} потребовала 44 итераций.

Дальнейшие $\{z_j\}$ практически совпадают с z_{27} . Как видно, они очень близки к опорному. То же можно сказать и о предельных вероятностях условных микросостояний.

В табл. 8 расчетное распределение числа заявок в системе сопоставлено с результатами ИМ (10 млн испытаний).

Результаты согласуются очень хорошо. Отметим также, что сумма полученных вероятностей, дополненная суммой образующих геометрическую прогрессию неучтенных членов, с точностью 10^{-7} совпала с единицей, причем условие нормировки полных вероятностей в расчетной схеме не использовалось. Это согласие является дополнительным подтверждением правильности расчетной схемы и ее программной реализации.

Имитационная модель допускала до 100 заявок в системе. Однако сходимость численного метода при увеличении j_{\max} заметно ухудшается. Поэтому целесообразно выполнять расчет при умеренных j_{\max} , а требуемое число дополнительных вероятностей вычислять по формуле $p_j = p_{j-1}/z_{j_{\max}}$.

Распределение пачек и их задержек

Выше отмечалась необходимость расчета распределения числа «непочатых» пачек в очереди. Такая задача является обратной по отношению к естественному переходу от распределения числа пачек к распределению числа заявок и как таковая обладает вычислительной неустойчивостью. Это подтвердили попытки решения соответствующего уравнения в дискретных свертках несколькими разными методами. В связи с этим ниже предлагается хорошо зарекомендовавший себя полуматричный метод.

Анализ результатов имитации показал, что отношение смежных вероятностей искомого распределения близко к интуитивно ожидаемому: $Z = z_\infty^f$. Полные пачки в очереди отсутствуют, если заявок в очереди нет или их количество не превышает среднего остатка пачки (скажем, k — поло-

■ Таблица 7. Расчетные отношения смежных вероятностей

j	z_j	j	z_j	j	z_j	j	z_j
0	.10000e+01	7	.11719e+01	14	.10929e+01	21	.10957e+01
1	.20589e+01	8	.10837e+01	15	.10940e+01	22	.10958e+01
2	.11822e+01	9	.10776e+01	16	.10952e+01	23	.10958e+01
3	.10695e+01	10	.10824e+01	17	.10958e+01	24	.10958e+01
4	.96669e+00	11	.10918e+01	18	.10958e+01	25	.10958e+01
5	.99908e+00	12	.10972e+01	19	.10956e+01	26	.10958e+01
6	.10468e+01	13	.10976e+01	20	.10956e+01	27	.10959e+01

■ Таблица 8. Распределения количества заявок в системе

<i>j</i>	Расчет	Имитация	<i>j</i>	Расчет	Имитация
0	.13692e+00	.13727e+00	20	.14154e-01	.14266e-01
1	.66499e-01	.65292e-01	21	.12917e-01	.13041e-01
2	.56251e-01	.58327e-01	22	.11788e-01	.11928e-01
3	.52598e-01	.52512e-01	23	.10757e-01	.10876e-01
4	.54410e-01	.54403e-01	24	.98164e-02	.98637e-02
5	.54460e-01	.54402e-01	25	.89579e-02	.90459e-02
6	.52027e-01	.51411e-01	26	.81745e-02	.82060e-02
7	.44397e-01	.44084e-01	27	.74594e-02	.74814e-02
8	.40967e-01	.40911e-01	28	.68069e-02	.67880e-02
9	.38016e-01	.37984e-01	29	.62115e-02	.61825e-02
10	.35122e-01	.35059e-01	30	.56681e-02	.56444e-02
11	.32169e-01	.32082e-01	31	.51723e-02	.51133e-02
12	.29320e-01	.29277e-01	32	.47198e-02	.46397e-02
13	.26713e-01	.26761e-01	33	.43070e-02	.42310e-02
14	.24442e-01	.24510e-01	34	.39302e-02	.38639e-02
15	.22341e-01	.22394e-01	35	.35864e-02	.35175e-02
16	.20400e-01	.20592e-01	36	.32726e-02	.32098e-02
17	.18617e-01	.18672e-01	37	.29864e-02	.29245e-02
18	.16990e-01	.17102e-01	38	.27251e-02	.26566e-02
19	.15507e-01	.15581e-01	39	.24867e-02	.24467e-02

вины ее среднего значения). Тогда можно считать, что вероятность этого события

$$q_0 = \sum_{i=0}^{n+\bar{k}} p_i,$$

а остальные вероятности образуют убывающую геометрическую прогрессию с начальным членом

q_1 и знаменателем Z . Замыкающую схему недостающую вероятность находим из условия нормировки: $q_1 = (1 - 1/Z)(1 - q_0)$.

В табл. 9 проводится сопоставление результатов расчета с имитацией.

Имея распределение числа непечатых пачек в очереди, можно по формуле Брюмелля выразить

■ Таблица 9. Распределения количества пачек в очереди

<i>j</i>	Расчет	Имитация	<i>j</i>	Расчет	Имитация
0	.47316e+00	.45458e+00	15	.16276e-02	.14718e-02
1	.14443e+00	.15136e+00	16	.11814e-02	.10227e-02
2	.10484e+00	.11139e+00	17	.85751e-03	.71961e-03
3	.76095e-01	.80331e-01	18	.62243e-03	.51405e-03
4	.55234e-01	.57622e-01	19	.45179e-03	.37014e-03
5	.40092e-01	.41471e-01	20	.32793e-03	.25248e-03
6	.29101e-01	.29731e-01	21	.23803e-03	.16101e-03
7	.21123e-01	.20981e-01	22	.17278e-03	.12383e-03
8	.15332e-01	.14868e-01	23	.12541e-03	.98421e-04
9	.11129e-01	.10666e-01	24	.91029e-04	.68672e-04
10	.80779e-02	.76942e-02	25	.66074e-04	.50771e-04
11	.58634e-02	.55641e-02	26	.47960e-04	.33102e-04
12	.42559e-02	.39279e-02	27	.34812e-04	.19618e-04
13	.30892e-02	.28407e-02	28	.25268e-04	.13579e-04
14	.22423e-02	.20432e-02	29	.18341e-04	.68985e-05

■ Таблица 10. Моменты распределения ожидания пачек

Способ определения	w_1	w_2	w_3
Расчет	.14013e+02	.54106e+03	.31337e+05
Имитация непосредственно	.14079e+02	.51615e+03	.28040e+05
Имитация через MFACT	.14067e+02	.51369e+03	.27659e+05

моменты распределения времени ожидания начала обслуживания пачки (табл. 10).

Заключение

Широкий круг практически важных систем обслуживания описывается математическими моделями с групповым потоком заявок. Предло-

женные расчетные зависимости впервые дают возможность получать основные вероятностные характеристики многоканальных систем. Имитационные эксперименты подтвердили достаточную точность расчета практически значимых вероятностей и моментов временных распределений.

Литература

1. Кудрявцев Е. М. GPSS World. Основы имитационного моделирования различных систем. М.: ДМК Пресс, 2004. 320 с.
2. Рыжиков Ю. И., Хомоненко А. Д. Итеративный метод расчета многоканальных систем с произвольным распределением времени обслуживания // Проблемы управления и теории информации. 1980. № 3. С. 32–38.
3. Рыжиков Ю. И. Теория очередей и управление запасами. СПб.: Питер, 2001. 384 с.
4. Рыжиков Ю. И. Имитационное моделирование. Теория и технологии. СПб.: КОРОНА принт, 2004. 384 с.
5. Рыжиков Ю. И. Современный Фортран. СПб.: КОРОНА принт, 2004. 288 с.
6. Саати Т. Л. Элементы теории массового обслуживания и ее приложения: Пер. с англ. М.: Сов. радио, 1965. 510 с.
7. Томашевский В. Н., Жданова Е. Г. Имитационное моделирование в среде GPSS. М.: Бестселлер, 2003. 416 с.
8. Chaudry M. L., Gupta U. C., Agarwal M. Computational analysis of the distribution of demands number in the system $M^X/G/1$ — an alternative approach // INFOR. 1992. Vol. 30. N 1. P. 30–43.
9. Krakowski M. Arrival and departure processes in queues — Pollaczek — Khintchine formulas for bulk arrivals and bounded systems // RFAIRO. 1974. Vol. 1. P. 45–56.
10. Loris-Teghem J. A. Modèles d'attente $M/G/1$ et $GI/M/1$ à arrivées et service en groupe. Berlin: Springer, 1969. 53 p.
11. Ommereen J. C. W., van. Simple approximation for the batch-size $M^X/G/1$ queue // OR. 1990. Vol. 38. N 4. P. 678–685.
12. Sivasamy R. A bulk service queue with accessible and non-accessible batches // Opsearch. 1990. Vol. 27. N 1. P. 46–54.
13. Takahashi Y., Takami Y. A numerical method for the steady-state probabilities of a $GI/G/c$ queuing system in a general class // J. of the Operat. res. soc. of Japan. 1976. Vol. 19. N 2. P. 147–157.
14. Wirth K. D. A remark on relations between batch delays and customer delays // J. of Information Processing and Cybernetics. 1985. Vol. 21. N 1/2. P. 65–67.