

УДК 519.872

## ОПЫТ РАСЧЕТА СЛОЖНЫХ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

**Ю. И. Рыжиков,**

доктор техн. наук, профессор

**А. В. Уланов,**

курсант

Военно-космическая академия им. А. Ф. Можайского

Анализируется опыт применения современных методов расчета систем массового обслуживания, в частности рассматривается роль высших моментов распределений времени обслуживания и интервалов между заявками входящего потока. Исследуются эффект дробления производительности, масштабный эффект, эффект очереди в многоканальных системах и приоритетное обслуживание. Показана практическая непригодность формул Эрланга и категорическая необходимость перехода к применению (и преподаванию) современных методов теории массового обслуживания.

**Ключевые слова** — теория очередей, пакет программ, немарковские системы массового обслуживания, численные результаты.

### Введение

В самых разнообразных областях человеческой деятельности часто приходится иметь дело с работой своеобразных систем, называемых системами массового обслуживания (СМО). Примерами таких систем могут служить взлетно-посадочная полоса аэродрома, автоматизированные системы сбора и обработки информации, сети связи, учреждения здравоохранения и т. п. По справедливому замечанию академика Б. В. Гнеденко, легче указать области, где не применяется теория массового обслуживания (ТМО), чем перечислить все области ее применения.

В классической (эрланговской) теории массового обслуживания [1] рассматриваются системы с простейшим входящим потоком и показательной распределенной длительностью обслуживания. Но на практике часто приходится иметь дело с системами, заметно отличающимися от простейших, и тогда, как будет показано ниже, погрешность от применения формул Эрланга может быть сколь угодно велика.

Однако в современной литературе по исследованию операций численным и аналитическим методам расчета немарковских СМО не уделяется должного внимания. Утверждается даже [2], что такие задачи можно решить лишь с помощью имитационного моделирования. Во многих вузах

преподавание ТМО ограничивается рассмотрением простейших марковских моделей, хотя известны более реалистичные аналитические и численные методы, позволяющие рассчитывать сложные, немарковские системы. Имеется множество журнальных публикаций по методам расчета немарковских систем, но результаты нигде не доводятся до числа, поэтому о необходимости применения «продвинутых» методов судить сложно.

В Военно-космической академии им. А. Ф. Можайского уже более четверти века преподаются машинные методы расчета СМО [3, 4]. Лекционный цикл дисциплины состоит из семи тем:

- 1) аппроксимация вероятностных распределений;
- 2) элементы модели теории очередей;
- 3) законы сохранения;
- 4) имитационное моделирование;
- 5) расчет немарковских систем;
- 6) приоритетное обслуживание;
- 7) сети обслуживания.

Практическая часть курса посвящена выводу следствий условия баланса заявок и построению диаграмм переходов для многоканальных систем. Часть этих занятий отводится на подготовку к лабораторным работам. Лабораторный практикум состоит из двух работ — проверки законов сохранения на имитационной модели и расчета многоканальных немарковских систем. Они ба-

зируются на учебной версии разработанного в академии пакета прикладных программ МОСТ (Массовое Обслуживание — СТАционарные задачи). Теоретическими основами пакета служат аппроксимация исходных распределений по методу моментов и законы сохранения (заявок, вероятностей, стационарной очереди). Он реализован на ПЭВМ, программно совместимых с IBM PC, и работает под управлением Windows 98. Для его применения необходима система программирования Fortran PowerStation 1.0, которая:

- реализует стандарт Фортрана-77 с критически важными расширениями из Фортрана-90;
- обеспечивает привычный для современных пользователей оконный интерфейс;
- довольствуется умеренными аппаратными ресурсами (ПЭВМ класса «486», оперативная память 6 МБ и дисковая — до 15 МБ).

Процедуры пакета оформлены в виде библиотеки объектных модулей и вызываются из составленной пользователем главной процедуры, записанной на расширенном Фортране-77. МОСТ не имеет ни зарубежных, ни отечественных аналогов.

В настоящее время заканчивается работа над профессиональной версией ППП МОСТ на Фортране-90. Она будет включать 180 процедур и 120 тестов и позволит решать не только практические, но и исследовательские задачи теории очередей с наибольшей точностью и эффективностью. Техника использования, как и в учебной версии, — сборочное программирование.

### Влияние коэффициента вариации немарковского распределения на показатели обслуживания в одноканальных системах

Важнейшей характеристикой любой СМО является распределение числа заявок в системе, поскольку все остальные вероятностно-временные показатели работы системы можно выразить через него. Рассмотрим систему с немарковским обслуживанием  $M/G/1$ . Расчет распределения числа заявок производится методом вложенных цепей Маркова [3]

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

где  $\{q_j\}$  — вероятности прибытия ровно  $j$  заявок простейшего потока за случайное время обслуживания:

$$q_j = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} dB(t), \quad (2)$$

здесь  $B(t)$  — распределение времени обслуживания;  $\lambda$  — интенсивность входящего потока.

Расчет систем  $GI/M/1$  производится также методом вложенных цепей Маркова, но по формулам

$$p_0 = 1 - \rho;$$

$$p_k = \rho(1 - \omega)\omega^{k-1}, \quad k = 1, 2, \dots, \quad (3)$$

где  $\rho = \lambda b_1$  — коэффициент загрузки системы;  $b_1$  — среднее время обслуживания;  $\omega$  — корень из интервала  $(0, 1)$  уравнения

$$\omega = \int_0^{\infty} e^{-\mu(1-\omega)t} dA(t),$$

здесь  $A(t)$  — распределение интервалов между заявками. Это уравнение решается методом простых итераций или методом Вегстейна [5]. Графики распределений числа заявок представлены на рис. 1 и 2. Оба рисунка свидетельствуют о категорической необходимости учитывать второй момент немарковского распределения, заметно отличающийся от показательного. Заметим, что для систем типа  $GI/M/1$  графики распределений вероятностей в полулогарифмическом масштабе при наличии в системе заявок представляют собой прямые, что согласуется с формулой (3). Чтобы исследовать влияние коэффициента вариации распределения времени обслуживания на дополнительную функцию распределения (ДФР) времени пребывания заявки в системе, были рассчитаны шесть моментов распределения времени пребывания, а затем по ним подобрана ДФР Вейбулла с поправочным многочленом [4]. Ее график (рис. 3) также свидетельствует о непригодности формул Эрланга для расчета немарковских систем.

Среднее время ожидания начала обслуживания для систем типа  $M/G/1$  находится по формуле Полячека — Хинчина (ФПХ)

$$w_1 = \frac{\lambda b_2}{2(1 - \rho)},$$

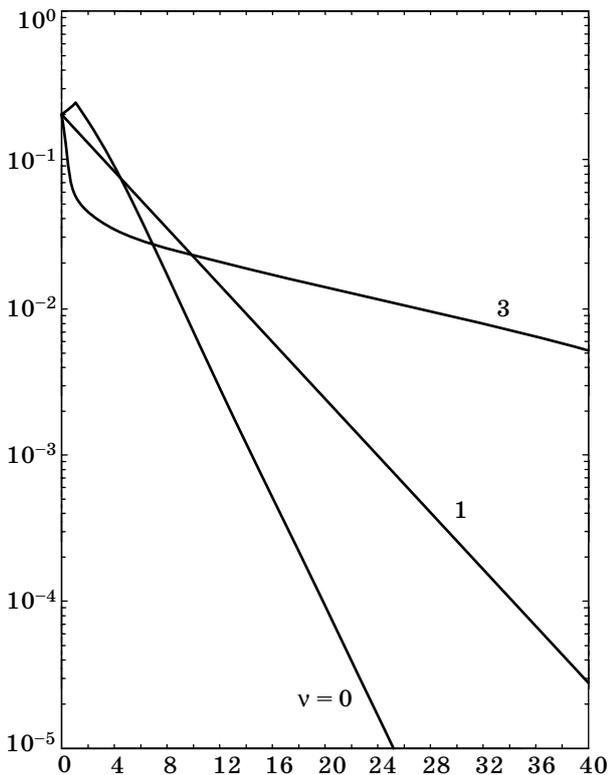
моменты более высокого порядка — согласно формуле

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

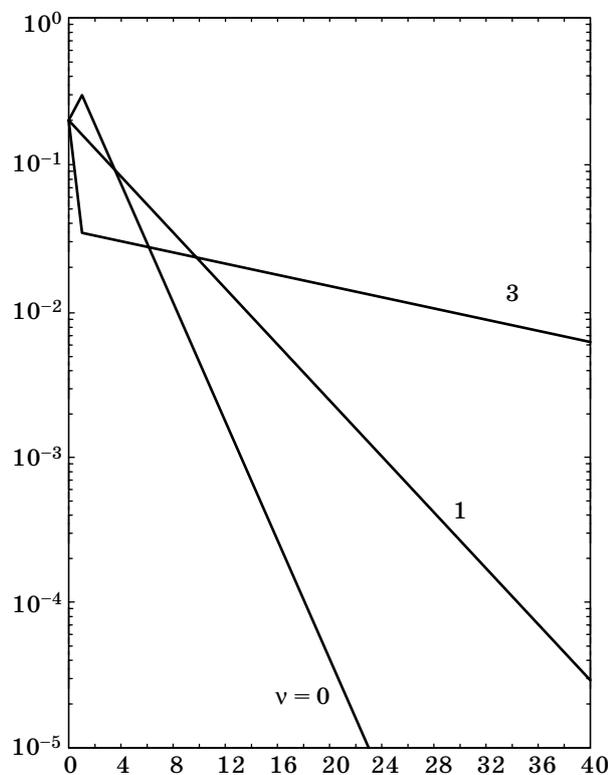
Возможен альтернативный вариант вычисления — через факториальные моменты распределения длины очереди  $\{q_{[k]}\}$ , которые рассчитывает входящая в ППП МОСТ процедура MFACT. В этом случае

$$w_k = q_{[k]} / \lambda^k. \quad (4)$$

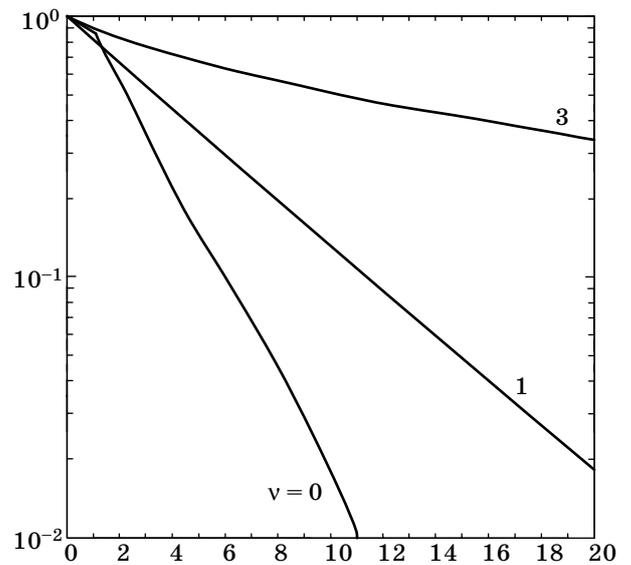
Для исследования влияния коэффициента вариации на среднее время ожидания начала об-



■ Рис. 1. Влияние коэффициента вариации времени обслуживания на распределение числа заявок в системах типа  $M/G/1$



■ Рис. 2. Влияние коэффициента вариации интервалов между заявками входящего потока на распределение числа заявок в системах типа  $GI/M/1$



■ Рис. 3. Влияние коэффициента вариации обслуживания на ДФР времени пребывания заявки в системах  $M/G/1$

служивания был выполнен непосредственный расчет по ФПХ, расчет через факториальные моменты и с помощью имитационной модели (1 млн испытаний). Результаты представлены в табл. 1.

Имитационное моделирование при расчете среднего времени ожидания дает наименее точный результат. Это вызвано, прежде всего, несовершенством программных датчиков случайных чисел и свидетельствует о предпочтительности применения и необходимости дальнейшего развития аналитических и численных методов для анализа СМО. Заметим, что расчет через факториальные моменты дает неточный результат для распределения времени обслуживания с большим коэффициентом вариации (с «толстым хвостом»). Это связано с тем, что при расчете факториальных моментов используется ограниченное число вероятностей. При неограниченной очереди предполагается, что вероятности «на хвосте» стационарного распределения представляют собой бесконечно убывающую геометрическую прогрессию, знаменатель которой находится как

■ Таблица 1. Влияние коэффициента вариации распределения времени обслуживания на среднее время ожидания

Способ расчета	Коэффициент вариации $v$		
	0	1	3
По ФПХ	.2000e+01	.4000e+01	.2000e+02
Через МФАСТ	.2000e+01	.4000e+01	.1878e+02
В модели	.1995e+01	.4029e+01	.1714e+02
В модели через МФАСТ	.1996e+01	.4026e+01	.1699e+02

отношение последней учтенной вероятности к предпоследней. В полулוגарифмическом масштабе такая зависимость представляется прямой линией. Так как относительная погрешность при вычислении редких событий на имитационной модели велика, поправка в факториальные моменты, вычисленные по данным моделирования, не вводится.

Также было установлено, что при увеличении числа каналов влияние коэффициента вариации на распределение числа заявок в системе уменьшается.

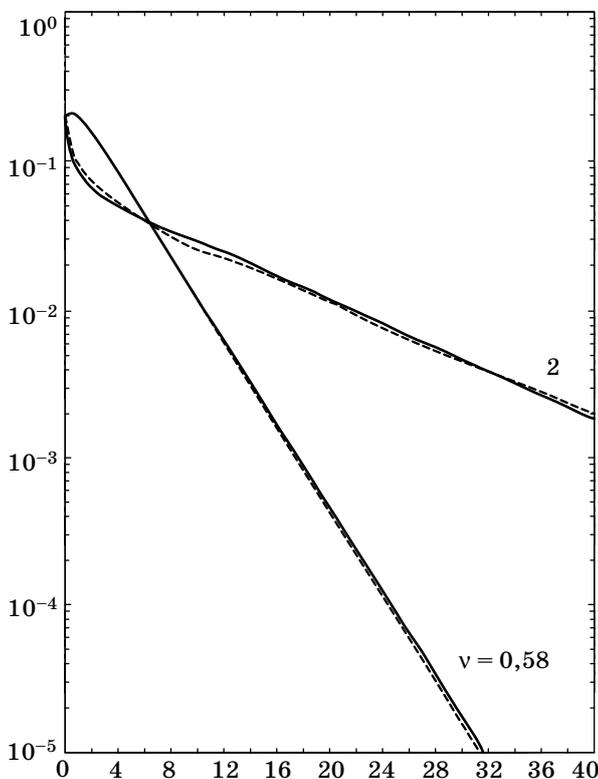
### Влияние высших моментов на распределение числа заявок в системах M/G/1

Для исследования влияния высших (начиная с третьего) моментов на распределение числа заявок в системе заменим исходное распределение времени обслуживания гамма-плотностью с поправочным многочленом [4], позволяющей выровнять более двух моментов:

$$f(t) = \frac{\mu(\mu t)^{\alpha-1}}{\Gamma(\alpha)} e^{-\mu t} \sum_{i=0}^N g_i t^i, \quad (5)$$

где  $N$  — число учитываемых моментов.

Далее необходимо найти вероятности  $\{q_j\}$  по формуле (2). Затем производится расчет ста-



■ Рис. 4. Влияние высших моментов на распределение числа заявок в системах с немарковским обслуживанием M/G/1

ционарного распределения с помощью рекуррентной формулы (1). По этой методике было исследовано влияние высших моментов на стационарное распределение в системах с немарковским обслуживанием для распределения Вейбулла ( $\nu = 2$ ) и равномерного распределения ( $\nu = 0,577$ ) при коэффициенте загрузки  $\rho = 0,8$ . Сплошные линии (рис. 4) соответствуют аппроксимации по двум моментам, т. е. «чистому» гамма-распределению, штриховые — по шести моментам с помощью формулы (5).

Видно относительно слабое влияние третьего и высших моментов на распределение вероятностей, особенно при малом коэффициенте вариации, что позволяет в инженерных расчетах ограничиться аппроксимацией гамма-плотностью. Также было установлено, что при уменьшении коэффициента загрузки это влияние ослабевает.

### Эффект дробления производительности и масштабный эффект

При исследовании обоих упомянутых эффектов предполагалось сохранение коэффициента загрузки  $\rho = 0,8$ . Выходными показателями являлись средние времена ожидания и пребывания в системе. В первом случае исследовалось влияние на них числа каналов, во втором — одинаковое увеличение интенсивности входящего потока и интенсивности обслуживания в каждом из каналов.

Принято считать, что средние времена ожидания начала обслуживания и пребывания заявки в системе должны оставаться неизменными. Этот тезис опроверг Б. В. Гнеденко для простейшей системы M/M/1, где среднее время пребывания  $\nu(1) = 1/(\mu - \lambda)$ , а при  $m$ -кратном увеличении интенсивности поступления и обслуживания  $\nu(m) = 1/(m\mu - m\lambda) = \nu(1)/m$ . Для исследования масштабного эффекта в немарковских многоканальных системах нами была рассчитана система  $E_3/M/n$ , где  $E_3$  — распределение Эрланга третьего порядка.

Расчет производился методом фиктивных фаз [3] через факториальные моменты и на имитационной модели для 1 млн испытаний (среднее время ожидания), среднее время пребывания в этом случае рассчитывалось как сумма средних времен ожидания и обслуживания. Результаты приведены в табл. 2. Как видно, реакция средних времен ожидания и пребывания на масштабирование интенсивностей поступления и обслуживания заявок аналогична той, что имеет место в простейшем случае.

Эффект дробления производительности рассмотрен на примере системы  $M/E_3/n$ ,  $n = 1, 2, 3$ . Расчет производился так же, как при исследова-

■ Таблица 2. Масштабный эффект

Показатель		$m = 1$	$m = 2$	$m = 3$
$w$	Расчет	.23944e+01	.11972e+01	.79812e+01
	Имитация	.23689e+01	.11845e+01	.78964e+01
$v$	Расчет	.43944e+01	.21972e+01	.14648e+01
	Имитация	.43689e+01	.21845e+01	.14563e+01

■ Таблица 3. Эффект дробления производительности

Показатель		$n = 1$	$n = 2$	$n = 3$
$w$	Расчет	.26667e+01	.23944e+01	.21941e+01
	Имитация	.26419e+01	.23689e+01	.21241e+01
$v$	Расчет	.36667e+01	.43944e+01	.51941e+01
	Имитация	.36419e+01	.43689e+01	.51241e+01

нии масштабного эффекта. Результаты приведены в табл. 3, откуда следует:

— с ростом числа каналов среднее время ожидания начала обслуживания уменьшается, но незначительно;

— среднее время обслуживания изменяется обратно пропорционально интенсивности обслуживания в одном канале, т. е. возрастает, причем возрастает быстрее, чем уменьшается среднее время ожидания;

— среднее время пребывания является суммой средних времен ожидания и обслуживания и поэтому с увеличением числа каналов оно также возрастает.

Таким образом, с точки зрения оперативности обработки информации целесообразнее иметь одно обслуживающее устройство, чем несколько с той же суммарной производительностью. Но поскольку в процессе эксплуатации технических устройств необходимо проводить профилактику и ремонт, оптимальное число каналов равно двум. Такой принцип (спарки) использовался при эксплуатации машин ЕС-1045.

### Эффект общей очереди

В бытовой сфере часто возникает вопрос, как при наличии нескольких обслуживающих устройств организовать очередь — отдельно к каждому устройству или общую. ППП МОСТ позволяет ответить на этот вопрос на базе количественных оценок.

Исследовались системы с простейшим входящим потоком и различными коэффициентами вариации  $v$  времени обслуживания при сложении  $n$  очередей. Расчет проводился при коэффициенте загрузки  $\rho = 0,7$  и  $0,9$ . Случай с одним ка-

■ Таблица 4. Среднее время ожидания в системе с общей очередью

$n$	$v = 0,5$		$v = 1,0$		$v = 2,0$	
	0,7	0,9	0,7	0,9	0,7	0,9
1	1.458	5.625	2.333	9.000	5.833	22.500
2	0.611	2.679	0.961	4.263	2.334	10.574
3	0.353	1.717	0.547	2.724	1.305	6.718
4	0.232	1.245	0.357	1.969	0.839	4.828
5	0.166	0.967	0.252	1.525	0.583	3.696

налом ( $n = 1$ ) соответствует изолированным очередям. Результаты (табл. 4) показывают, что при наличии нескольких обслуживающих устройств гораздо выгоднее организовать общую очередь. Этот интересный эффект объясняется тем, что ввиду случайного характера поступления заявок при разделении очередей могут возникать ситуации, когда к одному каналу возникла очередь, а другой канал свободен. При объединении очередей система становится многоканальной и такие ситуации исключены. При увеличении количества «суммируемых» очередей рассматриваемый эффект затухает.

### Приоритетное обслуживание

Известно [3], что введение приоритетов лишь перераспределяет ресурс системы в пользу более приоритетных заявок за счет менее приоритетных, причем относительный выигрыш первых больше, чем относительный проигрыш вторых. Классик теории очередей Л. Клейнрок утверждает [6], что при этом для консервативных (не создающих дополнительную работу в системе) дисциплин должно иметь место сохранение взвешенной суммы средних времен ожидания заявок с разными приоритетами  $\sum_i \rho_i w_i = \text{const}$  (инвариант объема работ). Проверим правильность этого утверждения для различных дисциплин обслуживания (FCFS — в порядке прибытия, NP — приоритет без прерывания, PR — с прерыванием и дообслуживанием) при изменении коэффициента вариации времени обслуживания и коэффициента загрузки системы. Длительности обслуживания подчинены гамма-распределению с общим параметром формы  $\beta$ . Результаты приведены в табл. 5.

При показательно распределенной длительности обслуживания ( $\beta = 1,0$ ) объем работ действительно не зависит от дисциплины обслуживания, для дисциплин FCFS и NP он также инвариантен. Но для дисциплины с прерыванием и дообслуживанием (PR) и немарковского обслуживания разбаланс значений все же есть, хотя и составляет доли процента. Этот результат — еще

Таблица 5. Сопоставление объемов работ

ρ	β = 0,5			β = 1,0			β = 3,0		
	PR	NP	FCFS	PR	NP	FCFS	PR	NP	FCFS
.50	1.268	1.319	1.319	0.879	0.879	0.879	0.619	0.586	0.586
.55	1.711	1.733	1.733	1.182	1.182	1.182	0.829	0.788	0.788
.60	2.299	2.373	2.373	1.582	1.582	1.582	1.104	1.055	1.055
.65	3.095	3.183	3.183	2.122	2.122	2.122	1.473	1.415	1.415
.70	4.204	4.307	4.307	2.872	2.872	2.872	1.983	1.914	1.914
.75	5.813	5.933	5.933	3.956	3.956	3.956	2.717	2.637	2.637
.80	8.300	8.439	8.439	5.626	5.626	5.626	3.843	3.751	3.751
.85	12.543	12.702	12.702	8.468	8.468	8.468	5.751	5.645	5.645
.90	21.180	21.360	21.360	14.240	14.240	14.240	9.614	9.494	9.494
.95	47.396	47.600	47.600	31.733	31.733	31.733	21.291	21.155	21.155

один аргумент в пользу применения современных методов.

Малая величина разбаланса позволяет в инженерных расчетах считать постоянство объема работ для всех консервативных (не создающих дополнительную работу в системе) дисциплин хорошим приближением.

**Проверка законов сохранения на имитационной модели**

Для верификации аналитических методов и проверки их теоретического фундамента (законов сохранения) на имитационной модели была рассчитана система  $M/C_2/2$  с коэффициентом загрузки  $\rho = 0,8$ , где  $C_2$  — распределение Кокса второго порядка. Модель реализована на Фортране-77, так как современные инструментальные средства моделирования (типа GPSS World) не являются столь функциональными (например, не позволяют получить третьи моменты распределения времени ожидания, реализовать вложенные прерывания) и не способствуют усвоению базовых принципов имитационного моделирования.

На имитационной модели проверялись:

- закон сохранения заявок в форме

$$\sum_{j=0}^{n-1} (n-j)p_j = n - \lambda b_1,$$

где  $n$  — число каналов;

— закон сохранения объема работ в виде обобщенной ФПХ на случай  $n$  каналов

$$w = \frac{\lambda b_2}{2n^2(1-\rho)} e^{-0,7(n-1)(1-\rho)};$$

— закон сохранения стационарной очереди для высших (до третьего включительно) моментов времени ожидания начала обслуживания — формула (4).

Оказалось, что невязка левых и правых частей приведенных формул составляет:

- по закону сохранения заявок 1,2 %;
- по закону сохранения объема работ 4,4 %;
- по закону сохранения очереди для первого, второго и третьего моментов времени ожидания 0,2; 3,4 и 11,5 % соответственно.

Расчет на имитационной модели подтверждает справедливость законов сохранения и дает основание для дальнейшего развития аналитических и численных методов, позволяющих рассчитывать характеристики СМО быстрее и точнее.

**Заключение**

Перечислим основные выводы, которые следуют из представленных в работе расчетов.

1. Учет только средних значений, т. е. показательная аппроксимация, при заметном отличии коэффициентов вариации исходных распределений от единицы приводит к недопустимым вычислительным погрешностям. Влияние высших моментов на показатели обслуживания увеличивается с ростом коэффициента загрузки и быстро убывает по числу учтенных моментов. При любом количестве последних расхождение на «хвосте» распределения (где вероятности  $p < 0,01$ ) все равно будет велико. К тому же с ростом порядка исходных моментов увеличивается ошибка их статистического вычисления. Поэтому в инженерных расчетах достаточно выровнять от двух до четырех моментов.

2. Дробление суммарной производительности при большом коэффициенте загрузки приводит к незначительному снижению среднего времени ожидания, но увеличивает суммарное время пребывания заявки в системе из-за большего увеличения среднего времени обслуживания. В случае пропорционального увеличения интенсивностей поступления и обслуживания заявок средние времена ожидания и обслуживания уменьшаются во столько же раз. При наличии нескольких обслуживающих устройств гораздо выгоднее иметь общую очередь, чем отдельную к каждому.

3. Введение приоритетов только перераспределяет ресурс системы в пользу более приоритетных заявок за счет менее приоритетных. Относительный выигрыш заявок высшего приоритета заметно выше, чем относительный проигрыш заявок низшего приоритета.

4. Имитационное моделирование вносит в результат статистические погрешности и не позволяет исследовать тонкие эффекты (например, влияние третьего и высших моментов на распределение числа заявок). Оно должно применяться лишь для апробации новых расчетных методов и для решения особо сложных задач. Использование моделирующих систем класса GPSS World не всегда позволяет реализовать нужные вычисления и не способствует усвоению базовых принципов имитационного моделирования. Первичное обучение имитационному моделированию следует проводить не на GPSS, а на языках типа Фортран.

### Литература

1. **Вентцель Е. С.** Исследование операций. М.: Высш. шк., 2007. 208 с.
2. **Исследование операций в экономике: Учеб. пособие для вузов / Под ред. Н. Ш. Кремера.** М.: Юнити, 2006. 408 с.
3. **Рыжиков Ю. И.** Компьютерное моделирование систем с очередями / ВКА им. А. Ф. Можайского. СПб., 2007. 164 с.
4. **Рыжиков Ю. И.** Машинные методы расчета систем массового обслуживания / ВИККИ им. А. Ф. Можайского. Л., 1979. 178 с.
5. **Ланс Дж. Н.** Численные методы для быстродействующих вычислительных машин: Пер. с англ. М.: ИЛ, 1962. 208 с.
6. **Клейнрок Л.** Вычислительные системы с очередями. М.: Мир, 1979. 600 с.

### ПАМЯТКА ДЛЯ АВТОРОВ

*Поступающие в редакцию статьи проходят обязательное рецензирование.*

При наличии положительной рецензии статья рассматривается редакционной коллегией. Принятая в печать статья направляется автору для согласования редакторских правок. После согласования автор представляет в редакцию окончательный вариант текста статьи.

Процедуры согласования текста статьи могут осуществляться как непосредственно в редакции, так и по e-mail (80x@mail.ru).

При отклонении статьи редакция представляет автору мотивированное заключение и рецензию, при необходимости доработать статью — рецензию. Рукописи не возвращаются.

*Редакция журнала напоминает, что ответственность за достоверность и точность рекламных материалов несут рекламодатели.*