УДК 519.872.3

ИССЛЕДОВАНИЕ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ С ПРИОРИТЕТНЫМИ ЗАЯВКАМИ

RESEARCH OF QUEUING SYSTEMS WITH PRIORITY REQUESTS

©Осипов Г. С.
SPIN–код: 7749-0840
д–р техн. наук, Сахалинский государственный университет
г. Южно–Сахалинск, Россия, _Osipov@rambler.ru
©Osipov G.
SPIN–code: 7749-0840
Dr. habil., Sakhalin State University
Yuzhno–Sakhalinsk, Russia, _Osipov@rambler.ru

Аннотация. В работе исследуются системы массового обслуживания с приоритетной дисциплиной очереди. Предложены методологии имитационного моделирования систем обслуживания на основе априорного разделения потоков по рангам приоритетов и с возможностью блокирования очередей с заявками не обладающими высоким приоритетом.

Исследуются эмпирические зависимости, позволяющие получить априорные оценки времени ожидания в очереди для систем с последействием. Практическая апробация исследования выполнена в среде имитационного моделирования *AnyLogic*.

Abstract. In operation queuing systems with priority discipline of queue are researched. Methodologies of simulation modeling of systems of service on the basis of prior division of flows according to ranks of priorities and with a possibility of blocking of queues with the requests which don't have a high priority are offered.

The empirical dependences allowing to receive prior estimates of wait time in queue for systems with an aftereffect are researched. Practical approbation of a research is executed in the environment of simulation modeling of AnyLogic.

Ключевые слова: системы массового обслуживания, заявки, приоритеты.

Keywords: queuing systems, requests, priorities.

Реально существующие системы массового обслуживания (СМО) функционируют, как правило, не только на базе элементарной дисциплины очереди типа «первым пришел — первым обслужен», но имеют в своем арсенале более практически значимую дисциплину, основанную на возможности предоставления приоритетного (рангового) обслуживания. Это позволяет повысить глобальную (системную) экономическую эффективность функционирования СМО.

К таким системам можно отнести, например, приоритетное обслуживание отдельных категорий пассажиров в аэропорту, организацию особых очередей в поликлиниках, торговых центрах, приоритетный порядок пропуска морских и речных судов в каналах и т. д.

Очевидно приоритетное обслуживание заявок приводит к появлению не пуассоновских потоков, что резко ограничивает возможности стандартного математического аппарата теории массового обслуживания.

Поэтому настоящая работа имеет своей целью изложить методологию имитационного моделирования СМО с приоритетными заявками. Исследуются различные варианты синтеза систем обслуживания в среде имитационного моделирования *AnyLogic*.

Материал и методика

Для одноканальных систем массового обслуживания с ожиданием с простейшим потоком заявок и временем обслуживания, распределенным по показательному закону, средняя продолжительность ожидания обслуживания определяется по известному выражению [1, 2].

$$T_q = \frac{\rho}{\mu - \lambda}, \qquad (1)$$

где ρ — приведенная интенсивность потока заявок (интенсивность нагрузки канала);
λ, μ — интенсивности потока заявок и обслуживания, соответственно.

На практике часто используются СМО с приоритетным обслуживанием заявок.

Пусть в систему поступают заявки двух рангов (заявки первого ранга обслуживаются в первую очередь). В таких случаях (поток заявок уже не является простейшим) среднюю длительность ожидания обслуживания для заявок первого t_1 и второго t_2 рангов можно определить по следующим эмпирическим выражениям:

$$t_1 = k_1 T_q; t_2 = k_2 T_q;$$

где $k_1 = k_1(a, \rho); k_2 = k_2(a, \rho)$.

Здесь $a = \frac{\lambda_1}{\lambda}$ (λ_1 — интенсивность потока заявок первого ранга);

$$k_1 = \frac{1-\rho}{1-a\rho}; k_2 = \frac{1}{1-a\rho}.$$

Окончательно:

$$t_1 = \frac{1-\rho}{1-a\rho} T_q; t_2 = \frac{1}{1-a\rho} T_q. \qquad (2)$$

На Рисунке 1 приведены графики зависимости коэффициентов k_1 и k_2 от величины a при фиксированном значении нагрузки канала.

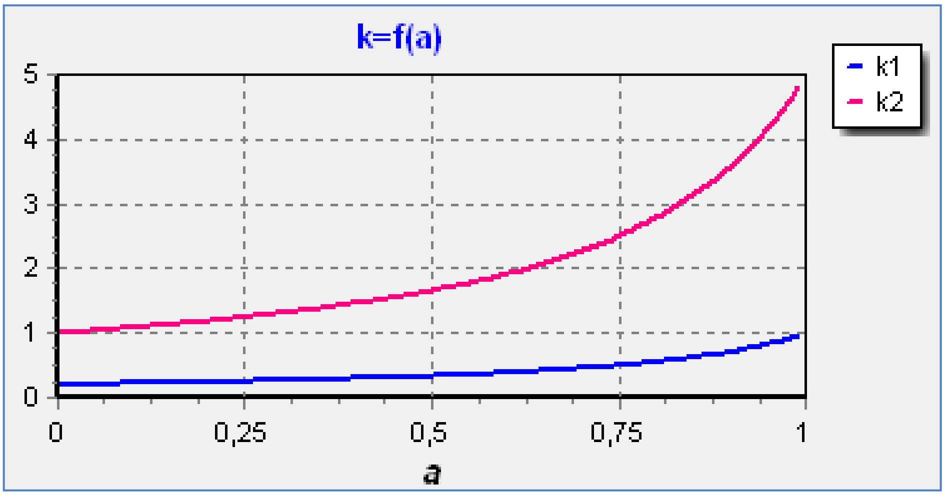

Рисунок 1. Графики эмпирических коэффициентов.

Одной из возможностей организации имитационного моделирования СМО с учетом наличия приоритетных заявок двух рангов является разделение (блок «Выбор» на Рисунке 2) входящего потока на две очереди в соответствии с процентным соотношением приоритетных и простых заявок. В одну из очередей направляются заявки с приоритетом.

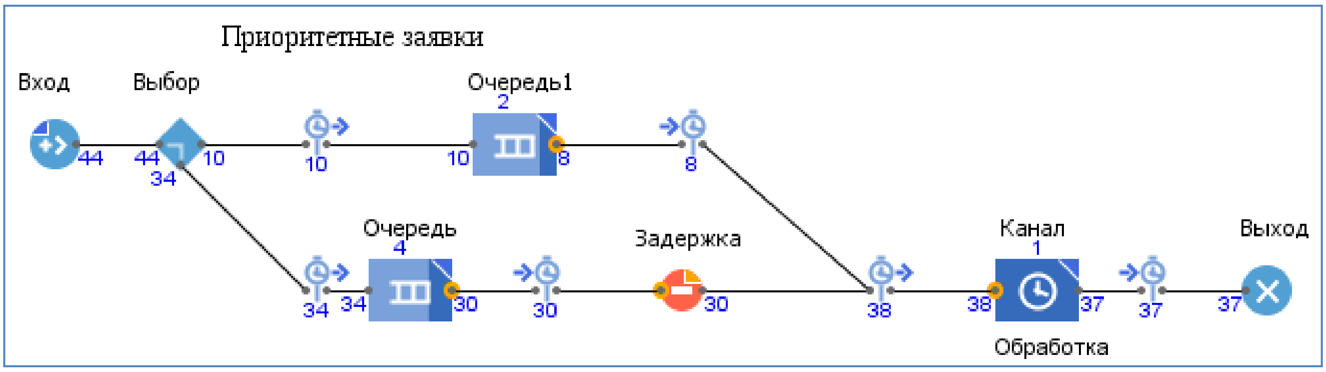

Рисунок 2. Схема СМО с приоритетами в *AnyLogic.*

В этом случае при нахождении в очереди приоритетных заявок альтернативная очередь блокируются до тех пор, пока обрабатывается эти заявки (блок «Задержка» на Рисунке 2).

Таким же образом можно организовать обслуживание многоранговых (три и более) приоритетных заявок.

Более унифицированным методом является подход, основанный на явном задании для каждой заявки ее приоритета. Тогда из очереди на обслуживание в первую очередь забираются все пришедшие к текущему моменту времени заявки с наиболее высоким приоритетом и т. д.

На Рисунке 3 представлен фрагмент имитационной модели в среде *AnyLogic*. Поток с заявками наивысшего приоритета обозначен как «Красный».

Особую практическую значимость имеет СМО в которых время обслуживания заявок в канале различно. В этом случае необходимо решать оптимизационную задачу [2] по упорядочению потока заявок на обслуживание для минимизации среднего (суммарного) времени ожидания в очереди и оперативно формировать систему приоритетов.

В более общем случае для минимизации стоимостных затрат за время ожидания порядок обслуживания должен учитывать не только длительность обслуживания, но и «стоимость «простоя» (C) заявки в единицу времени. Можно показать, что, в данном случае приоритеты в обслуживании заявок должны быть упорядочены по условию:

$$(\mu_1 C_1) > (\mu_2 C_2) > (\mu_2 C_3) > \cdots .$$

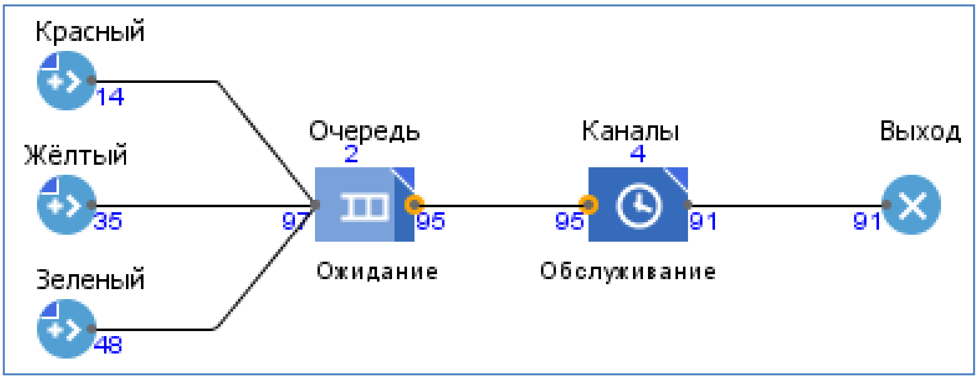

Рисунок 3. Модель трехрангового потока с приоритетами.

Результаты и их обсуждение

Решим задачу определения времени ожидания обслуживания при простейших потоках заявок. Пусть [3] в одноканальную систему поступает простейший поток заявок λ, обслуживание распределено по показательному закону со средней длительность $t_{\text{обсл}} = 0,5$ и коэффициент загрузки системы $\rho = 0,8$, тогда в соответствии с формулой (1) средняя длительность ожидания обслуживания составит:

$$T_q = \frac{\rho}{\mu - \lambda} = \frac{0,8}{2 - 1,6} = 2$$

На Рисунке 4 приведены информация о модельных и предельных значениях времени ожидания в очереди. Для расчетов использовалась среда имитационного моделирования *AnyLogic* [4].

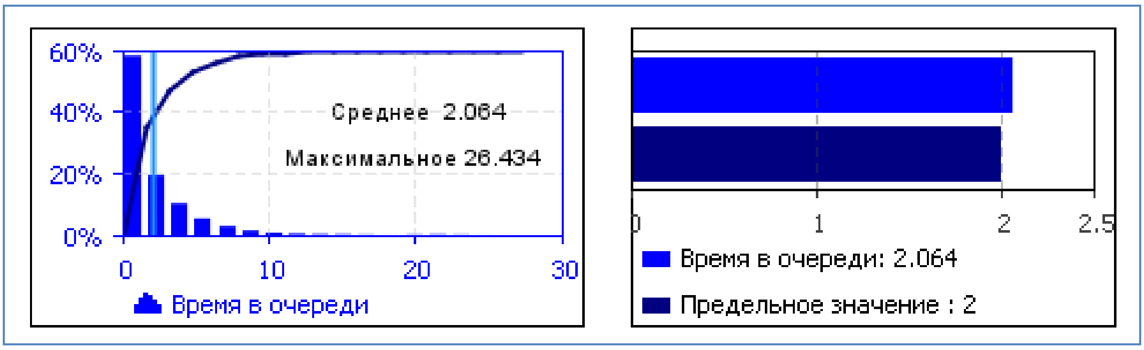

Рисунок 4. Результаты имитации простейшего потока заявок.

Исследуем поток с приоритетами [3]. Пусть в соответствии с формулой (2) 20% $(a = 0,2)$ заявок имеют право на приоритетное обслуживание, тогда:

$$t_1 = \frac{1-\rho}{1-a\rho} T_q = \frac{1-0,8}{1-0,2 \cdot 0,8} \cdot 2 \approx 0,48;\ t_2 = \frac{t_1}{1-\rho} \approx 2,38.$$

На Рисунке 5 представлена гистограмма распределения времени ожидания в очереди для потока с высшим приоритетом и сравнение экспериментального времени с эмпирической оценкой (2).

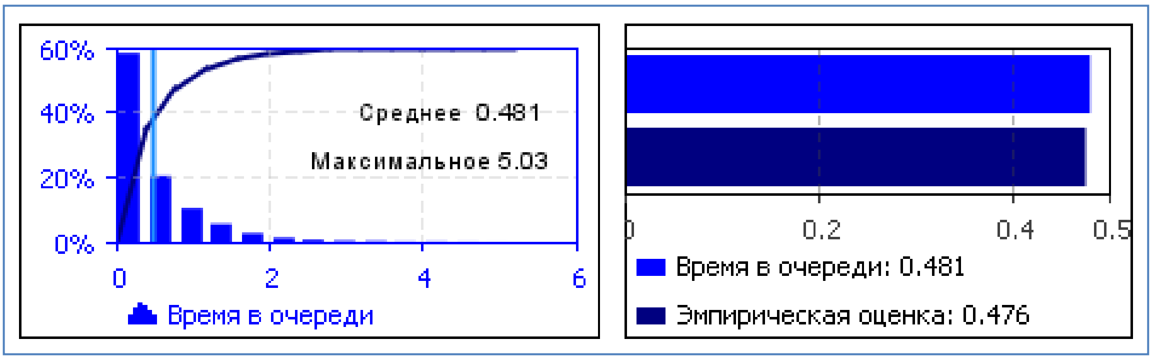

Рисунок 5. Характеристики времени ожидания для заявок с приоритетом.

На Рисунке 6 представлены аналогичные показатели для потока заявок без приоритета.

Результаты моделирования свидетельствуют об удовлетворительности эмпирических оценок (2).

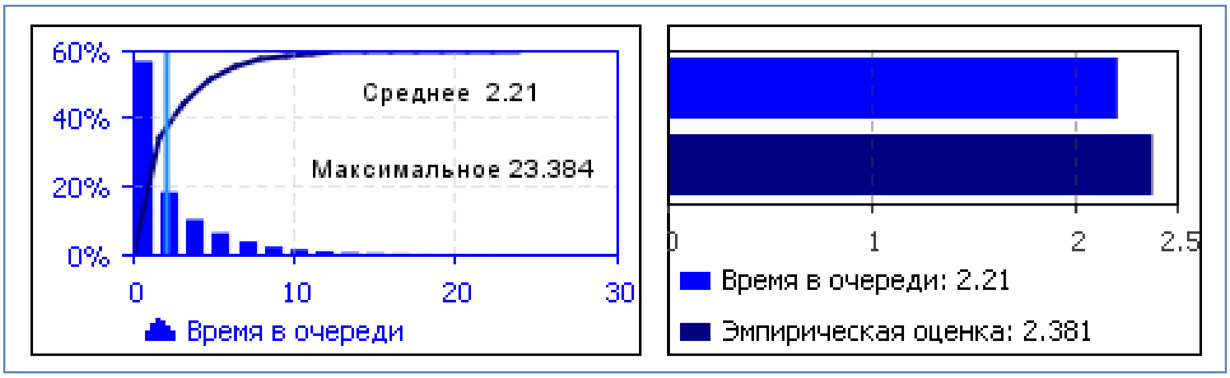

Рисунок 6. Характеристики времени ожидания для заявок без приоритета.

Как уже отмечалось универсальным способом организации обработки заявок является явное задание приоритета при формировании заявки. Проведем моделирование СМО в соответствии со схемой, представленной на Рисунке 3.

Пусть число каналов $n = 4$, наивысший приоритет имеют «красные» заявки.

Интервал времени между прибытиями задан треугольными распределениями (соответственно для «красных», «желтых» и «зеленых заявок»):

$$A_1 = (2,5;\ 2,75;\ 3);\ A_2 = (1;\ 1,25;\ 1,5);\ A_3 = (0,8;\ 0,9;\ 1)$$

На Рисунке 7 представлен фрагмент имитационного моделирования функционирования СМО. В данный момент времени в очереди две «зеленые» заявки, а обслуживаются в каналах одна «красная», одна «желтая» и две «зеленых заявки» (Рисунок 3).

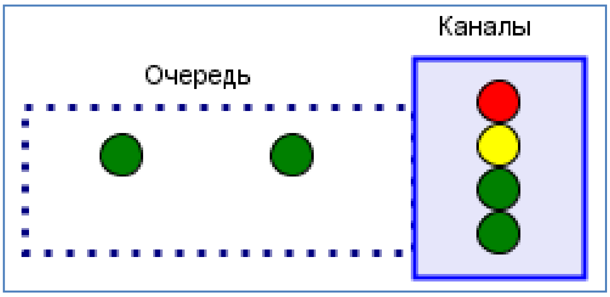

Рисунок 7. Текущее состояние системы.

Очевидно, при освобождении одного из каналов в первую очередь на обслуживание будут забираться «красные» заявки, если их в очереди не окажется, то «желтые» и, наконец, «зеленые», обладающие самым низким приоритетом.

Рисунок 8 содержит информацию об изменении среднего времени ожидания заявок в очереди.

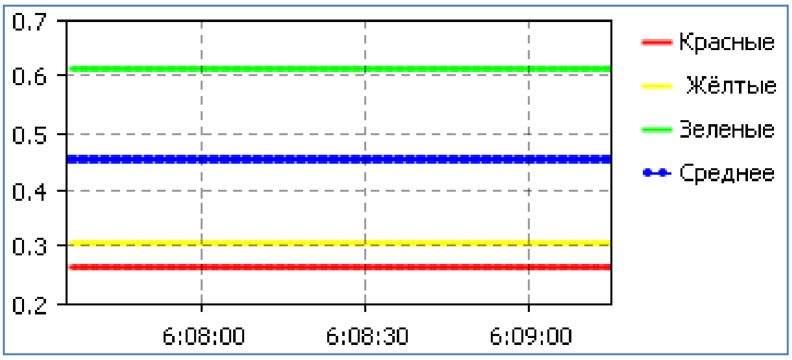

Рисунок 8. Временной график изменения среднего времени ожидания.

В соответствии с представленным графиком среднее время ожидания в очереди для «красных» заявок составляет 0,263, для «желтых» — 0,305, а для «зеленых» — 0,614.

В данном случае продолжительность обслуживания была постоянной $B = (1,25; 1,8; 2,2)$ для любых заявок вне зависимости от их ранга.

Исследуем влияние установки приоритетов на среднюю продолжительность пребывания заявок в очереди при разных временах обслуживания заявок.

Пусть интервал времени между прибытиями одинаков для всех трех групп заявок: $A = (1; 1,25; 1,5)$.

Продолжительность обслуживания определяется различными треугольными законами распределения:

$$B_1 = (1; 1,1; 1,2); B_2 = (1,25; 1,625; 2); B_3 = (2; 2,25; 2,5)$$

т. е. быстрее обслуживаются заявки из первой группы.

Если положить, что наивысшим приоритетом обладают заявки первой группы («быстрые заявки»), то соотношение времен пребывания (по уменьшению приоритета) в очереди будет таким:

$$\frac{(t_1 \vdots t_2 \vdots t_3)}{\bar{t}} = \frac{(0,223 \vdots 0,455 \vdots 1,124)}{0,601},$$

здесь $\bar{t}$ — среднее время пребывания заявок в очереди.

Если ранги групп заявок сменить на противоположные (первыми будут обслуживаться заявки с наибольшим временем обслуживания), то получим следующие результаты:

$$\frac{(t_1 \vdots t_2 \vdots t_3)}{\bar{t}} = \frac{(0,257 \vdots 0,462 \vdots 1,573)}{0,764},$$

т. е. среднее время пребывания в очереди увеличилось на 27%.

Соотношение времен пребывания заявок в СМО с изначальным приоритетом:

$$\frac{(1,323 \vdots 2,077 \vdots 3,374)}{2,258},$$

и после изменения приоритетов:

$$\frac{(2,507 \vdots 2,086 \vdots 2,673)}{2,422}.$$

Выводы

Результаты выполненного исследования позволяют производить моделирование, анализ и оптимизацию функционирования систем массового обслуживания с заявками, обладающими приоритетом.

Исследуемая методология моделирования является унифицированной и допускает возможность обобщения полученных положений на системы обслуживания с более сложными дисциплинами обслуживания — системы с ограниченной длительностью ожидания [5, 6].

Практическая апробация концепции подхода к моделированию СМО с не простейшими дисциплинами организации очереди осуществлена на базе среды моделирования *AnyLogic*, которая позволяет реализовывать все известные парадигмы имитационного моделирования.

Список литературы:

1. Вентцель Е. С. Исследование операций: задачи, принципы, методология. М.: Наука, 1980. 208 с.

2. Осипов Г. С. Оптимизация одноканальных систем массового обслуживания с неограниченной очередью // Бюллетень науки и практики. Электрон. журн. 2016. №9 (10). С. 63–71. Режим доступа: http://www.bulletennauki.com/osipov-gs (дата обращения 15.09.2016). DOI: 10.5281/zenodo.154304.

3. Пьяных С. М. Экономико–математические методы оптимального планирования работы речного транспорта. М.: Транспорт, 1988. 253 с.

4. Осипов Г. С. Исследование систем массового обслуживания с ожиданием в AnyLogic // Бюллетень науки и практики. Электрон. журн. 2016. №10 (11). С. 139–151. Режим доступа: http://www.bulletennauki.com/osipov-g-s (дата обращения 15.10.2016). DOI: 10.5281/zenodo.161072.

5. Osipov G., Osipova E. The study of queuing systems with “impatient” requests // International Scientific Conference “Science. Research. Practice”. Themed collection of papers from international conferences by HNRI “National development”. October 2016. St. Petersburg, HNRI “National development”, 2016. P. 41–44.

6. Осипов Г. С. Системы массового обслуживания с ограниченной длительностью ожидания // Бюллетень науки и практики. Электрон. журн. 2016. №12 (13). С. 28–36. Режим доступа: http://www.bulletennauki.com/osipov-g-1 (дата обращения 15.12.2016). DOI: 10.5281/zenodo.204623.

References:

1. Ventzel E S. Operations research: tasks, principles, methodology. Moscow, Nauka, 1980, 208 p.

2. Osipov G. Optimization of single–channel queuing system with the unlimited queue. Bulletin of Science and Practice. Electronic Journal, 2016, no. 9 (10), pp. 63–71. Available at: http://www.bulletennauki.com/osipov-gs, accessed 15.09.2016. (In Russian). DOI: 10.5281/zenodo.154304.

3. Pyanyh S. M. Economic–mathematical methods of optimum scheduling of river transport. Moscow, Transport, 1988. 253 p.

4. Osipov G. The study of queuing systems with waiting in AnyLogic. Bulletin of Science and Practice. Electronic Journal, 2016, no. 10 (11), pp. 139–151. Available at: http://www.bulletennauki.com/osipov-g-s, accessed 15.10.2016. DOI: 10.5281/zenodo.161072.

5. Osipov G., Osipova E. The study of queuing systems with “impatient” requests. International Scientific Conference “Science. Research. Practice”. Themed collection of papers from international conferences by HNRI “National development”. October 2016. St. Petersburg, HNRI “National development”, 2016, pp. 41–44.

6. Osipov G. Queuing systems with limited waiting times. Bulletin of Science and Practice. Electronic Journal, 2016, no. 12 (13), pp. 28–36. Available at: http://www.bulletennauki.com/osipov-g-1, accessed 15.12.2016. DOI: 10.5281/zenodo.204623.

Работа поступила в редакцию 10.12.2016 г.

Принята к публикации 14.12.2016 г.