УДК 519.872

Расчёт многоканальных СМО с отрицательными заявками

Р. С. Хабаров[1], к. т. н.
E-mail: xabarov.r@yandex.ru

В. А. Лохвицкий[1], д. т. н.
E-mail: lokhv_va@mail.ru

[1]Военно-космическая академия им. А. Ф. Можайского, Санкт-Петербург, Россия

Аннотация

Предложен численный метод расчёта многоканальных систем массового обслуживания с произвольным распределением времени обслуживания (аппроксимируемым гиперэкспоненциальным H_2 с комплексными параметрами) и двумя дисциплинами обработки отрицательных заявок: **RCS** (удаление заявки на обслуживании) и **катастрофы** (мгновенная очистка системы). Метод расширяет алгоритм Такахаси–Таками модификацией матриц переходов микросостояний и позволяет получать не только стационарные вероятности состояний, но и временные характеристики — начальные моменты времени ожидания и пребывания. Получены результаты для вариантов поведения положительных заявок при поступлении отрицательной заявки: при уходе из системы, при направлении на повторное обслуживание с перегенерацией длительности и при повторном обслуживании *без* пересемплирования (с фиксированной длительностью на заявку).

Ключевые слова: теория массового обслуживания, отрицательные заявки, катастрофы, RCS, метод Такахаси–Таками, фазовые аппроксимации, H_2, преобразование Лапласа–Стилтьеса.

Calculation of Multi-channel Queueing Systems with Negative Customers

Abstract. A numerical method is proposed for calculating multi-channel queueing systems with an arbitrary service time distribution (approximated by hyperexponential H_2 with complex parameters) and two negative customer handling disciplines: **RCS** (Remove Customer in Service) and **disasters** (instantaneous system clearing). The method extends the Takahashi–Takami algorithm by modifying microstate transition matrices and yields both stationary state probabilities and time-based performance measures — initial moments of waiting and sojourn times. Results are obtained for positive customers' behavior upon a negative arrival: removal from the system, restart with resampling, and restart without resampling (fixed service requirement per job).

Keywords: queueing theory, negative customers, disasters, RCS, Takahashi–Takami method, phase approximations, H_2, Laplace–Stieltjes transform.

1. Введение

Метод Такахаси–Таками [1] изначально предложен для общего класса систем $GI/G/c$; в монографии Рыжикова [2] подробно изложены многоканальные фазовые системы и итерационное решение векторно-матричных уравнений баланса с блоками A_j, B_j, C_j, D_j. Метод применим к достаточно широкому классу систем (в том числе $M/H_k/n$, $H_k/M/n$ и др.); в данной работе рассматриваются системы с *пуассоновским входом* интенсивности λ и *распределением времени обслуживания* B, аппроксимируемым гиперэкспоненциальным H_2 с комплексными параметрами, что позволяет описывать системы с произвольным коэффициентом вариации при n каналах. Предположения о пуассоновском потоке заявок и произвольном обслуживании являются достаточно общими для многих реальных систем [3; 4].

В последние десятилетия значительный интерес вызывает исследование систем с отрицательными заявками (negative customers), введёнными в работах Геленбе [5; 6]. Отрицательные заявки при поступлении могут удалять положительные заявки в соответствии с различными дисциплинами. В работе рассматриваются две дисциплины обслуживания: **RCS (Remove Customer in Service)** — воздействие на одну заявку в обслуживании (удаление); **катастрофы** — мгновенная очистка системы от всех положительных заявок. Далее через δ обозначается интенсивность потока отрицательных заявок, W и V — времена ожидания в очереди и пребывания в системе (случайные величины), w_1 и v_1 — их первые начальные моменты (математические ожидания).

Исследования СМО с отрицательными заявками обычно связывают с работами Э. Геленбе (G-сети с «сигналами»); ключевой результат — существование стационарных распределений в мультипликативной форме (product-form) при марковских предпосылках и ограничениях на механизмы удаления [5; 6]. G-сети обобщались на случаи зависящих от состояния интенсивностей, пакетных перемещений и многоклассовых заявок, сохраняя мультипликативную форму стационарных распределений при экспоненциальных предпосылках. Обзор подходов к сигналам и стратегиям удаления приведён в [7]. Для одноканальных систем $M/G/1$ сформировалась аналитическая ветка (дополнительная переменная, интегральные уравнения), позволяющая описывать стационарные распределения и временные характеристики [8]; развивались также факторизация Винера–Хопфа [9] и интерпретации удаления работы [10]. Отдельная линия — *катастрофы*: для $M/G/1$ получены аналоги формулы Поллачека–Хинчина для стационарного распределения числа заявок в системе и времени пребывания, подчёркивающие роль преобразования Лапласа–Стилтьеса (ПЛС), $T^*(s) = \mathbb{E}[e^{-sT}]$, $s \geq 0$, и корректной трактовки «обрезания» траектории катастрофой [11]. Модели с катастрофами обобщались на BMAP/SM/1 [12], дисциплину с разделением процессора [13], дискретное время [14], повторные попытки и отказы прибора [15; 16]. Для многоканальных моделей встречаются марковские частные случаи [17]; интерес к системам с отрицательными заявками и катастрофами сохраняется в последние годы (дискретное время, параллельные очереди, гетерогенные приборы), однако расчёт при произвольном обслуживании в многоканальном случае остаётся нетривиальным.

В данной работе предложено расширение метода Такахаси–Таками для многоканальных СМО с отрицательными заявками: вводятся правила формирования матриц переходов для дисциплин RCS и катастроф, а также алгоритмы вычисления первых начальных моментов времени ожидания и пребывания w_1 и v_1, согласованные с имитационным моделированием. В результате метод позволяет получать стационарные вероятности состояний и временные характеристики в рамках единого алгоритма; обслуживание аппроксимируется H_2 при любом коэффициенте вариации, при необходимости с комплексными параметрами.

2. Постановка задачи

Рассматривается многоканальная система массового обслуживания с пуассоновскими потоками положительных (интенсивность λ) и отрицательных (интенсивность δ) заявок, временем обслуживания B, аппроксимируемым гиперэкспоненциальным распределением H_2 с заданным средним и коэффициентом вариации, и числом каналов $n \geq 1$. Коэффициент загрузки $\rho = n^{-1}\lambda\,\mathbb{E}[B]$ предполагается меньше единицы.

Исследуются две дисциплины обработки отрицательных заявок:

1. **RCS** (Remove Customer in Service) — удаление одной заявки, находящейся в обслуживании;

2. **катастрофы** — мгновенная очистка системы от всех положительных заявок.

Для каждой дисциплины рассматриваются два варианта поведения вытесненных положительных заявок: уход из системы и направление на повторное обслуживание (рестарт).

Пространство состояний усекается по уровню $i \leq N - 1$ (N — максимальный уровень); сходимость итераций контролируется по изменению $x[i]$ или $t[i]$ с заданным допуском.

3. Метод Такахаси–Таками: базовые принципы

3.1. Структура микросостояний

Метод основан на представлении системы $M/H_2/n$ в виде марковского процесса с дискретным пространством состояний. Состояние описывается:

- i — число заявок в системе (уровень, level);

- j — число заявок, находящихся в фазе 2 H_2-аппроксимации (с интенсивностью μ_2; микросостояние, microstate).

Для уровня $i < n$ число микросостояний равно $i+1$, для $i \geq n$ постоянно и равно $n+1$.

3.2. Матрицы переходов

Процесс описывается четырьмя типами матриц переходов между микросостояниями соседних уровней: A (поступление), B (уход вниз), C (горизонтальные переходы), D (диагональная матрица интенсивностей выхода).

Пусть $y_1+y_2 = 1$, где y_1 и y_2 — вероятности выбора фаз 1 и 2 в H_2-аппроксимации, а μ_1, μ_2 — интенсивности экспоненциального обслуживания в фазах. На уровне $i < n$ все заявки находятся в обслуживании, а на уровне $i \geq n$ в обслуживании ровно n заявок, остальные ждут в очереди.

3.2.1. Матрица A (поступление положительных заявок)

Для $i < n$ при приходе новая заявка немедленно занимает свободный канал и выбирает фазу:

$$(i, j) \to (i+1, j) \text{ с интенсивностью } \lambda y_1, \qquad (i, j) \to (i+1, j+1) \text{ с интенсивностью } \lambda y_2.$$

Для $i \geq n$ пришедшая заявка становится в очередь, и микросостояние j (число заявок в фазе 2 среди обслуживаемых) не меняется:

$$(i, j) \to (i+1, j) \text{ с интенсивностью } \lambda.$$

3.2.2. Матрица B (уход на уровень вниз)

Для $i \le n$ уход вниз обусловлен завершением обслуживания одной из i заявок:

$(i,j) \to (i{-}1,j)$ с интенсивностью $(i{-}j)\mu_1$, $\qquad (i,j) \to (i{-}1,j{-}1)$ с интенсивностью $j\mu_2$.

Для $i > n$ уход вниз сопровождается тем, что освободившийся канал немедленно занимает заявка из очереди, выбирая фазу 1/2 с вероятностями y_1/y_2. Поэтому для $j \in \{0,1,\ldots,n\}$:

$$\begin{aligned}
(i,j) &\to (i-1,j) \text{ с интенсивностью } (n-j)\mu_1 y_1 + j\mu_2 y_2,\\
(i,j) &\to (i-1,j+1) \text{ с интенсивностью } (n-j)\mu_1 y_2, \qquad (j<n),\\
(i,j) &\to (i-1,j-1) \text{ с интенсивностью } j\mu_2 y_1, \qquad (j>0).
\end{aligned}$$

3.2.3. Матрица D (интенсивности выхода)

Диагональные элементы задаются суммарной интенсивностью выхода из микросостояния. Для $i < n$:

$$D(i,j) = \lambda + (i-j)\mu_1 + j\mu_2,$$

а для $i \ge n$:

$$D(i,j) = \lambda + (n-j)\mu_1 + j\mu_2.$$

Примеры для $n = 3$, уровень $i = 2$: матрицы A_1 (переход 1→2), B_2 (переход 2→1) и D_2 (диагональ выхода) можно компактно записать на одном уровне:

$$A_1 = \begin{pmatrix} \lambda y_1 & \lambda y_2 & 0 \\ 0 & \lambda y_1 & \lambda y_2 \end{pmatrix} \quad B_2 = \begin{pmatrix} 2\mu_1 & 0 \\ \mu_2 & \mu_1 \\ 0 & 2\mu_2 \end{pmatrix} \quad D_2 = \begin{pmatrix} \lambda + 2\mu_1 & 0 & 0 \\ 0 & \lambda + \mu_1 + \mu_2 & 0 \\ 0 & 0 & \lambda + 2\mu_2 \end{pmatrix}.$$

3.3. Итерационный алгоритм

Алгоритм Такахаси–Таками основан на итерационном вычислении вспомогательных величин для уровней $i = 0,1,\ldots,N-1$ и последующем восстановлении вероятностей уровней $p[i]$. В типичной реализации используются:

- вспомогательные матрицы $G_i = (D_i - C_i)^{-1}$, а также произведения $A_{i-1}G_i$ и $B_{i+1}G_i$;
- итерации по уровням для обновления векторов $t[i]$ и скаляров $x[i], z[i]$ до сходимости;
- восстановление вероятностей уровней по рекуррентной формуле $p[i] = p[i-1]x[i-1]$ при нормировке.

Построение матриц: для каждого уровня i по микросостоянию j заполняются блоки: (1) A — переходы вверх по правилам для $i < n$ и $i \ge n$; (2) B — переходы вниз из завершений обслуживания и выбора фазы при $i > n$; (3) C — горизонтальные переходы; (4) диагональ D так, чтобы сумма по строкам в $A + B + C - D$ равнялась нулю. Размеры блоков определяются числом микросостояний на соседних уровнях.

Выбор N, сходимость, сложность: N выбирают так, чтобы $p[N-1]x[N-1] < \varepsilon$. Итерации прекращают при $\max_i |x[i]^{\text{new}} - x[i]^{\text{old}}| < \varepsilon$. Время одного шага итерации — $O(N(n+1)^3)$, память — $O(N(n+1)^2)$.

3.4. Временные характеристики в базовой модели

Для базовой системы без отрицательных заявок моменты ожидания $w^{(k)}$ могут вычисляться по вероятностям уровней:

$$w^{(k)} = \frac{1}{\lambda^{k+1}} \sum_{j=k+1}^{N-n} j(j-1)\cdots(j-k)\, p[n+j],$$

где $k=0$ соответствует первому начальному моменту $w_1 = \mathbb{E}[W]$, $k=1$ — второму $w_2 = \mathbb{E}[W^2]$ и т. д. Моменты пребывания при отсутствии отрицательных заявок получаются стандартной свёрткой моментов ожидания и обслуживания. Для систем с отрицательными воздействиями моменты вычисляются либо по вероятностям уровней (в редуцированных аппроксимациях к $M/G/n$), либо восстанавливаются из ПЛС численным дифференцированием: $(-1)^k T^{*(k)}(0)$ аппроксимируется конечными разностями по s в окрестности нуля (шаг $h \sim 10^{-5}$–10^{-4}); устойчиво ограничены первые 2–4 момента.

4. Расширение метода для систем с отрицательными заявками

4.1. Общая идея

Расширение основано на модификации матриц переходов базового метода без изменения общей структуры итерационного процесса. В зависимости от дисциплины отрицательного воздействия переопределяются правила построения матриц B и D, а для катастроф — также структура микросостояний.

4.2. Дисциплина RCS (Remove Customer in Service)

При дисциплине RCS отрицательная заявка воздействует на одну заявку в обслуживании, выбирая занятый канал равновероятно. На уровнях $i \le n$ (очереди нет) отрицательные события добавляют к уходам вниз интенсивности

$$(i,j) \to (i-1,j) \text{ с интенсивностью } \delta\frac{i-j}{i}, \qquad (i,j) \to (i-1,j-1) \text{ с интенсивностью } \delta\frac{j}{i},$$

то есть эквивалентны дополнительной «опасности удаления» δ/i для каждой обслуживаемой заявки. Диагональные элементы дополняются интенсивностью отрицательных приходов:

$$D_{\mathrm{RCS}}(i,j) = D_{\mathrm{base}}(i,j) + \delta.$$

На уровнях $i > n$ отрицательное событие также удаляет одну из n заявок в обслуживании, после чего канал немедленно занимает заявка из очереди и выбирает фазу. В рамках структуры Такахаси–Таками это приводит к добавлению интенсивностей, зависящих от микросостояния j, в матрицу B_{n+1} (уровень, соответствующий полной занятости n каналов), а для всех уровней $i \ge n+1$ используются те же матрицы, поскольку число микросостояний стабилизируется и равно $n+1$.

4.3. Дисциплина катастроф

При базовой дисциплине обслуживания катастрофа очищает систему (удаляет все положительные заявки). Для согласования с матричной структурой метода вводится искусственное микросостояние D («состояние катастрофы») на каждом уровне $i > 0$. При поступлении отрицательной заявки система переходит в состояние D, а затем с большой искусственной интенсивностью $\gamma \gg \mu_1, \mu_2, \lambda$ — в нулевой уровень, что моделирует мгновенную очистку.

Структура микросостояний модифицируется так:

- уровень 0: одно состояние $(0,0)$;
- уровни $1 \le i < n+1$: $1+(i+1)$ микросостояний (первое — D, затем $(i,0),\dots,(i,i)$);
- уровни $i \ge n+1$: $1+(n+1)$ микросостояний (первое — D, затем $(i,0),\dots,(i,n)$).

Поступление положительных заявок (A) «пропускает» состояние D. Для уровня 0:

$$(0,0) \to (1,0) \text{ с интенсивностью } \lambda y_1, \qquad (0,0) \to (1,1) \text{ с интенсивностью } \lambda y_2.$$

Для уровня $i > 0$ положительный приход переводит (i,j) в соответствующее обычное состояние уровня $i+1$, сохраняя описанную выше структуру (без попадания в D).

Переход в катастрофическое состояние происходит из любого обычного микросостояния уровня $i > 0$ с интенсивностью δ. Состояние D имеет уход в $(0,0)$ с интенсивностью γ, поэтому диагональный элемент для этого состояния (суммарная интенсивность выхода) равен γ. Для уровня $i=1$ при $n=3$ микросостояния суть D, $(1,0)$, $(1,1)$; матрица B_1 имеет размер 3×1 (переход в уровень 0).

4.4. Примеры матриц переходов

Переходы по завершению обслуживания (интенсивности μ_1, μ_2) в базе $\mathrm{M/H_2}/n$ стандартны, поэтому ниже для каждой дисциплины приведены только переходы, возникающие из отрицательных воздействий (интенсивность δ и производные; для катастроф также γ).

4.4.1. RCS (удаление из системы)

При построении матриц B_i к «стандартным» интенсивностям обслуживания добавляются интенсивности отрицательных воздействий. В качестве наглядного примера приведём матрицу B_2 для дисциплины RCS при $n \ge 2$ (уровень $i=2$, микросостояния $j=0,1,2$ на уровне 2 и $j=0,1$ на уровне 1). Тогда

$$B_2 = \begin{pmatrix} 2\mu_1+\delta & 0 \\ \mu_2+\delta/2 & \mu_1+\delta/2 \\ 0 & 2\mu_2+\delta \end{pmatrix},$$

где, например, элемент $(2,1)\to(1,1)$ равен $\mu_1+\delta/2$: он складывается из обслуживания фазы 1 и удаления заявки в фазе 1 отрицательным прибытием (равновероятный выбор канала среди двух занятых). Матрица C_i совпадает с базовой (горизонтальные переходы отсутствуют при удалении из системы).

4.4.2. RCS (повторное обслуживание)

При повторном обслуживании отрицательное воздействие не меняет числа заявок в системе, а лишь прерывает обслуживание одной случайной заявки и возвращает её в очередь с рестартом. Для этой постановки временные характеристики (раздел 4) и все численные результаты (раздел 5) рассчитываются по аппроксимации через эффективное время обслуживания B_{eff}. Матричное описание через горизонтальные переходы в C_i для данного случая в расчётах не используется и здесь не приводится.

4.4.3. Катастрофы (очищение системы)

Для катастроф с очищением вводится дополнительное микросостояние D на каждом уровне $i > 0$. Приведём примеры матриц B_1 и C_1 при $n \geq 1$ (уровень $i = 1$, микросостояния: D, $(1,0)$, $(1,1)$; переход в уровень 0) — блок 3×1 и матрица горизонтальных переходов:

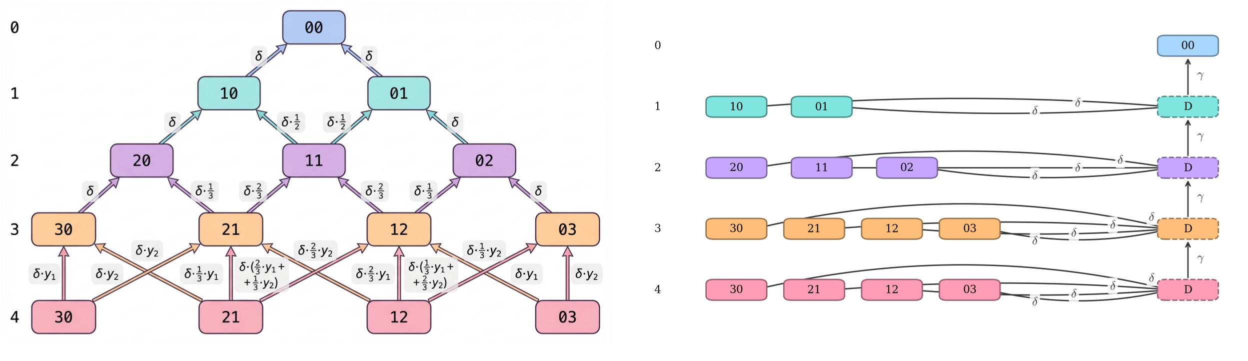
$$B_1 = \begin{pmatrix} \gamma \\ \mu_1 + \delta \\ \mu_2 + \delta \end{pmatrix} \quad C_1 = \begin{pmatrix} 0 & 0 & 0 \\ \delta & 0 & 0 \\ \delta & 0 & 0 \end{pmatrix}.$$

Первая строка соответствует состоянию D: система переходит в $(0,0)$ с интенсивностью $\gamma \gg \mu_1, \mu_2, \lambda$. Вторая и третья строки соответствуют обычным состояниям $(1,0)$ и $(1,1)$: уход в уровень 0 происходит либо по завершению обслуживания (μ_1, μ_2), либо по поступлению катастрофы (интенсивность δ добавляется к соответствующему переходу).

Матрица C_1 моделирует горизонтальные переходы из обычных состояний в катастрофическое состояние D, где элементы $(1,0) \to D$ и $(1,1) \to D$ равны δ (поступление катастрофы мгновенно переводит систему в состояние D). В программной реализации используется эквивалентный подход: переходы по катастрофе включены непосредственно в матрицу B (нисходящие переходы из обычных состояний в уровень ниже), при этом C остаётся нулевой; оба варианта дают одинаковые стационарные вероятности.

4.4.4. Катастрофы (повторное обслуживание)

При повторном обслуживании катастрофа не очищает систему, а прерывает все обслуживания и возвращает заявки в очередь с рестартом. Для этой постановки временные характеристики (раздел 4) и все численные результаты (раздел 5) рассчитываются по аппроксимации через эффективное время обслуживания B_{eff}. Матричное описание через горизонтальные переходы в C_i для данного случая в расчётах не используется и здесь не приводится.



(a) RCS (удаление заявки из обслуживания) (b) Катастрофы (интенсивности δ и γ)

Рис. 1: Переходы, вызванные отрицательными воздействиями, при $n = 3$ (уровни $0, \ldots, n$ для RCS и $0, \ldots, n+1$ для катастроф).

5. Расчёт временных характеристик

5.1. Дисциплина RCS

5.1.1. RCS с удалением из системы

Схема переходов, вызванных отрицательными воздействиями при дисциплине RCS (удаление заявки из обслуживания), приведена на рисунке 1a. Для типичной заявки, находящейся в обслуживании при m занятых каналах, интенсивность удаления аппроксимируется как $r(m) = \delta/m$. При дисциплине «удаление из обслуживания» время в обслуживании:

$$S = \min(B, Y), \qquad Y \sim \mathrm{Exp}(r),$$

и ПЛС времени обслуживания $S^*(s) = \frac{r}{s+r} + \frac{s}{s+r}\beta(s+r)$, где $\beta(s) = \mathbb{E}[e^{-sB}]$. При этом ожидающие в очереди положительные заявки отрицательными воздействиями не удаляются, поэтому W трактуется как время до *начала* обслуживания. Влияние δ на W учитывается косвенно: удаление одной из обслуживаемых заявок освобождает прибор и тем самым может ускорять запуск ожидания.

5.1.2. RCS с повторным обслуживанием

Если отрицательное событие интерпретируется как прерывание обслуживания с возвратом заявки в очередь (рестарт обслуживания и повторная генерация длительности), расчёт сводится к *эффективному времени обслуживания* B_{eff} при пуассоновских рестартах. Пусть B имеет ПЛС $\beta(s)$, рестарты — с интенсивностью r. Для модели «прерывание–повтор» время до завершения либо меньше первого рестарта (вклад в ПЛС даёт $\beta(s+r)$ с учётом условного распределения времени до рестарта), либо после рестарта процесс обновляется; из обновляющего уравнения для ПЛС времени завершения получается:

$$B^*_{\text{eff}}(s) = \beta(s+r) \cdot \frac{s+r}{s + r\,\beta(s+r)}.$$

Для RCS типичная заявка при m занятых каналах имеет интенсивность рестарта $r(m) = \delta/m$; в расчёте r оценивается самосогласованно (итерационно), начиная с приближения $r \approx \delta/n$. Система аппроксимируется стандартной $M/G/n$ без отрицательных выходов с распределением B_{eff}; тогда $V = W + B_{\text{eff}}$, $q = 1$, $V_{\text{served}} \equiv V$, $V_{\text{broken}} \equiv 0$. Моменты W и V вычисляются базовым алгоритмом Такахаси–Таками для $M/G/n$ с распределением B_{eff}.

Повтор без пересемплирования (фиксированная длительность). В ряде приложений естественна дисциплина, при которой длительность обслуживания B семплируется *один раз* для заявки, но при каждом прерывании обслуживание начинается заново и уже выполненная работа теряется. Тогда классическая формула для $B^*_{\text{eff}}(s)$ (рестарт с перегенерацией) неприменима. Для фиксированного $B = b$ и пуассоновских прерываний с интенсивностью r время завершения T имеет моменты

$$\mathbb{E}[T \mid b] = \frac{e^{rb} - 1}{r}, \qquad \mathbb{E}[T^2 \mid b] = \frac{2e^{rb}}{r^2}\left(e^{rb} - 1 - rb\right).$$

Безусловные моменты выражаются через значения производящей функции моментов $M_B(t) = \mathbb{E}[e^{tB}]$ при $t = r, 2r$. В частности, $\mathbb{E}[T] < \infty$ требует конечности $M_B(r)$, а $\mathbb{E}[T^2] < \infty$ — конечности $M_B(2r)$; поэтому при больших δ (и/или больших CV) моменты могут резко возрастать или становиться бесконечными. В численных экспериментах (раздел 5) для этой дисциплины используется ограниченный диапазон δ.

5.2. Дисциплина катастроф

5.2.1. Катастрофы с очищением

Схема переходов, вызванных катастрофическими воздействиями (интенсивности δ и γ), приведена на рисунке 1b. Пусть $Y \sim \mathrm{Exp}(\delta)$ — время до ближайшей катастрофы после прихода типичной заявки. Для получения результатов, согласованных с имитационным моделированием, вводятся гипотетические величины W_0 и $Z_0 = W_0 + B$, соответствующие динамике, в которой катастрофы *после прихода данной заявки* отключены (но состояние при приходе берётся из стационара исходной системы). ПЛС $W_0^*(s)$ вычисляется по стационарным вероятностям уровней $p[k]$ модели с катастрофами и переходным вероятностям базовой системы $M/H_2/n$ (без катастроф): при приходе в состояние с уровнем $k < n$ заявка сразу поступает на обслуживание ($W_0 = 0$); при $k \geq n$ вклад даёт сумма по уровням с весами, зависящими от времени до освобождения места в приборе. Тогда

$$W = \min(W_0, Y), \qquad V = \min(Z_0, Y),$$

и для любой независимой пары T и $Y \sim \mathrm{Exp}(\delta)$ справедливо:

$$\mathbb{E}[e^{-s\min(T,Y)}] = \frac{\delta}{s+\delta} + \frac{s}{s+\delta} T^*(s+\delta).$$

Отсюда:

$$W^*(s) = \frac{\delta}{s+\delta} + \frac{s}{s+\delta} W_0^*(s+\delta), \qquad V^*(s) = \frac{\delta}{s+\delta} + \frac{s}{s+\delta} Z_0^*(s+\delta).$$

Вероятность успешного обслуживания:

$$p_{\text{served}} = \mathbb{P}(Z_0 < Y) = \mathbb{E}[e^{-\delta Z_0}] = Z_0^*(\delta),$$

а ПЛС условных распределений:

$$\mathbb{E}[e^{-sV} \mid \text{served}] = \frac{Z_0^*(s+\delta)}{Z_0^*(\delta)}, \qquad \mathbb{E}[e^{-sV} \mid \text{broken}] = \frac{\delta}{s+\delta} \cdot \frac{1 - Z_0^*(s+\delta)}{1 - Z_0^*(\delta)}.$$

5.2.2. Катастрофы с повторным обслуживанием

При дисциплине «катастрофа прерывает обслуживание всех заявок в каналах и возвращает их в очередь с рестартом» прерванных в смысле удалённых заявок нет: $q = 1$, $V_{\text{served}} \equiv V$, $V_{\text{broken}} \equiv 0$. Эффективное время обслуживания B_{eff} при пуассоновских рестартах с интенсивностью δ (каждая катастрофа прерывает любую заявку в канале) задаётся той же формулой, что и выше, с $r = \delta$:

$$B_{\text{eff}}^*(s) = \beta(s+\delta) \cdot \frac{s+\delta}{s + \delta\,\beta(s+\delta)}.$$

Система в приближении сводится к стандартной $M/G/n$ без отрицательных выходов с распределением B_{eff}; $V = W + B_{\text{eff}}$, моменты получаются численным дифференцированием ПЛС (с гамма-аппроксимацией $\beta(s)$ по первым двум моментам B).

Повтор без пересемплирования (фиксированная длительность). Аналогично RCS, можно рассмотреть дисциплину, когда при каждой катастрофе обслуживание всех заявок в каналах начинается заново, но исходные длительности обслуживания сохраняются (каждая заявка имеет фиксированный B). В этом случае приведённые выше формулы для моментов $T \mid B = b$ применимы с $r = \delta$; далее система аппроксимируется эквивалентной $M/G/n$ с эффективными моментами B_{eff}, полученными из H_2-аппроксимации B. Для корректного сравнения с ИМ ограничиваем δ так, чтобы существовали первые 1–2 момента B_{eff}.

6. Численные результаты и сравнение с имитационным моделированием

6.1. Параметры тестов

В качестве базового примера использовались параметры: $n = 3$, $\lambda_{\text{pos}} = 1{,}0$, $\delta = 0{,}3$, $\rho = 0{,}7$, коэффициент вариации времени обслуживания $CV = 1{,}2$, размер выборки в имитационном моделировании 200 000. Дополнительно, чтобы показать применимость метода при различных коэффициентах вариации времени обслуживания, в таблице 1 приведены численные значения для $CV = 0{,}8$ и $CV = 1{,}2$. Для дисциплины «повтор без пересемплирования» (фиксированная длительность обслуживания на заявку) диапазон δ в графиках ограничивается: из-за возникновения дополнительной работы распределение времени завершения имеет тяжёлый хвост, а моменты зависят от существования $M_B(r)$ и $M_B(2r)$ (раздел 4). Поэтому для этой дисциплины в экспериментах используются меньшие значения δ по сравнению с базовым примером.

6.2. Сравнение численного метода и ИМ по w_1 и v_1

В таблице 1 приведены первые начальные моменты времени ожидания (w_1) и пребывания (v_1) для четырёх вариантов обработки отрицательных заявок (удаление/очищение и повторное обслуживание) при двух значениях коэффициента вариации времени обслуживания; для каждого значения CV приведено сравнение с имитационным моделированием (ИМ).

	w_1			v_1		
Дисциплина	Числ	ИМ	Отн.%	Числ	ИМ	Отн.%
Коэффициент вариации $CV = 0{,}8$						
RCS (удаление)	0.4205	0.4089	2.82	2.0952	2.0790	0.78
RCS (повторное обслуж.)	1.1906	1.1523	3.33	3.3685	3.3938	−0.75
Катастрофы (очищение)	0.1964	0.1922	2.17	1.4853	1.4814	0.27
Катастрофы (повторное обслуж.)	1.8268	1.8552	−1.53	4.1527	4.2008	−1.15
Коэффициент вариации $CV = 1{,}2$ (базовое значение)						
RCS (удаление)	0.3881	0.4002	−3.03	1.9127	1.9230	−0.54
RCS (повторное обслуж.)	1.0842	1.0451	3.74	3.0981	3.0900	0.26
Катастрофы (очищение)	0.1613	0.1693	−4.73	1.3096	1.3174	−0.59
Катастрофы (повторное обслуж.)	0.7583	0.7869	−3.64	2.6441	2.6800	−1.34

Таблица 1: Первые начальные моменты времени ожидания w_1 и пребывания v_1 при фиксированных $n = 3$, $\lambda_{\text{pos}} = 1{,}0$, $\delta = 0{,}3$, $\rho = 0{,}7$ и различных коэффициентах вариации времени обслуживания. Числ — численный расчёт, ИМ — имитационное моделирование, Отн.% — относительная погрешность.

В таблице 2 приведены среднеквадратичные отклонения σ_W и σ_V для тех же дисциплин обслуживания отрицательных заявок и значений коэффициента вариации.

Для дисциплины «повтор без пересемплирования» (фиксированная длительность обслуживания на заявку) базовое значение $\delta = 0{,}3$ использовать нельзя: моменты времени завершения быстро растут и могут не существовать (см. раздел 4). Поэтому отдельное сравнение численного метода и ИМ для этой дисциплины приведено в таблице 3 при меньшей интенсивности δ.

Заметны несколько устойчивых закономерностей. Во-первых, дисциплины с *удалением* (RCS) и *очищением* (катастрофы) дают наименьшие значения w_1 и v_1: отрица-

Дисциплина	σ_W			σ_V		
	Числ	ИМ	Отн.%	Числ	ИМ	Отн.%
Коэффициент вариации $CV = 0{,}8$						
RCS (удаление)	0.9291	0.9106	2.03	1.7136	1.6963	1.02
RCS (повторное обслуж.)	1.9534	1.9765	−1.17	2.6513	2.6955	−1.64
Катастрофы (очищение)	0.5615	0.5503	2.04	1.2682	1.2611	0.57
Катастрофы (повторное обслуж.)	2.7441	2.8083	−2.29	3.3921	3.4596	−1.95
Коэффициент вариации $CV = 1{,}2$ (базовое значение)						
RCS (удаление)	1.0082	1.0398	−3.04	2.0878	2.0935	−0.27
RCS (повторное обслуж.)	2.0361	2.0469	−0.53	3.1204	3.1239	−0.11
Катастрофы (очищение)	0.5599	0.5833	−4.01	1.4312	1.4404	−0.64
Катастрофы (повторное обслуж.)	1.5502	1.5761	−1.65	2.6529	2.6554	−0.10

Таблица 2: Среднеквадратичные отклонения σ_W и σ_V при фиксированных $n = 3$, $\lambda_{\text{pos}} = 1{,}0$, $\delta = 0{,}3$, $\rho = 0{,}7$ и различных коэффициентах вариации времени обслуживания.

Дисциплина	w_1			v_1		
	Числ	ИМ	Отн.%	Числ	ИМ	Отн.%
Коэффициент вариации $CV = 0{,}8$						
RCS (повтор без пересемпл.)	1.2077	1.1569	4.39	3.3783	3.3433	1.04
Катастрофы (повтор без пересемпл.)	1.8599	2.0551	-9.50	4.1563	4.3510	-4.47
Коэффициент вариации $CV = 1{,}2$						
RCS (повтор без пересемпл.)	2.0397	2.0017	1.90	4.2459	4.2367	0.22
Катастрофы (повтор без пересемпл.)	4.4891	4.8547	-7.53	6.9005	7.2376	-4.66

Таблица 3: Первые начальные моменты времени ожидания w_1 и пребывания v_1 для дисциплины «повтор без пересемплирования» при фиксированных $n = 3$, $\lambda_{\text{pos}} = 1{,}0$, $\delta = 0.05$, $\rho = 0{,}7$. Числ — численный расчёт, ИМ — имитационное моделирование, Отн.% — относительная погрешность.

тельные воздействия в этих постановках уменьшают «эффективную загрузку» системы за счёт преждевременного завершения обслуживания и/или мгновенного сброса заявок. Во-вторых, при *повторном обслуживании* (рестартах) значения w_1 и v_1 в среднем выше, особенно при малом CV, так как рестарт приводит к потере уже выполненной части обслуживания.

Согласование численного метода с имитационным моделированием в таблицах 1–2 в целом хорошее. В таблице 1 относительная погрешность для v_1, как правило, не превышает 1–2%, а для w_1 — нескольких процентов (в приведённых тестах максимум порядка 5%). Для среднеквадратичных отклонений (таблица 2) погрешности также остаются на уровне нескольких процентов (обычно $< 4\%$). Наибольшая чувствительность наблюдается для w_1 и σ_W при больших CV, что связано с влиянием хвостов распределений и приближениями, используемыми в расчёте (в частности, аппроксимацией интенсивностей рестартов/удалений и моментной аппроксимацией распределений).

Для дисциплины «повтор без пересемплирования» (таблица 3) расхождения с ИМ заметно выше, особенно для катастроф: относительные ошибки по w_1 и v_1 могут достигать порядка 5–10%. Это согласуется с тем, что при фиксированной длительности обслуживания и прерываниях средние и дисперсии становятся чувствительны к хвостам распределений и конечности производящей функции моментов $M_B(r)$ (см. раздел 4), а также к конечной длине имитационной выборки.

6.3. Графики параметрических зависимостей

На рисунке 2 приведены зависимости первого начального момента времени пребывания v_1 от числа каналов n для четырёх дисциплин обслуживания отрицательных заявок при фиксированной интенсивности отрицательных воздействий δ. В качестве примера использовано базовое значение $\delta = 0{,}3$. Отметим также, что для $CV = 0{,}8$ и $CV = 1{,}2$ зависимости $v_1(n)$ имеют близкий характер и отличаются сравнительно слабо; поэтому далее приводим графики только для базового значения $CV = 1{,}2$. Видно, что увеличение числа каналов приводит к росту v_1 при фиксированной ρ, что соответствует линейному росту $\mathbb{E}[B] = n\rho/\lambda_{\text{pos}}$. Влияние коэффициента вариации времени обслуживания показано в таблице 1.

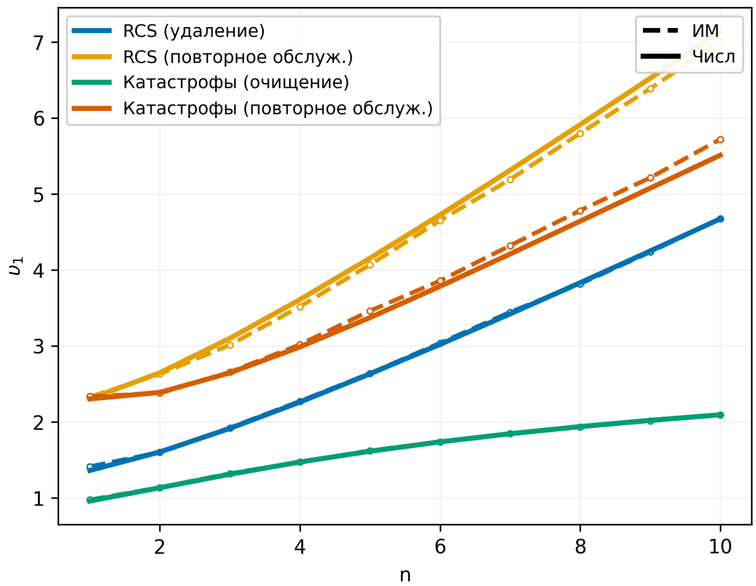
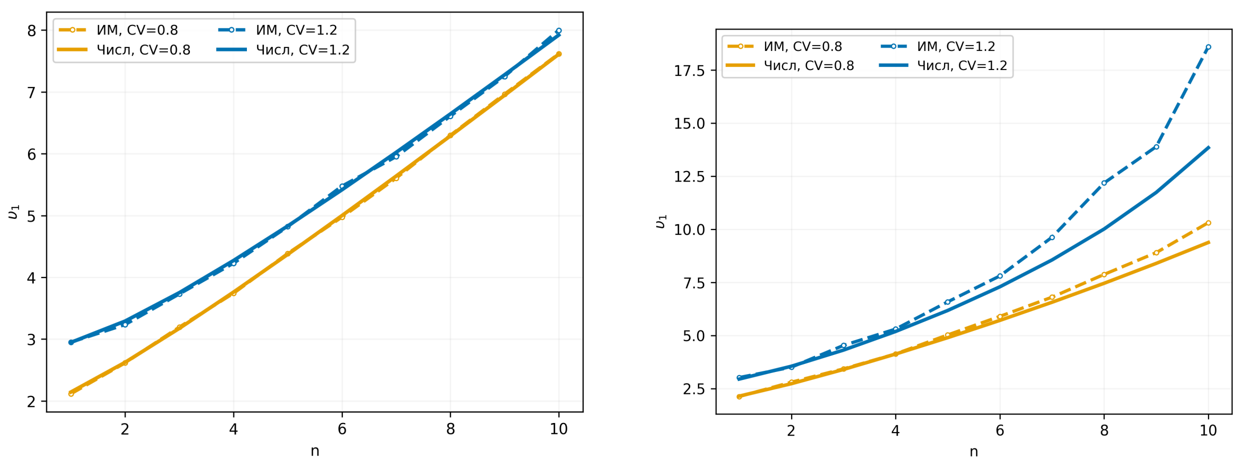


Рис. 2: Зависимость v_1 от числа каналов n при $\lambda_{\text{pos}} = 1{,}0$, $\delta = 0{,}3$, $\rho = 0{,}7$, $CV = 1{,}2$. Пунктир — ИМ, сплошная — численный расчёт.

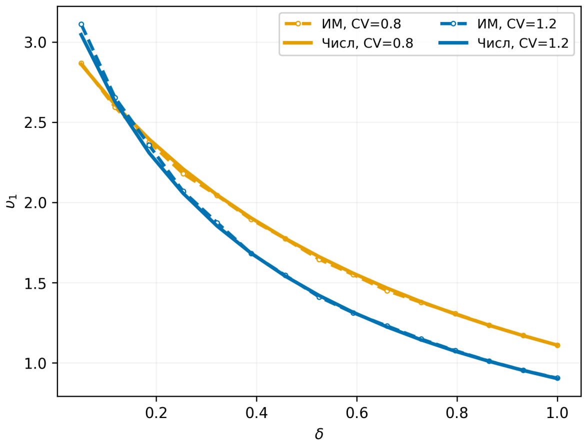


(a) RCS (повтор без пересемпл.)

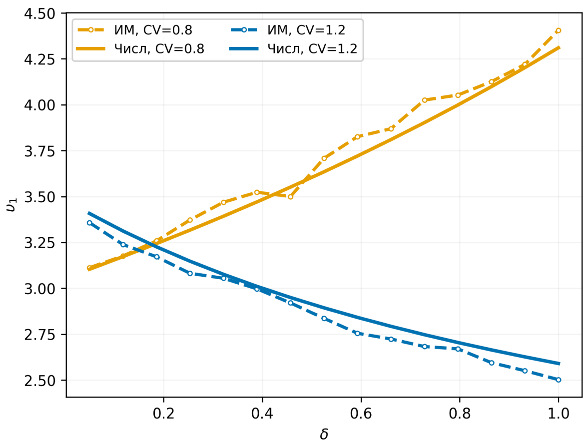
(b) Катастрофы (повтор без пересемпл.)

Рис. 3: Зависимость v_1 от числа каналов n для дисциплины «повтор без пересемплирования» при $\lambda_{\text{pos}} = 1{,}0$, $\delta = 0{,}02$, $\rho = 0{,}7$, $CV \in \{0{,}8;\ 1{,}2\}$.

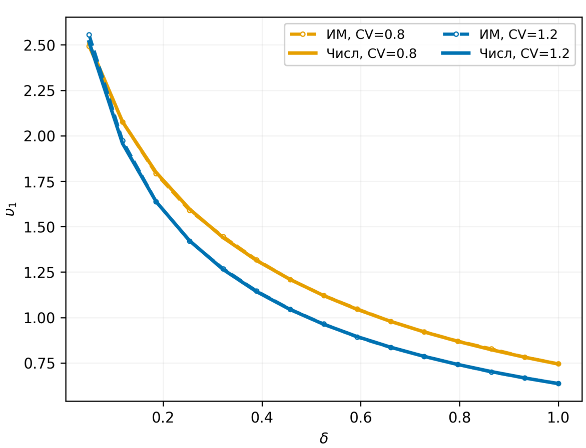
Отдельно отметим нетривиальный эффект для дисциплин с *повторным обслуживанием* (рестартом). Для постановки *вытесняющего повтора с перегенерацией* (preemptive-repeat with resampling) при прерывании заявка возвращается в очередь, а длительность обслуживания при следующем запуске генерируется заново. Для распределений с $CV > 1$ (существенная вариативность) более частые рестарты «отсекают» редкие очень длительные реализации обслуживания и повышают вероятность завершения по «быстрым» траекториям, в результате чего уменьшается *эффективное* время обслуживания B_{eff} и, как следствие, может снижаться υ_1 при увеличении δ. Для $CV < 1$ такой выигрыш отсутствует, и наблюдается ожидаемая зависимость: при увеличении частоты прерываний средние времена растут (рисунок 4).



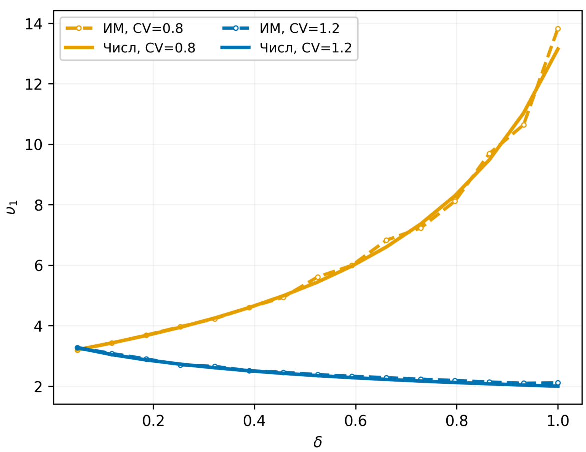
(a) RCS (удаление)



(b) RCS (повторное обслуж.)



(c) Катастрофы (очищение)



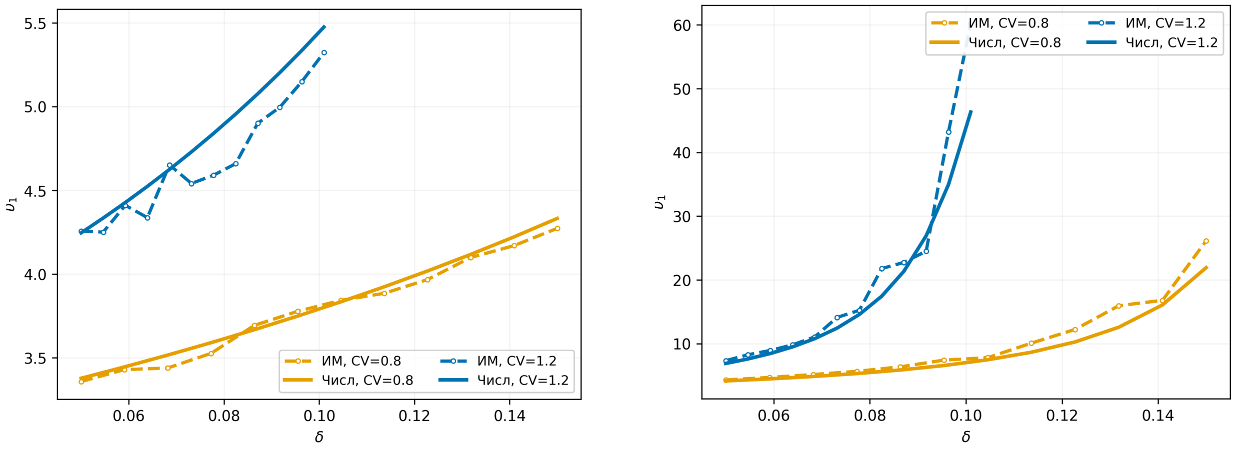
(d) Катастрофы (повторное обслуж.)

Рис. 4: Зависимость υ_1 от интенсивности отрицательных воздействий δ при $n = 3$, $\lambda_{\text{pos}} = 1{,}0$, $\rho = 0{,}7$, $CV \in \{0{,}8;\ 1{,}2\}$. Пунктир — ИМ, сплошная — численный расчёт.

Вероятность быть обслуженным q интерпретируется как доля положительных заявок, завершивших обслуживание. Для дисциплин с *удалением* (RCS) и *очищением* (катастрофы) q убывает с ростом δ, что отражает увеличение частоты «потерь» положительных заявок из системы. Для RCS (удаление) величина q в стационаре может быть также найдена из баланса потоков:

$$q \;=\; 1 - \frac{\delta}{\lambda}(1 - p_0),$$

где $p_0 = \mathbb{P}\{i = 0\}$ — вероятность пустой системы (отрицательное событие «теряется» при $i = 0$). Для дисциплин с *повторным обслуживанием* графики $q(\delta)$ не приводим: в

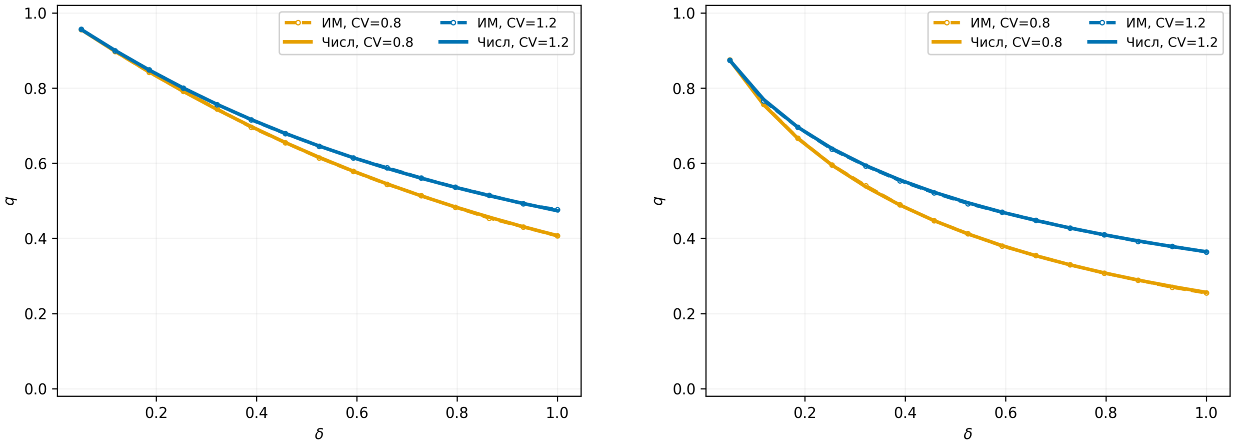


(a) RCS (повтор без пересемпл.)

(b) Катастрофы (повтор без пересемпл.)

Рис. 5: Зависимость v_1 от δ для дисциплины «повтор без пересемплирования» при $n = 3$, $\lambda_{\text{pos}} = 1{,}0$, $\rho = 0{,}7$, $CV \in \{0{,}8;\ 1{,}2\}$. Диапазон δ ограничен из-за быстрого роста моментов и условий их существования (см. раздел 4).

рассматриваемой постановке положительные заявки не уничтожаются, а лишь прерываются и возвращаются в очередь, поэтому $q \equiv 1$ и зависимость от δ тривиальна. На рисунке 6 показана зависимость $q(\delta)$ для дисциплин с удалением/очищением при $n = 3$, $\lambda_{\text{pos}} = 1{,}0$, $\rho = 0{,}7$, $CV \in \{0{,}8;\ 1{,}2\}$.



(a) RCS (удаление)

(b) Катастрофы (очищение)

Рис. 6: Зависимость вероятности быть обслуженным q от δ. Пунктир — ИМ, сплошная — численный расчёт.

7. Заключение

В данной работе предложено расширение метода Такахаси–Таками для расчёта многоканальных систем массового обслуживания с отрицательными заявками. Основные результаты состоят в следующем.

1. Разработаны правила модификации матриц переходов микросостояний для двух дисциплин обработки отрицательных заявок — RCS (удаление заявки из обслуживания) и катастроф (мгновенная очистка системы) — при двух вариантах поведения вытесненных заявок: уходе из системы и направлении на повторное обслуживание.

2. Предложены алгоритмы вычисления временных характеристик (первых начальных моментов времени ожидания и пребывания) через преобразование Лапласа–Стилтьеса и численное дифференцирование. Для дисциплины катастроф корректный расчёт времени пребывания основан на операции минимума с экспоненциальным временем до ближайшей катастрофы.

3. Проведено сравнение результатов численного метода с имитационным моделированием: относительная погрешность по первым моментам времени ожидания и пребывания не превышает нескольких процентов, аналогичный порядок точности наблюдается и для среднеквадратичных отклонений.

Вычислительное время предложенного метода существенно меньше по сравнению с прямым имитационным моделированием: для базового примера ($n = 3$, $\lambda_{\text{pos}} = 1{,}0$, $\delta = 0{,}3$, $\rho = 0{,}7$, $CV = 1{,}2$) численный расчёт занимает порядка 0,1 с, тогда как имитационное моделирование на 10^6 заявок — порядка 10–15 с, что делает метод пригодным для параметрического анализа и оптимизации.

В качестве направлений дальнейших исследований можно отметить обобщение предложенного подхода на системы с неоднородными каналами обслуживания, зависящими от состояния интенсивностями отрицательных воздействий, а также на многоклассовые модели с различными приоритетами. Программная реализация метода доступна в [18].

Список литературы

1. *Takahashi Y.*, *Takami Y.* A Numerical Method for the Steady-State Probabilities of a GI/G/c Queuing System in a General Class // Journal of the Operations Research Society of Japan. — 1976. — Т. 19, № 2. — С. 147—157.

2. *Рыжиков Ю. И.* Численные методы теории очередей. — Санкт-Петербург : Лань, 2018.

3. *Клейнрок Л.* Теория массового обслуживания. — Москва : Машиностроение, 1979.

4. *Бочаров П. П.*, *Печинкин А. В.* Теория массового обслуживания. — Москва : РУДН, 2004.

5. *Gelenbe E.* Product-form queueing networks with negative and positive customers // Journal of Applied Probability. — 1991. — Т. 28, № 3. — С. 656—663.

6. *Gelenbe E.*, *Schassberger R.* Stability of product form queueing networks with negative customers // Journal of Applied Probability. — 1992. — Т. 29, № 4. — С. 890—901.

7. *Artalejo J. R.* G-networks: A versatile approach for work removal in queueing networks // European Journal of Operational Research. — 2000. — Т. 126, № 2. — С. 233—249.

8. *Harrison P. G.*, *Pitel E.* The M/G/1 queue with negative customers // Advances in Applied Probability. — 1996. — Т. 28, № 2. — С. 540—566.

9. *Bayer N.*, *Boxma O. J.* Wiener–Hopf analysis of an M/G/1 queue with negative customers and of a related class of random walks // Queueing Systems. — 1996. — Т. 23, № 1—4. — С. 301—316.

10. *Boucherie R. J.*, *Boxma O. J.* The Workload in the M/G/1 Queue with Work Removal // Probability in the Engineering and Informational Sciences. — 1996. — Т. 10, № 2. — С. 261—277.

11. *Jain G.*, *Sigman K.* A Pollaczek–Khintchine formula for M/G/1 queues with disasters // Journal of Applied Probability. — 1996. — Т. 33, № 4. — С. 1191—1200.

12. *Dudin A.*, *Nishimura S.* A BMAP/SM/1 queueing system with Markovian arrival input of disasters // Journal of Applied Probability. — 1999. — Т. 36, № 3. — С. 868—881.

13. *Li Q.-L.*, *Lin C.* The M/G/1 processor-sharing queue with disasters // Computers & Mathematics with Applications. — 2006. — Т. 51, № 6/7. — С. 987—998.

14. *Atencia I.*, *Moreno P.* The discrete-time Geo/Geo/1 queue with negative customers and disasters // Computers & Operations Research. — 2004. — Т. 31, № 9. — С. 1537—1550.

15. *Artalejo J. R.*, *Gómez-Corral A.* On a single server queue with negative arrivals and request repeated // Journal of Applied Probability. — 1998. — Т. 35, № 3. — С. 907—918.

16. *Wang J.*, *Liu B.*, *Li J.* Transient analysis of an M/G/1 retrial queue subject to disasters and server failures // European Journal of Operational Research. — 2008. — Т. 189, № 3. — С. 1338—1353.

17. *Dharmaraja S.*, *Kumar R.* Transient solution of a Markovian queueing model with heterogeneous servers and catastrophes // Opsearch. — 2015. — Т. 52, № 4. — С. 810—824.

18. *Хабаров Р. С.*, *Лохвицкий В. А.* Most-Queue: Библиотека методов расчёта СМО с отрицательными заявками. — 2026. — Код расчётов для статьи: `works/negative_queues/`. Дата обращения: 14.02.2026. `https://github.com/xabarov/most-queue`.