

М. А. Гапшин, Д. Е. Огнева

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ДВУХКАНАЛЬНОЙ СМО С АДАПТИВНЫМИ И ПОСТОЯННЫМИ ПАРАМЕТРАМИ ОБСЛУЖИВАНИЯ В УСЛОВИЯХ АДАПТИВНОГО УПРАВЛЕНИЯ

Ключевые слова: Industry 4.0, система массового обслуживания, СМО без накопителя, адаптивное управление, управляемый параметр обслуживания, CPS-контроллер, имитационное моделирование, SimPy, пропускная способность, коэффициент ускорения β .

В работе исследуется двухканальная система массового обслуживания (СМО) без накопителя применительно к двум однотипным системам обслуживания. Первая система функционирует по классической схеме с постоянным временем обслуживания, тогда как вторая оснащена киберфизическим контроллером (CPS), адаптирующим управляемый параметр обслуживания и в зависимости от текущего числа заявок n в системе. Закон адаптации задаётся выражением $u(n) = \min\{u_0 \cdot (1 + \beta \cdot n); u_{\max}\}$, а время обслуживания представляется как сумма постоянной и переменной составляющих: $T(n) = T_{\text{const}} + T_{\text{var}} / (1 + \beta \cdot n)$. Базовые параметры заданы как модельный расчётный сценарий, достаточный для сопоставления двух политик управления без привязки к конкретному объекту. Сравнительный анализ выполнен методом дискретно-событийной имитации на языке Python с использованием библиотеки SimPy. Проведено 50 независимых прогонов на каждый из 30 сценариев при сетке $\lambda \in \{55, 60, 65, 70, 75, 80\}$ заявок/ч и $\beta \in \{0,05; 0,10; 0,15; 0,20; 0,25\}$, всего 1500 прогонов. Показано, что адаптивная система превосходит классическую во всех исследованных сценариях: прирост выходного потока составляет от 1,1 % при малой нагрузке до 5,8 % при высокой. Установлено, что наибольший эффект проявляется при высокой загрузке, когда вероятность блокировки входного потока возрастает. Выявлена оптимальная зона $\beta \in [0,15; 0,20]$ при $\lambda \geq 70$ заявок/ч, обеспечивающая прирост 3,5–4,8 % при сохранении заданных ограничений модели. Полученные результаты подтверждают применимость состояния-зависимого управления для повышения пропускной способности двухканальных систем без изменения их структуры.

М. А. Gapshin, D. E. Ogneva

COMPARATIVE ANALYSIS OF A TWO-CHANNEL QUEUEING SYSTEM WITH ADAPTIVE AND CONSTANT SERVICE PARAMETERS UNDER ADAPTIVE CONTROL CONDITIONS

Keywords: Industry 4.0, queueing system, queueing system without buffer, adaptive control, controllable service parameter, CPS controller, simulation modeling, SimPy, throughput, acceleration coefficient β .

This paper investigates a two-channel queueing system without a buffer applied to two equivalent service systems. System 1 operates according to a classical scheme with a constant service time, whereas System 2 is equipped with a cyber-physical controller (CPS) that adapts a controllable service parameter u based on the current number of jobs n in the system. The adaptation law is defined as $u(n) = \min\{u_0 \cdot (1 + \beta \cdot n); u_{\max}\}$, and the service time is represented as the sum of constant and variable components: $T(n) = T_{\text{const}} + T_{\text{var}} / (1 + \beta \cdot n)$. Baseline parameters are specified as a model calculation scenario sufficient for comparing two control policies without reference to a specific physical object. The comparative analysis was performed by discrete-event simulation in Python using the SimPy library. The experiment included 50 independent runs for each of 30 scenarios with $\lambda \in \{55, 60, 65, 70, 75, 80\}$ requests/h and $\beta \in \{0.05, 0.10, 0.15, 0.20, 0.25\}$, totalling 1,500 runs. The results show that the adaptive system outperforms the classical one in all investigated scenarios: throughput gain ranges from 1.1% under low load to 5.8% under high load. The greatest effect is observed at higher load levels, where the probability of input-flow blocking increases. The optimal zone $\beta \in [0.15; 0.20]$ at $\lambda \geq 70$ requests/h yields a 3.5–4.8% gain while satisfying the specified model constraints. These results confirm the applicability of state-dependent control for increasing the throughput of two-channel systems without changing their structure.

Введение

Четвёртая промышленная революция, получившая название Industry 4.0, связана с распространением киберфизических систем (CPS — Cyber-Physical Systems): встроенные датчики и контроллеры образуют замкнутый цифровой контур, в котором физический процесс и управляющее программное обеспечение взаимодействуют в режиме реального времени [1]. Практическую значимость такого подхода демонстрируют исследования цифровых двойников и киберфизических систем: объединение моделей, данных мониторинга и управляющих алгоритмов позволяет оценивать состояние системы и корректировать параметры её работы [2]. В таком контуре объект управления перестаёт быть

пассивным исполнителем заранее заданного расписания: он учитывает текущую загрузку и способен перераспределять ресурсы при изменении состояния системы.

По данным [3], адаптивное планирование относится к числу наиболее распространённых стратегий Industry 4.0 с уровнем внедрения 64%, а системы реального времени на основе обучения с подкреплением демонстрируют более высокие показатели эффективности по сравнению с эвристическими и экспертными подходами. Это свидетельствует о практической востребованности методов адаптивного управления сложными техническими системами.

Инструментом для анализа потоков заявок традиционно служит теория массового обслуживания (ТМО). Классические модели —

M/D/c, M/M/c, M/G/c — предполагают постоянство интенсивности входного потока λ и интенсивности обслуживания μ [4, 5]. Такое допущение справедливо для систем с жёстко фиксированным временем обслуживания, однако неприменимо к системам с обратной связью: здесь μ явно зависит от состояния системы n . Игнорирование этой зависимости занижает оценку пропускной способности адаптивной системы и делает невозможным количественное обоснование эффекта от внедрения CPS [5, 9].

Объектом исследования являются две двухканальные системы обслуживания без накопителя. Система 1 — классическая: время обслуживания постоянно. Система 2 — адаптивная: CPS-контроллер изменяет управляемый параметр обслуживания в зависимости от текущего числа заявок в системе.

Целью работы является количественная оценка прироста пропускной способности адаптивной системы над классической по результатам многопрогонной дискретно-событийной симуляции. Для достижения цели поставлены следующие задачи:

- формализовать обе системы как двухканальные СМО без накопителя с блокировкой при числе заявок $n = 2$;
- задать закон адаптации управляемого параметра обслуживания с учётом верхнего ограничения;
- реализовать дискретно-событийную имитационную модель (Python / SimPy);
- провести вычислительный эксперимент на сетке параметров (λ , β) из 30 сценариев, по 50 прогонов каждый;
- построить карту эффективности и выявить оптимальную зону коэффициента ускорения β .

Практическая значимость работы состоит в том, что полученные результаты позволяют обоснованно выбирать значение β при настройке CPS-контроллера без натурных экспериментов на конкретном объекте.

Научная новизна работы состоит в следующем. В отличие от классических моделей с постоянными параметрами [4, 5], впервые исследована двухканальная СМО без накопителя с состояние-зависимым управляемым параметром вида $u(n) = \min\{u_0 \cdot (1 + \beta \cdot n); u_{\max}\}$ в контексте CPS-управления. Новым результатом является количественная карта эффективности на сетке (λ , β), позволяющая выбирать коэффициент ускорения β без натурных экспериментов на конкретном объекте.

Описание моделируемой системы

Рассматриваются две однотипные системы обслуживания, каждая из которых содержит два параллельных канала. На вход поступает поток заявок, а каждый канал выполняет одну операцию обслуживания. На каждом из двух сравниваемых вариантов задано по два идентичных канала, работающих параллельно, — эта конфигурация и образует двухканальную СМО с числом каналов $c = 2$.

Конкретная физическая природа операции в модели не фиксируется: управляемый параметр u рассматривается как нормированная величина, влияющая на переменную часть времени обслуживания. Такой подход позволяет анализировать эффект адаптивного управления без привязки к конкретному типу оборудования, материалу или технологической операции.

Базовые (номинальные) параметры обслуживания заданы как расчётный сценарий дискретно-событийного эксперимента [6]. Переменная часть времени обслуживания зависит от управляемого параметра u , а постоянная часть отражает операции, не сокращаемые CPS-контроллером в рамках рассматриваемой модели.

Базовое значение управляемого параметра:

$$u_0 = 287 \text{ усл. ед./мин.} \quad (1)$$

Верхняя граница управляемого параметра:

$$u_{\max} = 4000 \text{ усл. ед./мин.} \quad (2)$$

Переменная часть времени обслуживания:

$$T_{\text{var}} = 0,52 \text{ мин.} \quad (3)$$

Постоянная часть времени обслуживания:

$$T_{\text{const}} = 0,80 \text{ мин.} \quad (4)$$

Базовое время обслуживания:

$$T_0 = T_{\text{var}} + T_{\text{const}} = 0,52 + 0,80 = 1,32 \text{ мин.} \quad (5)$$

Постоянная составляющая включает подготовительно-завершающие и контрольные действия, которые в данной модели не зависят от значения управляемого параметра.

$$T_0 = 1,32 \text{ мин./заявка.} \quad (6)$$

Базовая интенсивность обслуживания одного канала:

$$\mu_0 = 60 / T_0 = 60 / 1,32 \approx 45,4 \text{ заявок/ч.} \quad (7)$$

Суммарная пропускная способность двух каналов при базовых параметрах составляет $2 \cdot \mu_0 \approx 90,7$ заявок/ч. Диапазон входных потоков, исследуемых в настоящей работе ($\lambda = 55 \dots 80$ заявок/ч), соответствует коэффициентам загрузки $\rho = \lambda / (2\mu_0) = 0,61 \dots 0,88$. Таким образом, система работает в условиях высокой, но не предельной нагрузки — именно в этом диапазоне адаптивное управление наиболее эффективно, поскольку блокировка входного потока происходит регулярно, но не постоянно.

При адаптивном изменении параметров обслуживания необходимо соблюдать верхние ограничения, задающие область допустимых значений модели.

Максимальное значение управляемого параметра принимается равным:

$$u_{\max} = 4000 \text{ усл. ед./мин.} \quad (8)$$

Ограничение $u_{\max}$ принято в качестве верхней границы в законе адаптации. Такая постановка соответствует классу систем потерь с управляемой интенсивностью обслуживания, в которых повышение скорости обслуживания ограничивается допустимой областью управления [7].

Структурная схема исследуемой двухканальной СМО без накопителя представлена на рис. 1.

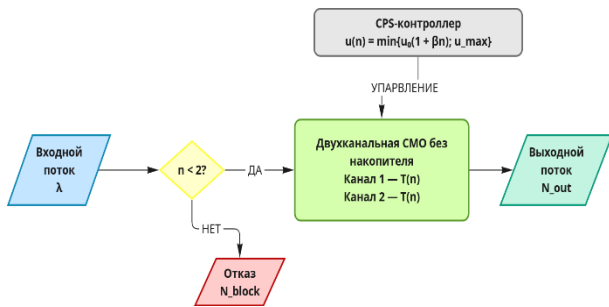


Рис. 1 - Структурная схема исследуемой двухканальной СМО без накопителя

Fig. 1 - Structural diagram of the investigated two-channel queueing system without a buffer

Имитационная модель

Обе сравниваемые системы описываются как двухканальная система массового обслуживания без накопителя (очереди). Данный тип системы характеризуется следующими структурными свойствами:

- число параллельных каналов обслуживания $c = 2$;
- накопитель (буфер) для ожидающих заявок отсутствует: ёмкость системы равна числу каналов;
- состояния системы $n \in \{0, 1, 2\}$, где n — число заявок, находящихся непосредственно на обслуживании;
- при $n = 2$ вновь поступившая заявка немедленно отклоняется: оба канала заняты и ждать негде.

Заявка, поступившая в занятую систему, выходит из неё без обслуживания. Принятое допущение стандартно в литературе по СМО с отказами [7, 8] и соответствует системе типа M/D/2/2 (для системы 1) и M/G/2/2 (для системы 2, где распределение времени обслуживания нестационарно).

Входной поток заявок на обе системы принимается пуассоновским с постоянной интенсивностью λ [заявок/ч].

Выбор линейного закона обусловлен тремя соображениями. Во-первых, при $n = 0$ закон вырождается в $u(0) = u_0$, обеспечивая непрерывность перехода от базового режима к адаптивному. Во-вторых, зависимость содержит единственный настраиваемый параметр β , что упрощает оптимизацию в инженерной практике. В-третьих, для пространства состояний $n \in \{0, 1, 2\}$ двухканальной СМО без накопителя более сложные законы — экспоненциальный $u(n) = u_0 \cdot \exp(\alpha \cdot n)$ или ступенчатый — не дают принципиального преимущества: при трёх возможных состояниях дополнительные степени свободы не используются.

CPS-контроллер системы 2 на каждом такте считывает текущее состояние системы n и вычисляет управляемый параметр обслуживания по следующему закону:

$$u(n) = \min\{u_0 \cdot (1 + \beta \cdot n); u_{\max}\}, \quad (9)$$

где $u_0 = 287$ усл. ед./мин — базовое значение управляемого параметра при $n = 0$; $\beta \geq 0$ — безразмерный коэффициент ускорения, настраиваемый параметр CPS-контроллера; $u_{\max} = 4000$ усл. ед./мин — верхнее ограничение (формула 8).

Физический смысл закона (9) состоит в следующем: при появлении хотя бы одной заявки в системе ($n \geq 1$) контроллер увеличивает значение управляемого параметра в пропорции $(1 + \beta \cdot n)$. При $n = 2$ — оба канала заняты, новые заявки блокируются — форсирование максимально: $u(2) = \min\{287 \cdot (1 + 2\beta); 4000\}$. Это сокращает переменную часть времени обслуживания и ускоряет освобождение каналов.

Время цикла при текущем состоянии n :

$$T(n) = T_{\text{const}} + T_{\text{var}} \cdot u_0 / u(n) = 0,80 + 0,52 / (1 + \beta \cdot n) \text{ мин.} \quad (10)$$

При $\beta = 0$ формула (10) вырождается в $T(n) \equiv T_0 = 1,32$ мин, то есть адаптивная система 2 становится идентичной классической системе 1. Это обеспечивает корректность сравнительного анализа: единственным отличием моделей служит ненулевое значение β .

В системе 1 время обслуживания постоянно:

$$T = T_0 = 1,32 \text{ мин при любом } n. \quad (11)$$

Дискретно-событийная имитационная модель реализована на языке программирования Python с использованием библиотеки SimPy версии 4.1.1 [9] — открытого симулятора, построенного на основе генераторов (coroutine-based). Ключевые архитектурные решения:

Генерация заявок. Функция-генератор создаёт поток заявок с постоянным интервалом $\Delta t = 60/\lambda$ мин. Перед обращением к ресурсу проверяется текущая загрузка: если оба слота ресурса заняты (`resource.count == 2`), заявка отклоняется без обработки. Счётчик отклонений ведётся отдельно для анализа потерь.

Ресурс обработки. Объект `simpy.Resource` с `capacity = 2` моделирует два параллельных канала. Очередь ожидания ресурса задана нулевой длиной (`queue_limit = 0`), что реализует логику блокировки без накопителя.

Время обслуживания. Для каждой заявки при захвате ресурса определяется текущее $n = \text{resource.count}$ (включая только что захваченный слот). Управляемый параметр вычисляется по формуле (9), затем разыгрывается случайный разброс:

$$T_{\text{реал}} = T(n) \cdot U[0,9; 1,1], \quad (12)$$

где $U[0,9; 1,1]$ — равномерно распределённая случайная величина, имитирующая случайные отклонения времени обслуживания. Разброс $\pm 10\%$ принят как модельное допущение вычислительного эксперимента.

Метрика эффективности. По истечении расчётного горизонта $\tau = 480$ мин фиксируется число фактически обслуженных заявок N_{out} . Параллельно фиксируются число заблокированных заявок N_{block} и суммарное время занятости каждого канала.

Многопрогонный эксперимент. На каждый сценарий (пара значений λ и β) запускается 50 прогонов с различными стартовыми значениями генератора псевдослучайных чисел (`seed = 0, 1, ..., 49`). Использование фиксированного набора `seed` обеспечивает воспроизводимость результатов. Схожий подход применён в [10], где DES-система оценивает несколько управляющих политик на

основе текущего состояния системы и выбирает наилучшую. Итоговая точечная оценка: среднее $\bar{N}_{out}$ по 50 прогонам. Для контроля статистической погрешности вычислялось также среднеквадратическое отклонение σ и 95%-й доверительный интервал по формуле Стьюдента; во всех сценариях ширина интервала не превысила ± 3 заявки за расчётный горизонт.

Пространство экспериментальных параметров задано следующими сетками:

— интенсивность входного потока $\lambda \in \{55, 60, 65, 70, 75, 80\}$ заявок/ч (шаг 5 заявок/ч), что соответствует коэффициентам загрузки $\rho \in \{0,61; 0,66; 0,72; 0,77; 0,83; 0,88\}$;

— коэффициент ускорения $\beta \in \{0,05; 0,10; 0,15; 0,20; 0,25\}$ (шаг 0,05).

Итого: $6 \times 5 = 30$ сценариев, $30 \times 50 = 1500$ прогонов. Полное время расчёта на стандартном вычислительном компьютере составило около 40 секунд, что подтверждает практичность метода для инженерной оптимизации в условиях цифровизации систем управления [12].

В каждом сценарии параллельно симулируются системы 1 и 2 с одинаковыми последовательностями seed, что исключает влияние «шума» генератора на разность выходных показателей. Такой приём (commonrandomnumbers, CRN) является стандартным для сравнительных имитационных экспериментов [13] и позволяет обнаружить статистически значимые различия при меньшем числе прогонов.

Результаты имитационного эксперимента

Анализ полного массива результатов (1500 прогонов) выявил два устойчивых закономерности, справедливых для всей исследованной области параметров.

Первая закономерность: прирост выходного потока адаптивной системы над классической монотонно возрастает с увеличением интенсивности входного потока λ . При малой нагрузке ($\rho \approx 0,61$) система большую часть времени находится в состоянии $n \leq 1$, то есть хотя бы один канал свободен. В этих условиях вероятность блокировки новой заявки невысока, и адаптивное ускорение срабатывает редко — разница в выходном потоке незначительна. С ростом λ доля времени в состоянии $n = 2$ увеличивается: оба канала чаще заняты, новые заявки систематически отклоняются. Именно в этом состоянии адаптация по формуле (9) наиболее ценна: сокращение $T(2)$ напрямую ускоряет освобождение канала и снижает вероятность блокировки следующей заявки.

Вторая закономерность: прирост монотонно возрастает с увеличением β , однако темп роста замедляется при $\beta > 0,20$. При достижении ограничения u_{max} дальнейшее увеличение β уже не меняет фактическое значение управляемого параметра — формула (9) насыщается: $u(n) = u_{max}$. В частности, при $\lambda = 80$ заявок/ч и $n = 2$ ограничение u_{max} достигается при $\beta \geq (u_{max}/u_0 - 1)/2 = (4000/287 - 1)/2 \approx 6,46$, что значительно превышает

исследованный диапазон. Тем не менее при $\lambda = 65\text{--}80$ заявок/ч и $n = 1$ ограничение достигается при $\beta \geq (4000/287 - 1)/1 \approx 12,9$, также вне диапазона. Следовательно, в исследованном диапазоне $\beta \leq 0,25$ насыщения формально нет, однако темп роста прироста замедляется по другой причине: при очень высоком β переменная часть времени обслуживания $T_{var}/(1 + \beta \cdot n)$ становится много меньше T_{const} , и дальнейшее её сокращение уже слабо влияет на суммарное $T(n)$, которое всё больше определяется постоянной составляющей $T_{const} = 0,80$ мин.

Зависимости прироста пропускной способности от интенсивности входного потока при различных значениях β представлены на рис. 2.

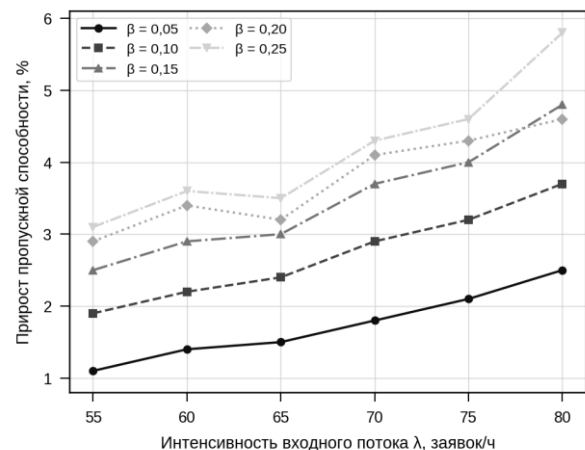


Рис. 2 - Зависимость прироста пропускной способности от интенсивности входного потока при различных β

Fig. 2 - Throughput gain versus input-flow intensity for different β values

Таблица 1 содержит сводные результаты по шести значениям λ при фиксированном $\beta = 0,20$ — центре выявленной оптимальной зоны. В столбцах приведено среднее число обслуженных заявок за расчётный горизонт по 50 прогонам для системы 1 (постоянный режим) и системы 2 (адаптивный режим), а также прирост в процентах.

Таблица 1 - Число обслуженных заявок за расчётный горизонт 480 мин при $\beta = 0,20$ (среднее по 50 прогонам)

Table 1 - Number of requests served over a 480-minute simulation period when $\beta = 0.20$ (average of 50 runs)

λ , заявок/ч	ρ	Система 1	Система 2	Прирост, %
55	0,61	329	338	+2,9
60	0,66	347	359	+3,4
65	0,72	366	378	+3,2
70	0,77	382	398	+4,1
75	0,83	397	414	+4,3
80	0,88	409	428	+4,6

Анализ таблицы 1 позволяет сделать следующие наблюдения. Во-первых, адаптивная система превосходит классическую во всех шести строках без

исключения, что подтверждает универсальность преимущества в исследованном диапазоне нагрузок. Во-вторых, абсолютный прирост выходного потока возрастает с λ : от 9 заявок за расчётный горизонт при $\lambda = 55$ до 19 заявок при $\lambda = 80$. В-третьих, при $\lambda = 65$ прирост (3,2 %) незначительно ниже, чем при $\lambda = 60$ (3,4 %): это объясняется статистическим разбросом при конечном числе прогонов и не противоречит общей монотонной тенденции, которая восстанавливается при $\lambda = 70$.

Практический смысл прироста на уровне 4–4,6 % состоит в том, что при неизменном числе каналов и неизменной структуре СМО адаптивная политика управления обеспечивает статистически устойчивое увеличение выходного потока. Например, при $\lambda = 75$ заявок/ч и $\beta = 0,20$ средний результат возрастает с 397 до 414 обслуженных заявок за расчётный горизонт 480 мин, то есть на 17 заявок. В относительном выражении это соответствует приросту 4,3 %, достигаемому только за счёт изменения закона управления параметром обслуживания.

Полная карта прироста выходного потока по всем 30 сценариям приведена в таблице 2. В каждой ячейке указан прирост адаптивной системы над классической в процентах от выходного потока классической системы. Строки соответствуют значениям λ , столбцы — значениям β .

Таблица 2 - Карта прироста выходного потока, % (система 2 над системой 1)

Table 2 - Output Flow Increase Chart, % (System 2 vs. System 1)

$\lambda \setminus \beta$	0,05	0,10	0,15	0,20	0,25
55 ($\rho=0,61$)	1,1	1,9	2,5	2,9	3,1
60 ($\rho=0,66$)	1,4	2,2	2,9	3,4	3,6
65 ($\rho=0,72$)	1,5	2,4	3,0	3,2	3,5
70 ($\rho=0,77$)	1,8	2,9	3,7	4,1	4,3
75 ($\rho=0,83$)	2,1	3,2	4,0	4,3	4,6
80 ($\rho=0,88$)	2,5	3,7	4,8	4,6	5,8

Карта наглядно иллюстрирует обе выявленные закономерности: значения прироста растут как при движении вниз по строкам (рост λ), так и при движении вправо по столбцам (рост β). Минимальный прирост (1,1 %) сосредоточен в левом верхнем углу (малая нагрузка, малое β), максимальный (5,8 %) — в правом нижнем (высокая нагрузка, $\beta = 0,25$).

Следует обратить внимание на ячейку $\lambda = 80$, $\beta = 0,20$, где прирост (4,6 %) ниже, чем в ячейке $\lambda = 80$, $\beta = 0,15$ (4,8 %) и при $\beta = 0,25$ (5,8 %). Это единственное отклонение от монотонности в карте; с высокой вероятностью оно объясняется статистическим шумом (ширина 95%-го ДИ составляет около $\pm 2,5$ % для данного сценария) и не свидетельствует о реальной немонотонности.

Графическое представление полной карты эффективности приведено на рис. 3.

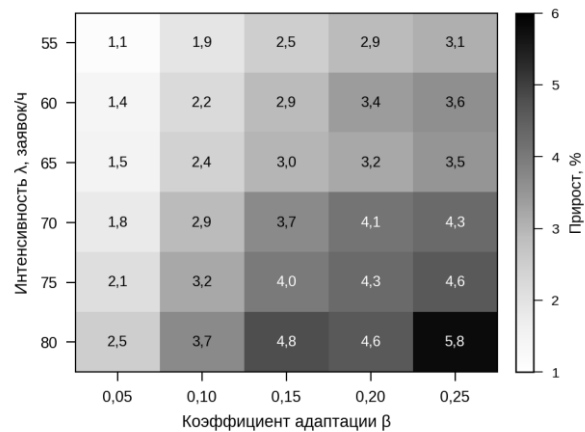


Рис. 3 - Карта прироста пропускной способности на сетке (λ , β)

Fig. 3 - Throughput-gain map over the (λ , β) grid

На основании анализа карты выявлена оптимальная зона управляющего параметра: $\beta \in [0,15; 0,20]$ при $\lambda \geq 70$ заявок/ч. В этой зоне достигается прирост 3,5–4,8 % при следующих дополнительных условиях:

— максимальное значение управляемого параметра при $n = 2$ составляет $u(2) = \min\{287 \cdot (1 + 0,20 \cdot 2); 4000\} = \min\{401,8; 4000\} = 401,8$ усл. ед./мин, что составляет лишь около 10 % от $u_{\max}$;

— при $\beta = 0,25$ и $n = 2$ значение $u(2) = 287 \cdot (1 + 0,25 \cdot 2) = 430,5$ усл. ед./мин — также находится далеко от верхнего ограничения, однако даёт меньший относительный выигрыш на единицу увеличения β ;

— дальнейшее увеличение β выше 0,25 требует отдельной проверки устойчивости модели и чувствительности результатов к выбранной верхней границе $u_{\max}$.

Таким образом, зона $\beta \in [0,15; 0,20]$ является оптимальной с точки зрения баланса «прирост пропускной способности — умеренность управляющего воздействия».

Обсуждение

Рассмотрим механизм преимущества адаптивной системы более детально. В двухканальной СМО без накопителя потеря производительности происходит исключительно в моменты блокировки: когда оба канала заняты ($n = 2$), любая прибывающая заявка отклоняется. Блокировка длится ровно столько, сколько остаётся до завершения ближайшего из двух одновременно выполняемых обслуживаний. Следовательно, сокращение $T(2)$ — времени обслуживания именно в состоянии $n = 2$ — напрямую уменьшает среднюю длительность блокировки и увеличивает долю заявок, принятых в обслуживание.

Помимо этого, ускорение при $n = 1$ также вносит положительный вклад: если единственный занятый канал освобождается быстрее, вероятность того, что следующая заявка застанет оба канала занятыми, снижается. Таким образом, адаптивный контроллер воздействует на обе «точки потерь» системы: уменьшает вероятность перехода в состояние $n = 2$ и сокращает время пребывания в нём.

В классической системе этот механизм отсутствует: μ постоянна вне зависимости от n , и система «не знает» о накоплении нагрузки. Именно невозможность обратной связи является принципиальным ограничением классической схемы управления с точки зрения Industry 4.0.

Схожее преимущество может проявляться в системах с переменными параметрами обслуживания — такой эффект был отмечен в исследованиях двухканальных СМО с каналами различной производительности [14]. Однако в отличие от стационарного случая с фиксированным соотношением μ_1/μ_2 , адаптивное управление позволяет динамически менять эффективную производительность μ в реальном времени в зависимости от состояния n , достигая больших преимуществ.

Для системы типа M/D/c/c ($c = 2$, постоянное время обслуживания, отсутствие накопителя) аналитическое решение стационарных вероятностей состояний получается по формулам Эрланга В для простейшего потока. При $\lambda = 70$ заявок/ч и $\mu_0 = 45,4$ заявок/ч коэффициент нагрузки на один канал $a = \lambda/\mu_0 = 70/45,4 \approx 1,542$. Вероятность блокировки по формуле Эрланга:

$$P_{\text{блок}} = a^2 / (2! + 2 \cdot a + a^2) = 2,378 / (2 + 3,084 + 2,378) \approx 0,319. \quad (13)$$

Ожидаемая пропускная способность системы: $\lambda_{\text{эфф}} = \lambda \cdot (1 - P_{\text{блок}}) = 70 \cdot 0,681 \approx 47,7$ заявок/ч. За расчётный горизонт 480 мин это соответствует $47,7 \cdot 8 \approx 382$ заявкам — в соответствии с результатом симуляции системы 1 при $\lambda = 70$ (382 заявки за расчётный горизонт по таблице 1). Это подтверждает корректность имитационной модели: для системы 1 с постоянным временем обслуживания результаты симуляции воспроизводят аналитическое решение.

Для системы 2 аналитического решения в замкнутом виде не существует, поскольку время обслуживания нестационарно (зависит от n) и нестепенное. Имитационный подход является в данном случае единственным практически применимым методом.

Предложенная модель имеет ряд допущений, которые следует учитывать при переносе результатов на прикладные системы.

Отсутствие учёта деградации ресурса. Модель предполагает постоянство связи между управляемым параметром u и временем обслуживания в течение всего расчётного горизонта. В реальных системах повышение интенсивности обслуживания может сопровождаться ростом затрат, снижением ресурса или ухудшением качества результата. Поэтому при автоматическом форсировании параметра обслуживания необходимо учитывать не только прирост выходного потока, но и ограничения по стоимости, ресурсу и качеству [15]. Это означает, что при длительной работе с $\beta = 0,20$ требуется дополнительная верификация модели на данных конкретной системы.

Мгновенная обратная связь. Предполагается, что CPS-контроллер мгновенно считывает n и мгновенно изменяет управляемый параметр. На практике имеется задержка обратной связи: считывание

сигналов датчиков, обработка в контроллере, отправка команды исполнительному уровню. Если эта задержка мала по сравнению со средним временем обслуживания, её влияние на результаты можно считать второстепенным; при больших задержках модель следует расширить.

Вместе с тем полученные результаты непосредственно применимы к классу двухканальных систем, в которых скорость обслуживания поддается программному регулированию: сборочные станции с ЧПУ, операторские рабочие места колл-центров с регулируемым темпом обработки заявок, вычислительные узлы с динамическим управлением тактовой частотой. В каждом из этих случаев выбор $\beta \in [0,15; 0,20]$ при $\rho \geq 0,77$ позволяет ожидать прироста пропускной способности порядка 3,5–4,8 % без изменения аппаратной конфигурации.

Равномерное распределение разброса времени цикла. Принятое распределение $U[0,9; 1,1]$ является упрощением. Реальный разброс времени обслуживания может иметь правостороннюю асимметрию из-за редких, но длительных задержек. Использование более реалистичных распределений, например логнормального или смешанного, способно несколько изменить абсолютные значения прироста, но вряд ли повлияет на качественные выводы.

Заключение

В настоящей работе проведён сравнительный анализ двух двухканальных систем обслуживания без накопителя. Каждая система формализована как СМО с блокировкой при $n = 2$. Система 1 работает с постоянным временем обслуживания $T_0 = 1,32$ мин/заявка; система 2 оснащена CPS-контроллером, адаптирующим управляемый параметр по закону $u(n) = \min\{u_0 \cdot (1 + \beta \cdot n); u_{\max}\}$.

Сравнительный анализ выполнен методом дискретно-событийной имитации (Python / SimPy) на сетке из 30 сценариев ($\lambda = 55 \dots 80$ заявок/ч, $\beta = 0,05 \dots 0,25$), по 50 прогонов длительностью 480 мин каждый, — итого 1500 прогонов. Получены следующие основные результаты.

Во-первых, адаптивная система превосходит классическую по числу обслуженных заявок за расчётный горизонт во всех тридцати исследованных сценариях без исключения. Прирост составляет от 1,1 % ($\lambda = 55$ заявок/ч, $\beta = 0,05$, $\rho = 0,61$) до 5,8 % ($\lambda = 80$ заявок/ч, $\beta = 0,25$, $\rho = 0,88$).

Во-вторых, прирост монотонно возрастает с увеличением нагрузки λ : при высокой загрузке ($\rho \rightarrow 1$) оба канала чаще находятся в состоянии $n = 2$, а именно в этом состоянии адаптация управляемого параметра даёт наибольший эффект за счёт сокращения длительности блокировки входного потока.

В-третьих, прирост монотонно возрастает с увеличением β , однако темп роста снижается при $\beta > 0,20$ из-за относительного увеличения доли постоянной составляющей T_{const} в суммарном цикле, которая не поддается сокращению программными средствами CPS.

В-четвёртых, выявлена оптимальная зона управляющего параметра $\beta \in [0,15; 0,20]$ при $\lambda \geq 70$ заявок/ч ($\rho \geq 0,77$): в этой зоне достигается прирост 3,5–4,8 % при $u(2) \leq 401,8$ усл. ед./мин, что составляет около 10 % от заданной верхней границы $u_{\max} = 4000$ усл. ед./мин.

Принципиальным достоинством предложенного решения является то, что прирост пропускной способности достигается за счёт изменения закона управления без увеличения числа каналов и без изменения структуры СМО. При $\lambda = 75$ заявок/ч и $\beta = 0,20$ средний выходной поток возрастает с 397 до 414 обслуженных заявок за расчётный горизонт 480 мин; относительный прирост составляет 4,3 %. Этот вывод подтверждает, что адаптивное управление может давать измеримый эффект даже при сохранении исходной структуры системы.

Направлениями дальнейших исследований являются: учёт ресурсных ограничений и включение плановых остановок как дополнительных событий симуляции; расширение на разнородные каналы ($\mu_1 \neq \mu_2$); построение модели с конечным буфером (накопителем ёмкостью k) для анализа промежуточных конфигураций; верификация результатов по обезличенным данным наблюдений.

Литература

1. H. Kagermann, W. Wahlster, J. Helbig, Recommendations for implementing the strategic initiative INDUSTRIE 4.0, acatech, Frankfurt am Main, 2013.
2. F. Tao, Q. Qi, L. Wang, A.Y.C. Nee, Engineering, 5, 4, 653–661 (2019). DOI: <https://doi.org/10.1016/j.eng.2019.01.014>.
3. D.A. Barua, S.A. Sami, L. Barua, Scientific Reports, 15, Article 41559 (2025). DOI: <https://doi.org/10.1038/s41598-025-25413-6>.
4. М.А. Плескунов, Теория массового обслуживания, Издательство Уральского университета, Екатеринбург, 2022.
5. Т.В. Черушева, Н.В. Зверовщикова, Теория массового обслуживания, Издательство ПГУ, Пенза, 2021.
6. A.M. Law, Simulation Modeling and Analysis, 5th ed., McGraw-Hill Education, New York, 2015.
7. M. Lefebvre, Axioms, 14, 2, Article 130 (2025). DOI: <https://doi.org/10.3390/axioms14020130>.
8. И.Н. Валеев, А.П. Кирпичников, Вестник Казанского технологического университета, 4, 75–79 (2006).
9. N. Matloff, Introduction to Discrete-Event Simulation and the SimPy Language, University of California, Davis, 2008.
10. Y. Zhang, Y. Kurkure, Y. Tao, M.E. Papka, Z. Lan, A Real-Time Digital Twin for Adaptive Scheduling, arXiv:2512.18894 (2025).
11. W.J. Hopp, M.L. Spearman, Factory Physics, McGraw-Hill, New York, 2008.
12. J. Xu, E. Huang, L. Hsieh, L.H. Lee, Q.S. Jia, C.H. Chen, Journal of Simulation, 10, 4, 310–320 (2016). DOI: <https://doi.org/10.1057/s41273-016-0037-6>.
13. D. Gross, J.F. Shortle, J.M. Thompson, C.M. Harris, Fundamentals of Queueing Theory, Wiley, Hoboken, 2008.
14. И.З. Самерханов, Е.А. Печеный, Н.К. Нуриев, Современные наукоемкие технологии, 3, 87–92 (2024). DOI: <https://doi.org/10.17513/snt.39951>.
15. G. Chen, Z. Liu, L. Xia, Journal of the Operational Research Society, 74, 3, 979–991 (2023). DOI: <https://doi.org/10.1080/01605682.2022.2053308>.

References

1. H. Kagermann, W. Wahlster, J. Helbig, Recommendations for implementing the strategic initiative INDUSTRIE 4.0, acatech, Frankfurt am Main, 2013.
2. F. Tao, Q. Qi, L. Wang, A.Y.C. Nee, Engineering, 5, 4, 653–661 (2019). DOI: <https://doi.org/10.1016/j.eng.2019.01.014>.
3. D.A. Barua, S.A. Sami, L. Barua, Scientific Reports, 15, Article 41559 (2025). DOI: <https://doi.org/10.1038/s41598-025-25413-6>.
4. M.A. Pleskunov, Queueing Theory, Ural University Press, Ekaterinburg, 2022.
5. T.V. Cherusheva, N.V. Zverovshchikova, Queueing Theory, PSU Publishing House, Penza, 2021.
6. A.M. Law, Simulation Modeling and Analysis, 5th ed., McGraw-Hill Education, New York, 2015.
7. M. Lefebvre, Axioms, 14, 2, Article 130 (2025). DOI: <https://doi.org/10.3390/axioms14020130>.
8. I.N. Valeev, A.P. Kirpichnikov, Herald of Kazan Technol. Univ., 4, 75–79 (2006).
9. N. Matloff, Introduction to Discrete-Event Simulation and the SimPy Language, University of California, Davis, 2008.
10. Y. Zhang, Y. Kurkure, Y. Tao, M.E. Papka, Z. Lan, A Real-Time Digital Twin for Adaptive Scheduling, arXiv:2512.18894 (2025).
11. W.J. Hopp, M.L. Spearman, Factory Physics, McGraw-Hill, New York, 2008.
12. J. Xu, E. Huang, L. Hsieh, L.H. Lee, Q.S. Jia, C.H. Chen, Journal of Simulation, 10, 4, 310–320 (2016). DOI: <https://doi.org/10.1057/s41273-016-0037-6>.
13. D. Gross, J.F. Shortle, J.M. Thompson, C.M. Harris, Fundamentals of Queueing Theory, Wiley, Hoboken, 2008.
14. I.Z. Samerkhanov, E.A. Pecheny, N.K. Nuriev, Modern High Technologies, 3, 87–92 (2024). DOI: <https://doi.org/10.17513/snt.39951>.
15. G. Chen, Z. Liu, L. Xia, Journal of the Operational Research Society, 74, 3, 979–991 (2023). DOI: <https://doi.org/10.1080/01605682.2022.2053308>.

© М. А. Гапшин – студент кафедры «Информатики и прикладной математики» (ИПМ), Казанский национальный исследовательский технологический университет (КНИТУ), Казань, Россия, mgapshin@yandex.ru; Д. Е. Огнева – студент кафедры ИПМ КНИТУ, dariya.ognewa@yandex.ru.

© М. А. Gapshin – Student, Department of Informatics and Applied Mathematics (DIAM), Kazan National Research Technological University (KNRTU), Kazan, Russia, mgapshin@yandex.ru; D. E. Ogneva – Student, DIAM KNRTU, dariya.ognewa@yandex.ru.

Дата поступления рукописи в редакцию – 01.06.26.

Дата принятия рукописи в печать – 20.06.26.