

МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра системного анализа и автоматического управления

**ОПТИМИЗАЦИЯ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С  
ГРУППОВЫМ ВХОДЯЩИМ ПОТОКОМ  
АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студента 4 курса 481 группы  
направления 27.03.03 Системный анализ и управление  
факультета компьютерных наук и информационных технологий  
Шапулина Данилы Дмитриевича

Научный руководитель  
старший преподаватель

\_\_\_\_\_

Н. В. Сергеева

Заведующий кафедрой  
к. ф.-м. н., доцент

\_\_\_\_\_

И. Е. Тананко

## ВВЕДЕНИЕ

**Актуальность темы.** Системы массового обслуживания используются для моделирования процессов, в которых заявки поступают в систему, ожидают обслуживания и покидают ее после завершения обслуживания. Такие модели применяются при анализе компьютерных и телекоммуникационных сетей, транспортных систем, складов, call-центров и других объектов, где возникают очереди.

Особый интерес представляют системы с групповым входящим потоком. В таких системах заявки поступают не по одной, а группами случайного размера. Такая ситуация характерна для передачи данных пакетами, поступления партий товаров на склад или одновременного прихода группы клиентов в систему обслуживания. В этом случае для анализа системы недостаточно учитывать только среднюю интенсивность входного потока. Необходимо также учитывать распределение размеров групп, поскольку оно влияет на длину очереди, среднее время ожидания и требуемую интенсивность обслуживания.

В работе рассматривается система массового обслуживания  $M^{[X]}/M/1/\infty$ . В данной системе группы заявок поступают по пуассоновскому потоку, время обслуживания одной заявки имеет экспоненциальное распределение, обслуживание выполняется одним прибором, а очередь считается неограниченной. Такая модель позволяет исследовать влияние группового поступления заявок на характеристики системы и рассмотреть задачу выбора минимальной интенсивности обслуживания при заданном ограничении на среднее время ожидания в очереди.

**Цель бакалаврской работы** – исследование системы массового обслуживания  $M^{[X]}/M/1/\infty$  с групповым входящим потоком и оптимизация интенсивности обслуживания при ограничении на среднее время ожидания в очереди.

Для достижения цели были поставлены следующие задачи:

1. изучение теоретических основ систем массового обслуживания с групповым потоком, их классификации и особенностей;
2. построение математической модели системы  $M^{[X]}/M/1/\infty$  и составление системы уравнений для стационарных вероятностей состояний;
3. применение метода производящих функций для получения аналитических выражений для основных характеристик системы;

4. постановка задачи оптимизации интенсивности обслуживания при ограничении на среднее время ожидания в очереди;
5. разработка программы на языке Python для расчета характеристик системы и поиска минимальной интенсивности обслуживания;
6. проведение численных экспериментов для оценки влияния параметров группового потока на характеристики системы.

**Методологические основы** работы составляют методы теории массового обслуживания, теория марковских процессов, метод производящих функций, формулы Литтла, а также методы численного расчета характеристик систем массового обслуживания. В качестве теоретической базы использовались работы О. А. Новикова, Л. Клейнрока, Б. В. Гнеденко, А. Кофмана, К. Е. Самуйлова, Дж. Медхи, П. П. Бочарова и А. В. Печинкина.

**Практическая значимость бакалаврской работы** состоит в том, что разработанный подход и программа позволяют рассчитывать характеристики системы с групповым входящим потоком и подбирать минимальную интенсивность обслуживания при заданном ограничении на среднее время ожидания. Такая постановка удобна в ситуациях, когда сложно заранее задать стоимостные коэффициенты, но можно установить допустимый уровень качества обслуживания. Полученные результаты могут использоваться для предварительной оценки параметров серверных, сетевых, складских и сервисных систем, в которых заявки поступают группами.

**Структура и объем работы.** Бакалаврская работа состоит из введения, пяти разделов, заключения, списка использованных источников и приложения. Общий объем работы – 45 страниц, из них 37 страниц – основное содержание, включая 10 рисунков и 2 таблицы. Список использованных источников содержит 21 наименование. В приложении представлен код программы.

## **КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Первый раздел «Теоретические основы систем массового обслуживания с групповым потоком»** посвящен рассмотрению основных понятий теории массового обслуживания и особенностей группового входящего потока.

В подразделе 1.1 приведены основные элементы системы массового обслуживания: входящий поток заявок, накопитель, прибор обслуживания и выходящий поток. Рассмотрена классификация систем массового обслуживания

по обозначениям Кендалла и введены основные показатели эффективности: вероятность состояния  $p_k$ , среднее число заявок в системе  $L_{\text{сист}}$ , средняя длина очереди  $L_{\text{оч}}$ , среднее время пребывания  $W_{\text{сист}}$ , среднее время ожидания  $W_{\text{оч}}$  и коэффициент загрузки  $\rho$ .

В подразделе 1.2 описан групповой входящий поток заявок. Размер группы обозначается случайной величиной  $X$ , а вероятность поступления группы размера  $k$  задается величиной  $l_k$ . Средний размер группы определяется выражением

$$l = \sum_{k=1}^{\infty} k l_k.$$

Также введен второй начальный момент размера группы

$$l^{(2)} = \sum_{k=1}^{\infty} k^2 l_k$$

и дисперсия размера группы

$$D[X] = l^{(2)} - l^2.$$

Показано, что при групповом поступлении заявок нагрузка на систему становится более неравномерной, поэтому очередь и время ожидания могут быть больше, чем в системе с ординарным потоком.

В подразделе 1.3 рассмотрены основные варианты систем массового обслуживания с групповым входящим потоком. Помимо системы  $M^{[X]}/M/1/\infty$ , описаны системы  $M^{[X]}/G/1$ ,  $M^{[X]}/M/n$ ,  $M^{[X]}/M/1/m$ ,  $M^{[X]}/G/n$ ,  $M^{[X]}/M/n/m$  и  $M^{[X]}/D/n$ . Эти модели отличаются законом распределения времени обслуживания, числом приборов и ограничениями на длину очереди.

В подразделе 1.4 рассмотрены задачи оптимизации в системах массового обслуживания. Описаны две основные постановки: оптимизация с ограничением на показатель качества обслуживания и оптимизация с экономическим критерием. В данной работе используется постановка с ограничением на среднее время ожидания в очереди:

$$W_{\text{оч}}(\mu) \leq W_{\text{оч}}^{\text{доп}}.$$

Такой подход позволяет определить минимальную интенсивность обслуживания без задания стоимостных коэффициентов.

**Второй раздел «Математическая модель системы  $M^{[X]}/M/1/\infty$ »** посвящен формальному описанию исследуемой системы и применению метода производящих функций для записи производящей функции стационарных вероятностей.

В подразделе 2.1 сформулированы основные параметры модели. Группы заявок поступают в систему по пуассоновскому потоку с интенсивностью  $\lambda$ . Размер группы является случайной величиной с распределением  $l_k$ . Обслуживание выполняется одним прибором с экспоненциальным временем обслуживания параметром  $\mu$ . Очередь считается неограниченной, а группы обслуживаются в порядке поступления.

В подразделе 2.2 составлена система уравнений Колмогорова для стационарных вероятностей состояний. Пусть  $p_n$  – стационарная вероятность того, что в системе находится  $n$  заявок. Тогда система уравнений имеет вид

$$\begin{cases} \lambda p_0 = \mu p_1, \\ (\lambda + \mu) p_n = \mu p_{n+1} + \lambda \sum_{k=1}^n l_k p_{n-k}, \quad n = 1, 2, 3, \dots \end{cases}$$

Отмечено, что процесс не является обычным процессом размножения и гибели, поскольку при поступлении группы возможен переход сразу на несколько состояний вправо.

В подразделе 2.3 для решения системы уравнений используется метод производящих функций. Введена производящая функция вероятностей состояний

$$P(z) = \sum_{n=0}^{\infty} p_n z^n$$

и производящая функция распределения размера группы

$$L(z) = \sum_{k=1}^{\infty} l_k z^k.$$

После преобразований производящая функция записывается в виде

$$P(z) = \frac{\mu p_0(1-z)}{\mu(1-z) - \lambda z(1-L(z))}.$$

Из условия нормировки  $P(1) = 1$  найдено

$$p_0 = 1 - \frac{\lambda l}{\mu} = 1 - \rho,$$

где  $\rho = \lambda l / \mu$  – коэффициент загрузки системы. Условие стационарности имеет вид  $\rho < 1$ . Окончательно производящая функция записывается как

$$P(z) = \frac{\mu(1-\rho)(1-z)}{\mu(1-z) - \lambda z(1-L(z))}.$$

**Третий раздел «Анализ основных характеристик системы»** посвящен выводу характеристик системы  $M^{[X]}/M/1/\infty$  и постановке задачи оптимизации.

В подразделе 3.1 рассмотрен расчет стационарных вероятностей состояний. Вероятности  $p_n$  могут быть найдены разложением производящей функции  $P(z)$  в ряд Тейлора. Также приведена рекуррентная формула, следующая из системы уравнений баланса:

$$p_{n+1} = \frac{\lambda + \mu}{\mu} p_n - \frac{\lambda}{\mu} \sum_{k=1}^n l_k p_{n-k}, \quad n = 1, 2, 3, \dots$$

Эта формула используется для последовательного вычисления вероятностей состояний.

В подразделе 3.2 получены основные показатели эффективности системы. Коэффициент загрузки равен

$$\rho = \frac{\lambda l}{\mu}.$$

Вероятность простоя прибора определяется формулой

$$p_0 = 1 - \rho.$$

Среднее число заявок в системе находится как производная производящей

функции:

$$L_{\text{сист}} = P'(1) = \frac{\lambda(l + l^{(2)})}{2(\mu - \lambda l)}.$$

С учетом соотношения  $l^{(2)} = D[X] + l^2$  эта формула может быть записана в виде

$$L_{\text{сист}} = \frac{\lambda(D[X] + l + l^2)}{2(\mu - \lambda l)}.$$

Из этой формулы следует, что при фиксированных  $\lambda$  и  $l$  увеличение дисперсии размера группы приводит к росту среднего числа заявок в системе.

Средняя длина очереди определяется выражением

$$L_{\text{оч}} = L_{\text{сист}} - \rho.$$

Среднее время пребывания и среднее время ожидания находятся по формулам Литтла:

$$W_{\text{сист}} = \frac{L_{\text{сист}}}{\lambda l}, \quad W_{\text{оч}} = \frac{L_{\text{оч}}}{\lambda l}.$$

В подразделе 3.3 рассмотрены распределения времени ожидания и времени пребывания заявки в системе. Для произвольной заявки среднее время ожидания в очереди записывается в виде

$$W_{\text{оч}} = \frac{l^{(2)} + l}{2\mu l(1 - \rho)} - \frac{1}{\mu}.$$

Так как  $1 - \rho = 1 - \lambda l/\mu$ , эту формулу можно переписать как

$$W_{\text{оч}}(\mu) = \frac{l^{(2)} + l}{2l(\mu - \lambda l)} - \frac{1}{\mu}.$$

В подразделе 3.4 сформулирована задача оптимизации интенсивности обслуживания. Требуется найти минимальное значение  $\mu$ , при котором выполняется ограничение на среднее время ожидания:

$$\mu_{\min} = \min \mu, \quad W_{\text{оч}}(\mu) \leq W_{\text{оч}}^{\text{доп}}, \quad \mu > \lambda l.$$

Условие  $\mu > \lambda l$  соответствует условию стационарности системы. Показано, что при увеличении  $\mu$  среднее время ожидания уменьшается, а значение  $\mu_{\min}$  задает

границу между допустимым и недопустимым режимами работы системы.

**Четвертый раздел «Программная реализация»** посвящен описанию разработанной программы для расчета характеристик системы и поиска минимальной интенсивности обслуживания.

В подразделе 4.1 описана структура программы. Программа разработана на языке Python с использованием библиотек NumPy и Matplotlib. Основная функция программы выполняет расчет характеристик системы  $M^{[X]}/M/1/\infty$  для заданных параметров. Она вычисляет средний размер группы, второй момент, дисперсию, коэффициент загрузки, вероятность простоя, среднее число заявок в системе и очереди, среднее время пребывания и ожидания. Функция также проверяет условие стационарности.

Отдельная функция выполняет поиск минимальной интенсивности обслуживания  $\mu_{\min}$ . Поиск начинается от нижней границы стационарности  $\mu = \lambda l$  и выполняется последовательным перебором с шагом 0.01 до тех пор, пока не будет выполнено условие

$$W_{\text{оч}} \leq W_{\text{оч}}^{\text{доп}}.$$

В подразделе 4.2 приведены основные идентификаторы программы: интенсивность потока групп, интенсивность обслуживания, массив вероятностей размера группы, средний размер группы, второй момент, дисперсия, коэффициент загрузки, массив стационарных вероятностей, среднее число заявок в системе, средняя длина очереди, среднее время пребывания, среднее время ожидания и минимальная интенсивность обслуживания.

В подразделе 4.3 приведен пример работы программы для системы с высокодисперсным распределением размера группы  $l_k = [0,75; 0; 0; 0; 0,25]$  при  $\lambda = 0,6$  и  $\mu = 1,5$ . Для этого случая программа получила  $l = 2,0$ ,  $D = 3,0$ ,  $\rho = 0,8$ ,  $L_{\text{сист}} \approx 9,0$ ,  $W_{\text{оч}} \approx 6,83$ . При допустимом значении  $W_{\text{оч}}^{\text{доп}} = 5,0$  минимальная интенсивность обслуживания составила  $\mu_{\min} = 1,60$ .

**Пятый раздел «Численное исследование системы и анализ результатов»** посвящен сравнению различных распределений размера группы и анализу результатов оптимизации.

В подразделе 5.1 выбраны пять базовых случаев численного исследования:

- ординарный поток  $M/M/1$ ;
- групповой поток с распределением размера группы  $[0,5; 0; 0,5]$ ;
- поток с фиксированным размером группы 4;

- поток с низкой дисперсией  $[0,1; 0,8; 0,1]$ ;
- поток с высокой дисперсией  $[0,75; 0; 0; 0; 0,25]$ .

Во всех случаях исходный коэффициент загрузки выбран равным  $\rho = 0,8$ . Это позволяет сравнивать варианты при одинаковой относительной загрузке.

В подразделе 5.2 выполнено сравнение характеристик системы для различных распределений размера группы. При переходе от ординарного потока к групповому потоку с распределением  $[0,5; 0; 0,5]$  среднее число заявок в системе увеличивается с 4,0 до 7,0, средняя длина очереди – с 3,2 до 6,2, среднее время пребывания – с 5,0 до 8,75, а среднее время ожидания – с 4,0 до 7,75. Для потока с фиксированным размером группы 4 показатели ухудшаются еще сильнее:  $L_{\text{сист}} = 10,0$ ,  $L_{\text{оч}} = 9,2$ ,  $W_{\text{сист}} = 12,5$ ,  $W_{\text{оч}} = 11,5$ .

Также проведено сравнение двух случаев с одинаковым средним размером группы  $l = 2$ , но разной дисперсией. При увеличении дисперсии с  $D = 0,2$  до  $D = 3,0$  среднее число заявок в системе возрастает с 6,2 до 9,0, средняя длина очереди – с 5,4 до 8,2, среднее время пребывания – с 5,17 до 7,5, а среднее время ожидания – с 4,5 до 6,83. Это подтверждает, что дисперсия размера группы является самостоятельным фактором, влияющим на характеристики системы.

В подразделе 5.3 рассмотрено решение задачи оптимизации для случая высокой дисперсии. При исходном значении  $\mu = 1,5$  среднее время ожидания превышает допустимое значение  $W_{\text{оч}}^{\text{доп}} = 5,0$ . В результате подбора получено значение

$$\mu_{\min} \approx 1,60.$$

При  $\mu < \mu_{\min}$  ограничение на среднее время ожидания не выполняется, а при  $\mu \geq \mu_{\min}$  система работает в допустимом режиме.

В подразделе 5.4 выполнен сравнительный анализ результатов оптимизации. Построены зависимости  $W_{\text{оч}}(\mu)$  для различных распределений размера группы, зависимость минимальной интенсивности обслуживания от допустимого среднего времени ожидания, а также зависимость  $\mu_{\min}$  от дисперсии размера группы. Полученные результаты показывают, что при одном и том же ограничении на среднее время ожидания разные распределения размера группы требуют различных значений интенсивности обслуживания.

Для базовых случаев при ограничении  $W_{\text{оч}}^{\text{доп}} = 5,0$  были получены следующие значения минимальной интенсивности обслуживания: 0,97 для ординарного потока, 1,10 для группового потока с распределением  $[0,5; 0; 0,5]$ , 1,24

для фиксированного размера группы 4, 1,48 для низкодисперсного группового потока и 1,60 для высокодисперсного группового потока.

В подразделе 5.5 сформулированы выводы по экспериментам. Показано, что групповой характер входного потока ухудшает характеристики системы по сравнению с ординарным потоком. Также подтверждено, что увеличение дисперсии размера группы приводит к росту среднего времени ожидания и требует увеличения интенсивности обслуживания для выполнения заданного ограничения.

## ЗАКЛЮЧЕНИЕ

В ходе выполнения бакалаврской работы была исследована система массового обслуживания с групповым входящим потоком  $M^{[X]}/M/1/\infty$  и рассмотрена задача оптимизации интенсивности обслуживания при ограничении на среднее время ожидания в очереди.

В теоретической части работы были рассмотрены основные понятия систем массового обслуживания, особенности группового входящего потока и классификация систем с групповым поступлением заявок. Для исследуемой системы была построена математическая модель, записана система уравнений Колмогорова для стационарных вероятностей состояний и применен метод производящих функций. На основе полученной производящей функции были записаны выражения для основных характеристик системы: коэффициента загрузки, вероятности простоя, среднего числа заявок в системе и очереди, среднего времени пребывания и ожидания.

В работе была сформулирована задача поиска минимальной интенсивности обслуживания  $\mu_{\min}$ , при которой среднее время ожидания в очереди не превышает заданного допустимого значения. Такая постановка позволяет выбирать минимально необходимую производительность обслуживающего прибора без введения стоимостных коэффициентов.

Для проведения расчетов была разработана программа на языке Python. Программа рассчитывает характеристики системы  $M^{[X]}/M/1/\infty$ , проверяет условие стационарности и выполняет численный поиск минимальной интенсивности обслуживания методом последовательного перебора.

Проведенные численные эксперименты показали, что групповой характер входного потока существенно влияет на работу системы. При одинаковом

коэффициенте загрузки переход от ординарного потока к групповому приводит к увеличению среднего числа заявок в системе, средней длины очереди, среднего времени пребывания и среднего времени ожидания. Это означает, что использование модели с ординарным потоком для анализа систем с групповым поступлением может приводить к заниженной оценке требуемого ресурса обслуживания.

Также было показано, что при одинаковом среднем размере группы важную роль играет дисперсия размера группы. Чем выше дисперсия, тем больше среднее время ожидания и тем большая интенсивность обслуживания требуется для выполнения одного и того же ограничения. Следовательно, при проектировании систем с групповым входящим потоком необходимо учитывать не только средний размер группы, но и распределение размеров групп.

Полученные результаты и разработанная программа могут быть использованы для предварительного анализа систем массового обслуживания с групповым входящим потоком и выбора минимальной интенсивности обслуживания при заданных требованиях к среднему времени ожидания.

#### **Основные источники информации:**

1. Новиков, О. А. Прикладные вопросы теории массового обслуживания / О. А. Новиков, С. И. Петухов. — М. : издательство «Советское радио», 1969. — 400 с.
2. Клейнрок, Л. Теория массового обслуживания : учебник / Л. Клейнрок. — М. : Машиностроение, 1979. — 432 с.
3. Гнеденко, Б. В. Введение в теорию массового обслуживания : учебник / Б. В. Гнеденко, И. Н. Коваленко. — Москва : Издательство «Наука», ГРФМЛ, 1966. — 432 с.
4. Кофман, А. Массовое обслуживание. Теория и приложения / А. Кофман, Р. Крюон. — М. : Мир, 1965. — 303 с.
5. Самуйлов, К. Е. Построение и анализ моделей системы с групповым обслуживанием заявок / К. Е. Самуйлов, Н. В. Першаков, И. А. Гудкова // Вестник РУДН. Серия Математика. Информатика. Физика. — 2007. — Т. 32, № 3–4. — С. 45–52.
6. Клейнрок, Л. Вычислительные системы с очередями : учебник / Л. Клейнрок. — М. : Мир, 1979. — 600 с.
7. Medhi, J. Stochastic Models in Queueing Theory / J. Medhi. — 2nd ed. —

Amsterdam : Elsevier, 2003. — 501 p.

8. Бочаров, П. П. Теория массового обслуживания : учебник / П. П. Бочаров, А. В. Печинкин. — Москва : Изд-во РУДН, 1995. — 529 с.