

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Саратовский национальный исследовательский государственный университет
имени Н.Г. Чернышевского»

На правах рукописи



Осипов Олег Александрович

СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ ПРОИЗВОЛЬНОЙ ТОПОЛОГИИ
С ДЕЛЕНИЕМ И СЛИЯНИЕМ ТРЕБОВАНИЙ

05.13.18 — Математическое моделирование, численные методы
и комплексы программ

Диссертация
на соискание учёной степени
кандидата физико-математических наук

Научный руководитель
кандидат физико-математических наук,
доцент Тананко Игорь Евстафьевич

Саратов — 2018

Оглавление

Введение	4
Глава 1. Обзор основных результатов исследования сетей массового обслуживания с делением и слиянием требований	
1.1 Сети параллельных систем обслуживания	13
1.2 Сети с произвольной топологией	27
1.3 Моделирование реальных систем	29
1.4 Длительность пребывания требований в сетях массового обслуживания	30
Глава 2. Сети обслуживания с бесконечноприборными базовыми системами	
2.1 Описание сети массового обслуживания	33
2.1.1 Типы систем обслуживания	34
2.1.2 Сигнатура фрагмента	35
2.1.3 Маршрутизация фрагментов	37
2.2 Анализ сети обслуживания	38
2.2.1 Потоки в сети обслуживания	38
2.2.2 Длительность пребывания требований в сети обслуживания	40
2.3 Пример анализа сети обслуживания с делением и слиянием требований	47
2.4 Стационарные характеристики для элементарной сети с делением и слиянием требований	49
Глава 3. Сети обслуживания, зависящие от нагрузки	
3.1 Метод анализа сети обслуживания, состоящей из одноприборных базовых систем	56

3.2	Задача оптимизации сети обслуживания с делением и слиянием требований	62
3.3	Моделирование сетей с многопутевой маршрутизацией	68
Глава 4. Комплекс программ имитационного моделирования и численного анализа сетей массового обслуживания с делением и слиянием требований		
	с делением и слиянием требований	73
4.1	Структура имитационной модели	73
4.2	Программа численных расчётов	77
4.3	Вычислительные аспекты нахождения параметров длительности пребывания требований в сети обслуживания	78
4.4	Уменьшение вычислительной сложности на основе уменьшения числа фаз	79
4.5	Аспекты практического использования комплекса	80
Заключение		88
Список литературы		89

Введение

Актуальность и степень разработанности темы исследования. Реальные системы, в которых имеет место параллельная и распределённая обработка [1] (многопроцессорные системы, GRID-системы, распределённые базы данных, сети передачи данных), получают всё большее распространение. В таких системах поступающие для обработки задачи делятся на подзадачи более простые для выполнения, которые распределяются по системе, занимая выделенные для них ресурсы. После завершения своего выполнения подзадачи освобождают выделенные им ресурсы. При этом исходная задача считается выполненной только после завершения выполнения всех её подзадач.

Структурная и функциональная специфика систем такого класса требует разработки новых эффективных моделей и методов для использования при решении задач анализа, синтеза и оптимизации. Для описания и анализа таких систем используются разнообразные математические абстракции: модель акторов, сети Петри, потоки работ, модели теории массового обслуживания [2; 3].

Сети массового обслуживания являются математическими моделями, используемыми для анализа дискретных стохастических систем с сетевой структурой, эффективность применения которых обусловила интенсивное развитие в течение последних пяти десятилетий теории сетей массового обслуживания [4—41].

Большой вклад в развитие теории, методов анализа, оптимизации и синтеза сетей массового обслуживания внесли Г. П. Башарин, А. А. Боровков, П. П. Бочаров, В. М. Вишневецкий, В. А. Ивницкий, Ю. И. Митрофанов, А. В. Печинкин, В. В. Рыков. Среди зарубежных специалистов необходимо отметить значительный вклад таких учёных как F. Baccelli, K. Chandy, P. Harrison, J. Jackson, F. Kelly, L. Kleinrock, M. Reiser, D. Towsley, J. Walrand.

Сети массового обслуживания с делением и слиянием требований (fork-join queueing networks) [42—44] являются математическими моделями, используемыми для анализа дискретных стохастических систем с параллельным и рас-

пределённым принципами функционирования. Ключевой особенностью в сетях обслуживания с делением и слиянием требований является деление поступающих требований на части — фрагменты, которые обслуживаются параллельно в системах сети обслуживания, и последующее объединение обслуженных фрагментов в исходные требования. Требование считается выполненным и покидает сеть только после окончания обслуживания всех его фрагментов.

В большинстве работ по сетям обслуживания с делением и слиянием требований [45—55] рассматриваются сети обслуживания, состоящие из множества параллельных систем массового обслуживания, а основным результатом является определение длительности пребывания требований в сети обслуживания. Поиск стационарного распределения осложнён тем, что оно не имеет простой мультипликативной формы даже в случае сетей, образованных множеством параллельных систем обслуживания. Некоторые частные случаи расширения классической топологии параллельных систем обслуживания обсуждаются в работах [56—59].

Стоит отметить, что реальным системам в большинстве случаев свойственна более сложная структура, а процесс выполнения задач в них может включать несколько этапов обработки, а также многократное деление и объединение получаемых при этом подзадач. Однако существующие математические модели сетей с параллельной топологией не позволяют адекватно описывать такое многообразие реальных систем с параллельным и распределённым принципами функционирования.

Таким образом, актуальной является задача, связанная с построением математических моделей сетей массового обслуживания произвольной топологии с делением и слиянием требований и разработкой методов анализа указанных сетей обслуживания.

В диссертационном исследовании рассматривается класс открытых сетей массового обслуживания произвольной топологии с делением и слиянием требований.

Цель и задачи исследования. Целью данной работы является изучение сетей массового обслуживания произвольной топологии с делением и слиянием требований, получение стационарных характеристик этих сетей обслуживания, оптимизация их параметров, использование сетей указанного класса для моделирования реальных систем.

Для достижения поставленной цели были решены следующие **задачи**:

1. Формальное построение сетей массового обслуживания произвольной топологии с делением и слиянием требований.
2. Разработка методов анализа сетей массового обслуживания произвольной топологии с делением и слиянием требований.
3. Исследование зависимости характеристик сетей массового обслуживания от их параметров.
4. Разработка и реализация алгоритмов для расчёта стационарных характеристик и оптимизации сети обслуживания.

Научная новизна результатов, представленных в диссертации.

Постановка задач и полученные результаты являются новыми.

1. Впервые предложены математические модели сетей массового обслуживания с делением и слиянием требований, позволяющие учитывать произвольную топологию, многократное деление и объединение фрагментов, зависимость маршрутизации от типа фрагментов, а также наличие сложных взаимосвязей между фрагментами одного требования, что позволяет повысить адекватность представления реальных систем с параллельным и распределённым принципами функционирования.
2. Доказано, что длительность пребывания требований в сети массового обслуживания произвольной топологии с делением и слиянием требований в случае бесконечноприборных базовых систем имеет фазовое распределение. Разработаны алгоритмы, позволяющие найти параметры фазового распределения длительности пребывания требований в сети обслуживания.

3. Получена форма стационарного распределения вероятностей состояний элементарной сети обслуживания с бесконечноприборными базовыми системами. Для определения стационарного распределения предложено использовать специальную сеть размещений. Разработан алгоритм построения сети размещений для элементарной сети обслуживания. Показано, что сеть размещений является сетью Джексона. Найдена связь между стационарными распределениями исходной сети с делением и слиянием требований и соответствующей сети размещений.
4. Для сетей обслуживания с одноприборными базовыми системами предложен подход для приближенного нахождения стационарных характеристик, основанный на идее декомпозиции и агрегирования, который позволяет использовать ранее полученные автором результаты, связанные с анализом сетей с бесконечноприборными базовыми системами. Проведено исследование точности получаемых характеристик посредством разработанного комплекса программ имитационного и численного анализа сетей массового обслуживания с делением и слиянием требований.

Методы исследования. В диссертационной работе использовались результаты теории вероятностей, случайных процессов, теории цепей Маркова, теории массового обслуживания, теории сетей массового обслуживания, методы имитационного моделирования.

Теоретическая и практическая значимость работы. Предложенные модели существенно расширяют круг задач, решаемых в теории массового обслуживания, поскольку позволяют рассмотреть особенности структуры и функционирования сетей обслуживания, возникающие в случае произвольной топологии сетей и связанные прежде всего с усложнением применяемых методов маршрутизации фрагментов, возможностью многократного деления и объединения фрагментов, а также наличием зависимостей между фрагментами одного требования.

Получен вид функции распределения длительности пребывания требований в сети обслуживания произвольной топологии с делением и слиянием требований. Указанная характеристика сети обслуживания является ключевым показателем качества обслуживания для реальных систем, моделируемых посредством моделей теории массового обслуживания. Найдено стационарное распределение вероятностей состояний элементарной сети обслуживания с бесконечноприборными базовыми системами путём построения соответствующей сети размещений. Показано, что сеть размещений является сетью Джексона, что существенно уменьшает вычислительную сложность нахождения стационарного распределения.

Представленные в диссертационной работе результаты могут быть применены для математического моделирования стохастических систем с параллельным и распределённым принципами функционирования, а также в задачах оптимизации и синтеза указанных систем. В качестве примера в диссертации рассмотрена модель сети передачи данных с многопутевой маршрутизацией в виде сети массового обслуживания произвольной топологии с делением и слиянием требований.

Предложенные математические модели также могут быть применены для решения задач:

- проектирования сети передачи данных с заданной пропускной способностью,
- расчёта характеристик производительности GRID-системы и оптимизации её архитектуры,
- нахождения условий для обеспечения некоторого уровня качества обслуживания пользователей распределённой базы данных.

Связь работы с крупными научными проектами. В основу диссертации положены результаты научных исследований, выполненных при участии автора в Саратовском государственном университете, по включённой в план НИР СГУ теме: «Развитие теории и методов анализа сетей массового обслуживания с групповыми переходами требований, распределением нагруз-

ки и нестационарными структурами, разработка методов управления сетями и методов анализа сетей с управлением» (шифр «Ресурс», регистрационный № АААА-А17-117110220045-8).

Достоверность полученных точных результатов обеспечивается корректными доказательствами всех приведённых в работе утверждений. Достоверность приближённых методов подтверждают результаты имитационного моделирования.

Положения, выносимые на защиту. На защиту выносятся следующие результаты исследования:

1. Математические модели сетей массового обслуживания произвольной топологии с делением и слиянием требований.
2. Нахождение вероятностных характеристик сетей массового обслуживания произвольной топологии с делением и слиянием требований.
3. Комплекс программ имитационного и численного анализа сетей массового обслуживания произвольной топологии с делением и слиянием требований.

Личное участие автора в получении результатов, изложенных в диссертации. Личное участие автора заключается в исследовании рассматриваемых моделей, получении аналитических результатов для них, разработке программ численного и имитационного моделирования.

Соответствие паспорту специальности Диссертационная работа выполнена в соответствии с паспортом специальности 05.13.18 «Математическое моделирование, численные методы и комплексы программ» и включает оригинальные результаты в области математического моделирования, численных методов и комплексов программ.

Исследование, представленное в работе, соответствует следующим разделам паспорта специальности: п. 1 (Разработка новых математических методов моделирования объектов и явлений), п. 2 (Развитие качественных и приближённых аналитических методов исследования математических моделей), п. 4 (Реализация эффективных численных методов и алгоритмов в виде комплек-

сов проблемно-ориентированных программ для проведения вычислительного эксперимента), п. 8 (Разработка систем компьютерного и имитационного моделирования).

Публикации. Основные результаты по теме диссертации изложены в 9 работах, из них 3 статьи в журналах, входящих в Перечень рецензируемых научных изданий, рекомендованных Высшей аттестационной комиссией при Минобрнауки Российской Федерации для опубликования основных научных результатов диссертации, 5 — в сборниках тезисов и материалов конференций, также получено 1 свидетельство о регистрации программы для ЭВМ.

В работах, опубликованных в соавторстве, вклад соискателя состоит в получении теоретических результатов и их интерпретации, а вклад научного руководителя — в постановке задач и обсуждении методов их решения.

Степень достоверности и апробация результатов. Основные результаты работы докладывались и обсуждались на следующих научных семинарах и конференциях:

- научные семинары кафедры системного анализа и автоматического управления Саратовского государственного университета (2015–2018 г.),
- XXII Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2015» (2015 г., МГУ, Москва),
- XVI Всероссийский симпозиум по прикладной и промышленной математике (осенняя сессия), (2015 г., Сочинский государственный университет, Сочи – Дагомыс),
- Всероссийская конференция с международным участием «Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем» (2016, 2017 г., РУДН, Москва),
- Международная научная конференция «Компьютерные науки и информационные технологии» (2016 г., СГУ, Саратов),
- XVI Международная конференция имени А.Ф. Терпугова «Информационные технологии и математическое моделирование» (2017 г., Казань)

ский национальный исследовательский технологический университет, Казань).

Структура и объём диссертации. Диссертация состоит из введения, четырёх глав и заключения. Полный объём диссертации составляет 102 страницы, включая 20 рисунков и 7 таблиц. Список литературы содержит 123 наименования.

Краткое содержание диссертационной работы.

Во введении обосновывается актуальность темы диссертации, формулируются цели и задачи исследования, а также указываются положения, выносимые на защиту.

В первой главе содержится обзор основных результатов по теории сетей массового обслуживания с делением и слиянием требований. Описаны известные приближённые и точные методы анализа сетей обслуживания, приведены примеры использования сетей обслуживания с делением и слиянием требований в качестве моделей реальных систем.

Вторая глава посвящена анализу сетей обслуживания с делением и слиянием требований произвольной топологии в случае бесконечноприборных базовых систем обслуживания. Введено формальное описание для сетей обслуживания с произвольной топологией.

Доказано, что длительность пребывания требований в сети обслуживания имеет распределение фазового типа и предложен метод для нахождения его параметров. Также для элементарной сети обслуживания исследовано стационарное распределение вероятностей состояний сети обслуживания.

В третьей главе обсуждаются зависимые от нагрузки сети с делением и слиянием требований. Предполагается, что в этом случае интенсивность обслуживания фрагментов на каждом приборе базовой системы зависит от интенсивности поступающего в эту систему потока. Представлен пример, демонстрирующий, что посредством введения такой зависимости можно получить приближённые результаты для сетей обслуживания с делением и слиянием требований, в которых все базовые системы являются одноприборными. Рассмотрена задача

оптимизации распределения потоков в сети обслуживания. Приведён пример модели сети передачи данных с многопутевой маршрутизацией в виде сети массового обслуживания с делением и слиянием требований.

В четвертой главе описан разработанный комплекс численного и имитационного моделирования на основе полученных в главах 2 и 3 результатов. Приводятся примеры исследования гипотетических сетей массового обслуживания с делением и слиянием требований, обсуждаются полученные результаты.

В заключении резюмируются полученные в диссертационной работе результаты, а также описываются возможные направления дальнейших исследований.

Благодарности. Выражаю глубокую благодарность научному руководителю кандидату физико-математических наук, доценту И. Е. Тананко за участие в постановке задач и руководство ходом исследований, а также искренне благодарю коллектив кафедры системного анализа и автоматического управления Саратовского государственного университета за помощь, оказанную при подготовке диссертации.

Глава 1. Обзор основных результатов исследования сетей массового обслуживания с делением и слиянием требований

Данная глава представляет собой обзор основных опубликованных результатов исследования сетей массового обслуживания с делением и слиянием требований. В разделе 1.1 приведены результаты работ, посвящённых сетям массового обслуживания, состоящим из параллельных систем обслуживания. В разделе 1.2 обсуждаются сети обслуживания с произвольной топологией. Примеры использования сетей данного класса для моделирования реальных систем, обсуждаются в разделе 1.3. В разделе 1.4 приведены методы нахождения длительности пребывания требований в сетях массового обслуживания.

1.1 Сети параллельных систем обслуживания

В большинстве работ по сетям с делением и слиянием требований [45—53] рассматриваются сети массового обслуживания, состоящие из M параллельных систем массового обслуживания S_1, \dots, S_M . Будем называть такие сети обслуживания классическими. Поступающее требование делится некоторым абстрактным устройством F (fork-point) на M фрагментов, которые затем обслуживаются в параллельных системах обслуживания (рисунок 1.1).

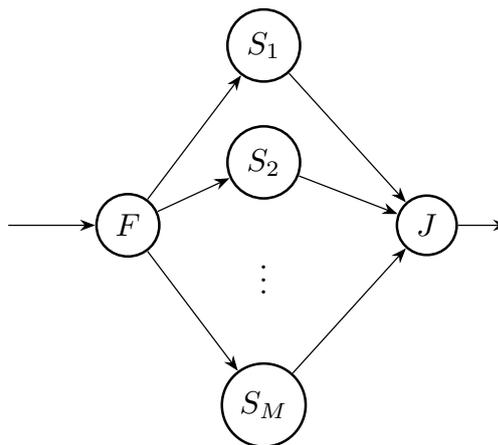


Рисунок 1.1 — Сеть с классической топологией

После окончания обслуживания всех фрагментов исходного требования они снова объединяются другим абстрактным устройством J (join-point) в требование, которое покидает сеть обслуживания. В зависимости от возможных

вариантов деления требований при поступлении и слияния фрагментов после завершения обслуживания выделяют три основных класса сетей обслуживания с делением и слиянием требований [46]:

- с центральным делением без синхронизирующей очереди (centralized splitting model without synchronization queue, split-merge model),
- с центральным делением и синхронизирующей очередью (centralized splitting model with synchronization queue),
- с распределённым делением и синхронизирующей очередью (distributed splitting model with synchronization queue).

Также отметим такие родственные модели как, например, модель независимых приборов (independent server model) [60; 61] и модель командного обслуживания (team service model) [62].

Распределение фрагментов поступающих требований по системам сети с делением и слиянием требований, как и число фрагментов, получаемых при делении одного требования, может быть задано некоторой детерминированной или вероятностной стратегией [45; 48; 63–66]. Так, в работах [45; 48; 64–66] требования каждый раз делятся на одинаковое число фрагментов, по одному фрагменту для каждой системы. С другой стороны, в [63] рассматривается модель, в которой каждое требование может быть поделено на случайное число фрагментов, поступающих в системы в соответствии с некоторым распределением вероятностей.

Далее будут рассматриваться сети обслуживания с *распределённым делением и синхронизирующей очередью*. Поступающее требование мгновенно делится в F на M фрагментов, которые сразу же распределяются по M параллельным системам обслуживания. После завершения обслуживания фрагмента в системе, он покидает эту систему обслуживания и направляется в J , где находится до тех пор, пока не будет завершено обслуживание всех фрагментов, полученных при делении исходного требования. Как только все фрагменты требования оказываются в J происходит их объединение в исходное требование, которое сразу же покидает сеть обслуживания.

Пусть в рассматриваемой сети все параллельные системы обслуживания являются одноприборными системами с дисциплиной FCFS. Положим, что последовательность $\{a_n\}_{n=1}^{\infty}$ задаёт длительности интервалов времени между последовательно поступающими в сеть требованиями, то есть момент поступления A_n для n -го требования

$$A_n = \sum_{i=1}^n a_i, \quad n = 1, 2, \dots, \quad (1.1)$$

полагаем, что $A_0 = 0$. После поступления в систему n -го требования оно делится на M фрагментов. Для n -го требования обозначим через x_n^m длительность обслуживания его m -го фрагмента, тогда для длительности W_{n+1}^m ожидания начала обслуживания m -го фрагмента $n + 1$ -го требования справедливо

$$W_{n+1}^m = (W_n^m + x_n^m - a_{n+1})^+, \quad m = 1, \dots, M, n = 0, 1, \dots, \quad (1.2)$$

где $(y)^+ = \max(y, 0)$.

Тогда длительность R_n^m пребывания m -го фрагмента n -го требования будет определена как [58; 67]

$$R_n^m = W_n^m + x_n^m, \quad m = 1, \dots, M, n = 0, 1, \dots \quad (1.3)$$

Длительность R_n пребывания n -го требования в сети

$$R_n = \max_{m=1, \dots, M} R_n^m, \quad n = 0, 1, \dots \quad (1.4)$$

Одним из первых исследований, посвящённых таким сетям обслуживания является работа [45], в которой рассматривается сеть обслуживания с пуассоновским входящим потоком требований интенсивности $\lambda = 1$. Каждое из требований при поступлении делится на $M = 2$ фрагмента, каждый из которых поступает в одну из двух систем массового обслуживания типа $M/M/1$ с интенсивностями обслуживания фрагментов μ_1 и μ_2 для первой и второй системы массового обслуживания соответственно.

Состояние сети определяется парой (i, j) , где i — число фрагментов в первой системе обслуживания, j — во второй. Обозначим через $p(i, j)$ стационарную вероятность состояния (i, j) . Стационарный режим в сети существует, когда

$\min\{\mu_1, \mu_2\} > 1$. Уравнения глобального равновесия для стационарного распределения имеют вид

$$(1 + \mu_1 \chi(i \geq 1) + \mu_2 \chi(j \geq 1)) p(i, j) = \mu_1 p(i + 1, j) + \mu_2 p(i, j + 1) + \chi(i \geq 1, j \geq 1) p(i - 1, j - 1), \quad i, j \geq 0. \quad (1.5)$$

Здесь $\chi(A)$ — индикатор события A ,

$$\chi(A) = \begin{cases} 1, & \text{если событие } A \text{ выполнено,} \\ 0, & \text{иначе.} \end{cases}$$

Основным результатом работы стало нахождение производящей функции $P(z, w) = \sum_{i, j \geq 0} p(i, j) z^i w^j$ стационарного распределения, поиск которой приводит к функциональному уравнению

$$Q(z, w) P(z, w) = N(z, w), \quad |z|, |w| \leq 1, \quad (1.6)$$

где

$$Q(z, w) = (1 + \mu_1 + \mu_2) z w - \mu_1 w - \mu_2 z - z^2 w^2, \quad (1.7)$$

$$N(z, w) = \mu_2 z (w - 1) P(z, 0) + \mu_1 w (z - 1) P(0, w). \quad (1.8)$$

Здесь функции $P(z, 0)$, $P(0, w)$ должны быть определены. В работе получено выражение для производящих функций, которое в случае $\mu_1 = \mu_2 = \mu$ имеет вид

$$P(z, 0) = P(0, z) = \frac{(\mu - 1)^{3/2}}{\mu(\mu - z)^{1/2}}. \quad (1.9)$$

В работе [68] изучена зависимость между числом фрагментов в каждой из двух систем обслуживания и получено несколько асимптотических результатов.

Для сетей обслуживания с произвольным числом параллельных систем, получены только приближённые результаты [48; 57; 58; 69—73]

В работах [57; 58] изучаются граничные значения для математического ожидания длительности пребывания требований в сети обслуживания. Подход основан на поиске точных решений известными методами для приближённо

сформулированной исходной модели. Так анализируются две различные системы, методы анализа которых известны. Для нахождения нижней границы заменяют исходную сеть массового обслуживания сетью с детерминированным входящим потоком [74].

В работе [48] предложен приближённый метод, названный авторами «метод аппроксимации масштабированием» (scaling approximation method), который применяется для анализа экспоненциальных сетей с произвольным числом параллельных систем обслуживания. Метод позволяет найти математическое ожидание T_M длительности пребывания требования в сети обслуживания, состоящей из M параллельных систем.

Предполагается, что в сеть поступает пуассоновский поток с интенсивностью λ , все системы в сети суть одноприборные системы с экспоненциальной длительностью обслуживания, интенсивность обслуживания фрагментов во всех системах одинакова и равна μ . Для рассматриваемой сети, состоящей из двух систем, было найдено *точное* значение для T_2

$$T_2 = \frac{12 - \rho}{8} T_1, \quad (1.10)$$

где $T_1 = 1/(\mu - \lambda)$, $\rho = \lambda/\mu$.

Предложенный метод основан на существовании нижней и верхней границ для величины T_M . Верхняя граница получена в предположении о независимости длительностей пребывания фрагментов в системах обслуживания, тогда как нижняя граница получена при допущении отсутствия очередей фрагментов в системах обслуживания. Тогда в силу того, что длительность пребывания требования в сети обслуживания определяется как максимальное значение из длительностей пребывания его фрагментов в системах, получим

$$H_M \frac{1}{\mu} \leq T_M \leq H_M T_1, \quad (1.11)$$

где H_M обозначает M -ую частичную сумму гармонического ряда

$$H_M = \sum_{i=1}^M \frac{1}{i}.$$

Учитывая одинаковую скорость роста $O(\ln M)$ этих границ, получено следующее приближённое выражение для T_M

$$T_M \approx \left(\frac{H_M}{H_2} + \frac{4}{11} \left(1 - \frac{H_M}{H_2} \right) \rho \right) T_2, \quad M \geq 2. \quad (1.12)$$

Этот метод был расширен и применён для анализа сетей, в которых распределение длительности обслуживания в системах сети не является экспоненциальным в работе [69].

Для произвольных входящего потока и длительностей интервалов обслуживания в работе [70] описан подход, основанный на анализе сети обслуживания в условиях слабой и сильной нагрузки. На основе предельных значений нагрузки выполняется интерполяция математического ожидания длительности пребывания требований в сети обслуживания.

Для случая экспоненциальной сети обслуживания выражение для T_M имеет вид

$$T_M \approx \left[H_M + \left(\left(\sum_{i=1}^M \binom{M}{i} (-1)^{i-1} \sum_{m=1}^i \binom{i}{m} \frac{(m-1)!}{i^{m+1}} \right) - H_M \right) \frac{\lambda}{\mu} \right] \frac{1}{\mu - \lambda}.$$

Переходный режим функционирования сети обслуживания рассматривается в [71]. Сеть обслуживания состоит из двух одноприборных систем, распределения длительности интервалов времени между последовательно поступающими требованиями и длительности обслуживания могут быть экспоненциальными, Эрланга 2-го порядка, гиперэкспоненциальными. В работе представлен метод для нахождения характеристик сети обслуживания в терминах виртуального времени ожидания [75].

В работе [46] рассматриваются вопросы анализа работоспособности трёх основных типов сетей с делением и слиянием требований, указанных выше. Предполагается, что каждая из M параллельных систем может отказывать. Длительность наработки на отказ и длительность восстановления имеют экспоненциальное распределение. Под структурным состоянием w сети обслуживания будем понимать число работоспособных систем обслуживания. Если сеть обслуживания находится в структурном состоянии w , то каждое поступающее

в сеть требование делится на w фрагментов, которые распределяются для обслуживания по w системам обслуживания, и каждый фрагмент обслуживается с интенсивностью $\frac{w\mu}{M}$.

Для каждого структурного состояния w , определён показатель его прибыльности $r(w)$.

Пусть в начальный момент времени сеть обслуживания находилась в некотором структурном состоянии $w_0 \in \{0, \dots, M\}$. Рассмотрим процесс эволюции сети обслуживания на промежутке времени $[0, t]$. Работоспособность всей сети в целом за указанный промежуток времени при начальном структурном состоянии w_0 определяется показателем накопленной прибыли $R(t, w_0)$, который задан выражением

$$R(t, w_0) = \sum_{w=0}^M r(w)u(w),$$

где $u(w)$ обозначает длительность интервала времени, в течение которого сеть обслуживания была в структурном состоянии w . Очевидно, что

$$\sum_{w=0}^M u(w) = t.$$

Далее в работе предложен подход, основанный на методе униформизации [76], который позволяет произвести приближённый численный расчёт функции распределения величины накопленной прибыли $R(t, w)$ для различных вариаций сети обслуживания.

Применение диффузионной аппроксимации к исследованию сетей изучаемого класса отражено в работах [72; 73]. В [72] рассмотрена экспоненциальная сеть массового обслуживания, функционирующая в режиме высокой нагрузки. Получены приближённые выражения для стационарного распределения числа фрагментов в сети, а также для длительности остаточного периода занятости.

Замкнутые сети массового обслуживания с делением и слиянием требований рассматриваются в статьях [51; 52; 77—80].

В работах [77; 78] изучается замкнутая экспоненциальная сеть массового обслуживания, состоящая из M систем обслуживания, в которой циркулиру-

ют Q требований. Ключевым отличием от ранее рассмотренных сетей обслуживания является тот факт, что деление требований в F не мгновенное. Здесь под F понимается система массового обслуживания с одним обслуживающим прибором.

Циркулирующее по сети обслуживания требование проходит последовательно следующие стадии: обслуживание в F , деление на M фрагментов, обслуживание фрагментов на параллельных системах, объединение фрагментов, поступление полученного требования снова в F . Для введённой сети рассматриваются два метода анализа: первый основан на точном решении уравнений равновесия и применим для задач малой размерности, второй — приближённый метод анализа, который основан на теореме Нортонa [81].

Приближённый метод анализа сети обслуживания сводится к замене сети массового обслуживания с M параллельными системами обслуживания сетью с двумя параллельными системами.

На первом шаге рассматривается замкнутая сеть массового обслуживания аналогичная изучаемой, состоящая только из двух параллельных систем обслуживания S_1, S_2 , однако требования в F не задерживаются (fork-point «закорачивается»). Такая сеть обслуживания имеет мультипликативную форму стационарного распределения. Пусть $\Lambda(i)$ есть пропускная способность сети, когда в ней находится i требований, $i = 1, \dots, Q$. Заменяем эту сеть эквивалентной системой обслуживания C_1 , интенсивность обслуживания которой зависит от числа требований в ней и равна $\Lambda(i)$ для i требований, то есть совпадает с пропускной способностью этой сети. Теперь образуем новую сеть из двух параллельных систем C_1 и S_3 .

Аналогично эта подсеть сводится к новой системе обслуживания C_2 , продолжая этот процесс получается сеть обслуживания, состоящая из композиционной системы C_{M-2} и системы обслуживания S_M .

После этого параллельная подсеть исходной сети заменяется на подсеть, состоящую из двух параллельных систем.

Отмечено, что данный подход даёт не только хорошую точность (относительная погрешность, полученная в результате сравнения с разработанной численной моделью составила 3%), получаемые оценки для математического ожидания длительности пребывания требований в сети и пропускной способности сети являются верхней и нижней границами для искомых параметров соответственно. Аналогичный подход был также применён для анализа открытых сетей обслуживания в [82].

Экспоненциальные сети обслуживания с многоприборными системами рассмотрены в [47], получено выражение для аппроксимации функции распределения $F(t)$ длительности пребывания требований в сети для стационарного режима функционирования. Пусть κ_i и μ_i , $i = 1, \dots, M$, задают число приборов в системах обслуживания сети и интенсивность обслуживания для каждого прибора системы соответственно, тогда справедливо

$$F(t) \approx 1 - \sum_{i=1}^M c_i (1 - F_i(t)) \prod_{j=1}^{i-1} F_j(t) = \sum_{i=1}^M (c_i - c_{i+1}) \prod_{j=1}^i F_j(t),$$

где $F_i(t)$ — функция распределения длительности пребывания фрагментов в системе обслуживания $M/M/\kappa_i$, $c_{M+1} = 0$, $c_1 = 1$, $c_2 = 1 - q_1/4$, и

$$c_i = \left(1 - \frac{q_{i-2}}{8}\right) \left(1 - \frac{q_{i-1}}{4}\right), \quad 3 \leq i \leq M, \quad (1.13)$$

здесь q_i — стационарная вероятность того, что поступающий в систему $M/M/\kappa_i$ фрагмент будет ожидать начала обслуживания в очереди, выражение для которой получено в [75].

Также для случая одноприборных систем обслуживания в работе получено стационарное распределение числа требований в сети обслуживания.

Модификация классической сети с M параллельными системами обслуживания при предположении о детерминированном входящем потоке рассматривается в [50]. Каждая из параллельных систем является системой массового обслуживания типа $D/M/1$. Пусть a обозначает длительность интервалов времени между последовательными требованиями. Для данной модели, используя

метод вложенных цепей Маркова, получено точное выражение для стационарного распределения числа требований в сети обслуживания в терминах производящих функций. Для случая $M = 2$ стационарные вероятности имеют вид

$$\begin{aligned} p(i, j) &= C(1 - \sigma_1)\sigma_1^{i-1}(1 - \sigma_2)\sigma_2^{j-1}, \quad i, j \geq 1, \\ p(0, j) &= (1 - \sigma_2)\sigma_2^{j-1}(\rho_2 - C), \quad j \geq 1, \\ p(i, 0) &= (1 - \sigma_1)\sigma_1^{i-1}(\rho_1 - C), \quad i \geq 1, \\ p(0, 0) &= 1 - \rho_1 - \rho_2 + C, \end{aligned}$$

где

$$\begin{aligned} C &= \frac{1 - \sigma_1\sigma_2}{\rho_1^{-1}(1 - \sigma_1) + \rho_2^{-1}(1 - \sigma_2)}, \\ \rho_m &= (a\mu_m)^{-1}, \quad m = 1, 2; \end{aligned}$$

σ_m — единственный действительный корень в интервале $(0, 1)$ для уравнения

$$x = e^{-a\mu_m(1-x)}, \quad m = 1, 2,$$

μ_m — интенсивность обслуживания в m -ой системе обслуживания.

В работах [65; 83—85] для анализа различных модификаций сетей массового обслуживания с делением и слиянием требований используются матричные методы.

В работе [65] рассматривается один из вариантов обобщения классических сетей обслуживания, состоящих из параллельных систем. Предполагается, что каждая из M систем обслуживания имеет один обслуживающий прибор, а число мест для ожидания в очереди каждой из систем (локальные FCFS очереди) конечно. Требование, поступающее в очередь может быть поделено в F только в том случае, когда в очереди каждой из систем есть свободное место, иначе требование поступает в неограниченную по вместимости очередь (глобальная FCFS очередь). Как только в каждой локальной очереди появится свободное место для нового фрагмента, требование переходит из глобальной очереди и делится

на фрагменты, которые распределяются по локальным очередям. Фрагменты, которые образуют единое требование, требуют для своего выполнения несколько фаз обслуживания с наличием взаимосвязи между фрагментами. Рассматривается несколько дисциплин обслуживания, которые направлены на повышение производительности сети. Для сетей, в которых число состояний невелико, используется точный метод, для случая, когда точный метод оказывается слишком затратным, используется разработанный приближённый метод.

Работа [83] основана на преобразовании уравнений глобального равновесия для сети параллельных систем к такой системе уравнений, решение которой может быть найдено используя матрично-геометрические методы. На основании этого получены точные решения для модифицированных сетей обслуживания, которые являются верхней и нижней границами для характеристик исходной сети обслуживания. Так, верхняя граница получена пренебрегая возможностью больших расхождений в значениях компонент процесса $\{(n_1(t), \dots, n_M(t)), t \geq 0\}$, в котором $n_i(t)$ задаёт число фрагментов в системе обслуживания с номером i в момент времени t , и обнулением соответствующих интенсивностей перехода.

Аналогичный подход, связанный с усечением пространства состояний, применяется в работе [86] для нахождения распределения длительности пребывания требований в сети обслуживания. Для сети с марковским входящим потоком (МАР, Markovian arrival process) [87], в которой длительность обслуживания имеет распределение фазового типа, построен модельный процесс описывающий её эволюцию, а длительность пребывания требования в сети обслуживания находится путём отдельного изучения процессов ожидания в очередях и обслуживания на приборах.

В том случае, когда системы сети достаточно загружены и усечение пространства приводит к задаче большой вычислительной сложности, предлагается эффективный вычислительный метод решения.

В статье [84] рассматривается система обслуживания, состоящая из M обслуживающих приборов и очереди бесконечной вместимости с дисциплиной

обслуживания FCFS. В систему обслуживания поступает пуассоновский поток требований с интенсивностью Λ . Обозначим через w класс требования и положим, что для всех требований, поступающих из источника, $w = 1$. Требование, поступающее на обслуживание, делится на фрагменты так, что они занимают все свободные в данный момент времени обслуживающие приборы: если имеется $1 \leq d \leq M$ свободных приборов, то поступающее требование разделяется на d идентичных фрагментов с классом $w = d$. Каждый из фрагментов мгновенно занимает свободный прибор и начинает обслуживаться. То есть деление требования на $d > 1$ фрагментов происходит в том случае, когда в момент поступления этого требования очередь системы была пуста, при этом имелось более одного свободного прибора.

Длительность обслуживания любого фрагмента с классом w на приборе i , $i = 1, \dots, M$, есть экспоненциально распределённая случайная величина с параметром $w\mu_i$, здесь μ_i — интенсивность обслуживания фрагмента (требования) класса $w = 1$.

Фрагмент, обслуживание которого завершено, освобождает обслуживающий его прибор. Требование будет считаться обслуженным только после того, как будет завершено обслуживание всех его фрагментов. Сразу после этого фрагменты требования мгновенно объединяются в исходное требование, которое покидает систему обслуживания.

Для описанной системы обслуживания с использованием матрично-геометрического метода получены точные выражения для основных стационарных характеристик. Особое внимание уделено длительности пребывания требований в системе обслуживания.

Повторные вызовы, возникающие в сетях обслуживания, обсуждаются в [85]. Рассматривается сеть массового обслуживания с делением и слиянием требований, состоящая из M параллельных одноприборных систем обслуживания, каждая с конечным числом B мест для ожидания в очереди. В сеть обслуживания из внешнего источника поступает пуассоновский поток требований с интенсивностью Λ . Вновь поступающее требование, застающее сеть в

состоянии, когда в каждой очереди имеются свободные места для ожидания, делится на M фрагментов, которые распределяются по системам сети и ожидают своего обслуживания в соответствии с дисциплиной FCFS в очередях. Длительность обслуживания фрагментов на приборе системы i имеет экспоненциальное распределение с параметром μ_i . Фрагмент, завершивший своё обслуживание, освобождает обслуживающий его прибор. Требование будет считаться обслуженным только после того, как будет завершено обслуживание всех его фрагментов. Сразу после этого фрагменты требования мгновенно объединяются в исходное обслуженное требование.

В том случае, когда хотя бы одна очередь полностью заполнена, поступившее требование не может быть поделено и переходит в источник повторных вызовов (ИПВ), где после случайной задержки вновь пытается обратиться к системам сети для получения обслуживания. Если требование снова не может получить обслуживание, то оно вновь возвращается в ИПВ. Предполагается, что если в ИПВ есть требования, то длительность интервала времени между повторными вызовами имеет экспоненциальное распределение с параметром γ .

Границы для функции распределения длительности пребывания требований в сети обслуживания рассматриваются в [49]. Отметим, что в работе границы были получены не только в том случае, когда входящий поток образует процесс восстановления, но и в случае МАР-потока. В случае сети обслуживания с пуассоновским входящим потоком интенсивности λ , состоящей из M параллельных одноприборных систем, каждая с интенсивностью обслуживания μ , для функции распределения $F(t)$ длительности пребывания требований в сети была получена следующая граница

$$F(t) \leq 1 - \frac{M\mu}{\lambda} e^{-(\mu-\lambda)t}, \quad t \geq 0.$$

Сети с бесконечноприборными системами обслуживания в условиях большой нагрузки обсуждаются в [88]. Здесь предполагается, что приборы систем обслуживания являются ненадёжными, а длительности обслуживания фрагмен-

тов одного требования на приборах систем могут быть зависимыми случайными величинами.

Проблема уменьшения разброса между фрагментами одного требования рассмотрена в [89; 90]. Под разбросом между фрагментами одного требования понимается длительность интервала времени между завершением обслуживания первого и последнего фрагментов требования. С этой целью в сети вводится управление, заключаемое в детерминированной задержке фрагментов, которые завершили своё обслуживание намного раньше остальных. Показано, что при оптимальном выборе задержек, определяемых в зависимости от текущего состояния сети обслуживания, длительность пребывания требований в сети практически не возрастает, однако разброс между фрагментами значительно сокращается.

Управление в сетях обслуживания с делением и слиянием требований рассмотрено также в работах [91; 92]. В [92] рассматривается следующая модификация классической сети обслуживания, представленная на рисунке 1.2.

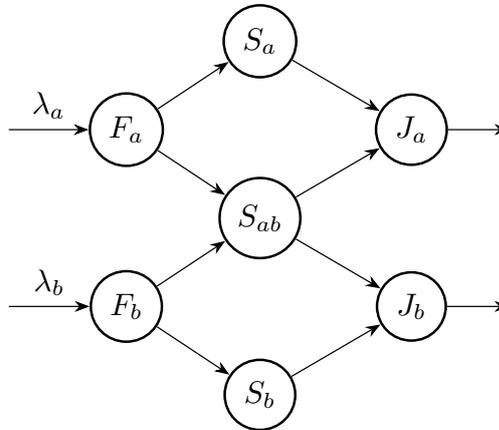


Рисунок 1.2 — Сеть обслуживания с управлением

В сеть обслуживания поступает два класса требований с заданными интенсивностями λ_a и λ_b . Требования каждого из классов делятся в F_a и F_b соответственно на два фрагмента. Фрагменты поступают в одноприборные системы обслуживания в соответствии с дисциплиной FCFS, однако система обслуживания S_{ab} имеет две очереди для фрагментов каждого класса. Фрагменты обслуживаются с интенсивностями μ_a и μ_b в соответствии со своим классом.

После того как система S_{ab} завершит обслуживание некоторого фрагмента, должно быть принято решение, заключающееся в выборе класса фрагмента, который получит обслуживание следующим.

Целью управления является минимизация длительности пребывания требований в сети обслуживания (в том числе минимизация стоимости пребывания).

1.2 Сети с произвольной топологией

В работах [56—59] рассматриваются сети обслуживания, топология которых отличается от классической топологии параллельных систем обслуживания.

В статье [59] изучается тандемная сеть обслуживания, состоящая из K последовательных подсетей, представляющих собой классические сети обслуживания с делением и слиянием требований, каждая из которых состоит из двух одноприборных экспоненциальных систем обслуживания. Предполагается, что в сеть поступает пуассоновский поток требований. Основной целью работы является поиск распределения длительности пребывания требований в сети обслуживания, с этой целью последовательно рассматривается каждая из подсетей и исследуются потоки, выходящие из них.

В работе [56] рассматривается сеть массового обслуживания, топология которой задана ориентированным ациклическим графом $G = (\mathbf{V}, \mathbf{E})$, где \mathbf{V} — множество из M вершин, а \mathbf{E} — множество дуг.

Каждая вершина представляет собой одноприборную систему массового обслуживания с дисциплиной обслуживания FCFS.

Для системы обслуживания S_i , $i = 1, \dots, M$, определяются множества номеров систем обслуживания её непосредственных предшественников $\text{pred}(i)$ и потомков $\text{succ}(i)$:

$$\text{pred}(i) = \{j \in \{1, \dots, M\} : (j, i) \in \mathbf{E}\}; \quad (1.14)$$

$$\text{succ}(i) = \{j \in \{1, \dots, M\} : (i, j) \in \mathbf{E}\}. \quad (1.15)$$

Пусть $\text{succ}(0)$ множество номеров тех систем обслуживания, соответствующие вершины которых не имеют входящих дуг. Предполагается, что фрагменты поступают только в системы с номерами из $\text{succ}(0)$. Аналогично $\text{pred}(M + 1)$ — множество номеров тех систем, соответствующие вершины которых не имеют исходящих дуг.

Для каждой системы обслуживания S_i , $i = 1, \dots, M$, определена последовательность $\{x_n^i\}_{n=0}^\infty$, где x_n^i есть длительность обслуживания для n -го поступившего в неё фрагмента на единственном обслуживающем приборе системы S_i .

Определение 1. Сеть массового обслуживания, определённая выше, называется ациклической сетью с делением и слиянием требований, если выполнены следующие условия:

1. Имеется единственный входящий поток требований, который задан последовательностью $A_0 = 0 < A_1 < \dots < A_n < \dots$

Фрагменты поступают в сеть в момент времени A_n во все обслуживающие системы с номерами из множества $\text{succ}(0)$ (деление фрагментов).

2. Завершение обслуживания некоторого фрагмента в системе S_i не влечёт мгновенное поступление этого фрагмента в системы обслуживания из $\text{succ}(i)$. Фрагмент поступает в систему S_j , $j \in \text{succ}(i)$, в момент времени, который совпадает с моментом завершения обслуживания последнего фрагмента в системах с номерами из $\text{pred}(j)$ (объединение фрагментов).
3. Те фрагменты, которые закончили своё обслуживание в системах с номерами из множества $\text{pred}(B + 1)$ возвращаются в источник требований в момент, когда все фрагменты из $\text{pred}(B + 1)$ закончили своё обслуживание.

То есть полагается, что наличие нескольких исходящих дуг из некоторой вершины приводит каждый раз к делению фрагмента на части, которые попадают во все соответствующие смежные вершины. Наличие нескольких входящих в вершину дуг означает факт объединения фрагментов в этой вершине.

Для заданной сети обслуживания получено условие существования стационарного режима, разработаны приближённые методы для получения длительности пребывания требований в сети обслуживания.

Сети обслуживания такого типа рассматривались также в работах [93; 94]. Отличительной особенностью этих сетей является наличие блокировок из-за ограничения на вместимость очереди в каждой системе сети.

1.3 Моделирование реальных систем

Сети с делением и слиянием требований эффективно применяются для исследования реальных систем. Поэтому во многих работах обсуждается анализ производительности реальных систем с использованием математических моделей в виде сетей обслуживания с делением и слиянием требований.

Результаты по моделированию RAID-массивов представлены в [42; 95; 96]. Полученные в [48] результаты используются в работе также для анализа производительности распределённой базы данных и многопроцессорной ЭВМ. Рассматриваемая база данных состоит из M серверов, выделяются два типа запросов: запрос на чтение и запрос на запись. Предполагается, что запрос на запись приводит к дублированию данных на каждый сервер, то есть происходит его разделение и выполнение на каждом сервере. Запрос на чтение может быть в таком случае обработан любым из серверов. Построенная модель массового обслуживания анализируется, используя метод «аппроксимации масштабированием», вычисляется среднее время выполнения запроса.

В работе [49] сеть массового обслуживания с делением и слиянием требований применяется для анализа метода многопутевой маршрутизации. Поступающий от отправителя в сеть передачи данных поток состоит из групп пакетов одинакового размера, которые передаются по многочисленным маршрутам к получателю. Предполагается, что каждый пакет в группе передаётся по единственному маршруту. Группа будет считаться успешно переданной получателю как только будут доставлены все пакеты группы. Описанная сеть передачи дан-

ных с многопутевой маршрутизацией моделировалась с использованием открытой сети массового обслуживания с делением и слиянием требований, состоящей из M параллельных систем обслуживания.

Обсуждению систем облачных вычислений и анализу параллельных алгоритмов посвящены работы [64; 97—100]. В работе [98] построена модель для GRID-системы.

В [97] для анализа работы производительности параллельных программ исследуется замкнутая сеть обслуживания с делением и слиянием требований. Предложены два оригинальных приближённых метода анализа. Первый основан на декомпозиции сети обслуживания, второй — на замене исходной сети обслуживания последовательностью сетей обслуживания, имеющих мультипликативную форму стационарного распределения.

Статья [64] рассматривает облачный центр, состоящий из нескольких физических машин, которые резервируются пользователями в порядке поступления запросов и могут использоваться совместно для их обработки посредством метода виртуализации. В систему поступают запросы пользователей, содержащие несколько задач (подзапросов), для обработки каждой из которых требуется одна виртуальная машина. Запрос считается выполненным, то есть пользователь получает отклик системы, после обработки всех его подзапросов.

Отметим, что родственные модели (independent server model, split-merge model) используются для анализа вычислительных систем, организованных на базе BOINC (Berkeley Open Infrastructure for Network Computing) [98; 101; 102].

Сети обслуживания исследуемого класса находят применение и для анализа процессов выполнения заказов на складах (в задачах теории управления запасами) [103].

1.4 Длительность пребывания требований в сетях массового обслуживания

Поскольку основной характеристикой, представляющей интерес при исследовании сетей обслуживания с делением и слиянием требований, является

длительность пребывания требований в сети, рассмотрим некоторые известные методы нахождения длительности пребывания требований в сетях обслуживания [104], в которых требования не делятся на фрагменты.

Будем рассматривать открытые сети обслуживания, в которых очереди систем имеют неограниченную вместимость.

Для сетей, стационарное распределение которых имеет мультипликативную форму, возможно получение решения в виде преобразования Лапласа-Стилтьеса для длительности пребывания требований в сети [105]. Также известны результаты для сетей с различными специфическими видами топологии, например: тандем, звезда, система с обратной связью и так далее. Однако получаемые посредством этих методов результаты являются достаточно громоздкими.

Метод ИФТА (independent flow time approximation), предложенный в [106], основан на предположении, что для некоторого помеченного требования, проходящего через последовательность систем в сети обслуживания, распределения вероятностей состояний для каждой системы в моменты поступления этого требования независимы и совпадают с соответствующими стационарными распределениями вероятностей состояний систем обслуживания на момент поступления произвольного требования.

Основанный на методе ИФТА способ был описан в [107] и заключается в нахождении входящих потоков в каждую систему и исследовании каждой системы по отдельности. Если известна длительность пребывания в каждой системе, то посредством «склеивания» этих длительностей определяется длительность пребывания во всей сети обслуживания.

Схожий метод описан в работе [108]. Предполагается, что длительность пребывания требований в каждой системе обслуживания есть случайная величина с фазовым распределением [109], то есть может быть определена как время до поглощения цепи Маркова. Тогда для всей сети в соответствии с топологией можно построить такую цепь Маркова, в которой поглощение будет эквивалентно уходу требования из сети обслуживания. Таким образом будет получен другой процесс — цепь Маркова с непрерывным временем, который аппрок-

симирует исходное распределение. Очевидным достоинством данного метода является то, что он представляет ясную интерпретацию и приводит к задаче вычисления времени до поглощения цепи Маркова.

Глава 2. Сети обслуживания с бесконечноприборными базовыми системами

В данной главе рассматривается класс открытых сетей массового обслуживания произвольной топологии, который является развитием классических сетей массового обслуживания с делением и слиянием требований. Системы обслуживания в рассматриваемой сети поделены на три типа в зависимости от их назначения: бесконечноприборные базовые системы, дивайдеры, интеграторы.

Формально определены процессы, происходящие с фрагментами требований в системах обслуживания каждого типа, а также способ идентификации каждого фрагмента, находящегося в сети обслуживания. Возможные переходы фрагментов в сети заданы при помощи набора матриц передач, определяющих число фрагментов и вероятности переходов фрагментов между системами сети. Получены уравнения потоков фрагментов требований. Показано, что длительность пребывания требований в рассматриваемой сети массового обслуживания с делением и слиянием требований имеет фазовое распределение. Предложен метод определения параметров этой функции распределения. Разработан алгоритм нахождения стационарного распределения вероятностей числа фрагментов в базовых системах сети.

2.1 Описание сети массового обслуживания

Рассматривается открытая сеть массового обслуживания \mathcal{N} с одним классом требований и множеством $\{N_1, \dots, N_L\}$ систем массового обслуживания [110; 111]. В сеть из источника N_0 поступает пуассоновский поток требований с интенсивностью Λ_0 . Любое из требований может многократно делиться на части — фрагменты. Полученные фрагменты обслуживаются независимо друг от друга, переходят по системам сети обслуживания. Каждый из фрагментов может снова разделиться на новые фрагменты. Перед уходом из сети все фрагменты объединяются в исходное обслуженное требование.

Далее будем использовать термин «фрагмент» для описания и требования, и его частей.

2.1.1 Типы систем обслуживания

Пусть $X = \{1, \dots, L\}$ — множество номеров систем обслуживания. Системы в сети обслуживания \mathcal{N} поделены на три непустых, непересекающихся множества в соответствии с их номерами:

1. $X_{\mathcal{S}} = \{x_1^{\mathcal{S}}, \dots, x_M^{\mathcal{S}}\}$ — множество номеров *базовых систем*;
2. $X_{\mathcal{F}} = \{x_1^{\mathcal{F}}, \dots, x_K^{\mathcal{F}}\}$ — множество номеров *дивайдеров*;
3. $X_{\mathcal{J}} = \{x_1^{\mathcal{J}}, \dots, x_K^{\mathcal{J}}\}$ — множество номеров *интеграторов*.

Заметим, что число дивайдеров K совпадает с числом интеграторов, а $X = X_{\mathcal{S}} \cup X_{\mathcal{F}} \cup X_{\mathcal{J}}$. Обозначим через

- S_i — базовая система $N_{x_i^{\mathcal{S}}}$, $i = 1, \dots, M$;
- F_j — дивайдер $N_{x_j^{\mathcal{F}}}$, $j = 1, \dots, K$;
- J_k — интегратор $N_{x_k^{\mathcal{J}}}$, $k = 1, \dots, K$.

Системы обслуживания, принадлежащие одному множеству, имеют сходное назначение и особенности.

1. $\mathcal{S} = \{N_{x_1^{\mathcal{S}}}, \dots, N_{x_M^{\mathcal{S}}}\} = \{S_1, \dots, S_M\}$ — множество базовых систем обслуживания — множество систем массового обслуживания с бесконечным числом одинаковых обслуживающих приборов. Длительность обслуживания фрагмента на любом приборе базовой системы S_i имеет экспоненциальное распределение с параметром μ_i , $i = 1, \dots, M$.
2. $\mathcal{F} = \{N_{x_1^{\mathcal{F}}}, \dots, N_{x_K^{\mathcal{F}}}\} = \{F_1, \dots, F_K\}$ — множество дивайдеров — множество одноприборных систем массового обслуживания. Процесс обслуживания в дивайдере F_k , $k = 1, \dots, K$, заключается в мгновенном делении поступающего фрагмента на фиксированное число $d_k \geq 2$ новых фрагментов. Каждый из d_k фрагментов мгновенно покидает дивайдер. Пример дивайдера изображён на рисунке 2.1.

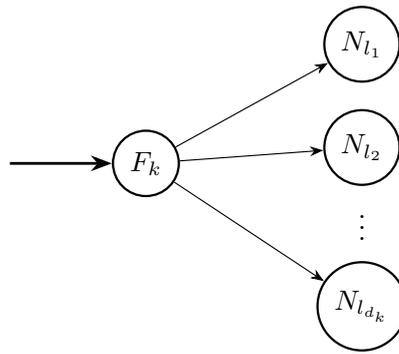


Рисунок 2.1 — Пример дивайдера

3. $\mathcal{J} = \{N_{x_1^{\mathcal{J}}}, \dots, N_{x_K^{\mathcal{J}}}\} = \{J_1, \dots, J_K\}$ — множество интеграторов — множество систем массового обслуживания с бесконечным числом приборов. Процесс обслуживания в интеграторе заключается в следующем. Каждый поступающий фрагмент занимает свободный прибор и ожидает момента поступления в интегратор последнего из всех фрагментов, принадлежащих до разделения тому же самому ранее единому фрагменту. Сразу после этого эти фрагменты мгновенно освобождают обслуживающие приборы, объединяются в исходный фрагмент, который покидает интегратор. Пример интегратора изображён на рисунке 2.2.

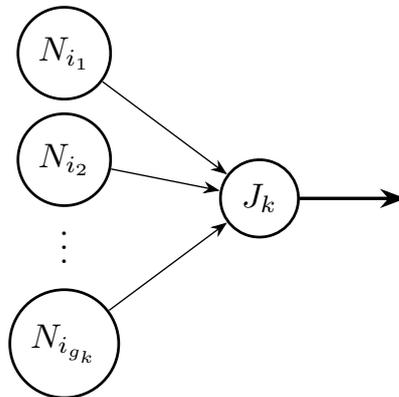


Рисунок 2.2 — Пример интегратора

2.1.2 Сигнатура фрагмента

Каждый из фрагментов в сети обслуживания характеризуется сигнатурой, которая идентифицирует фрагмент и отражает связи с другими фрагментами.

Определение 2. Пусть σ — некоторый произвольный фрагмент в сети обслуживания, который получен непосредственно в результате деления фрагмента σ^* в дивайдере F_k . Пусть i — номер фрагмента σ при делении, то есть $i \in \{1, \dots, d_k\}$. *Сигнатурой фрагмента σ* назовём тройку (σ^*, k, i) . В том случае, когда σ является поступающим из источника требованием, сигнатура имеет вид $(\sigma, 0, 1)$.

Если фрагмент σ имеет сигнатуру (σ^*, k, i) , то будем записывать это $\sigma \sim (\sigma^*, k, i)$.

Опишем формально процессы, происходящие с фрагментами в системах сети обслуживания.

Предположим, что фрагмент σ поступает в произвольный дивайдер F_k , $k \in \{1, \dots, K\}$, который делит фрагмент σ на d_k фрагментов $\{\sigma_1, \dots, \sigma_{d_k}\}$, $\sigma_i \sim (\sigma, k, i)$, $i = 1, \dots, d_k$. Полученные в результате деления фрагменты мгновенно покидают дивайдер и поступают в системы обслуживания. Базовые системы не меняют сигнатуру фрагмента, а каждый из полученных фрагментов может снова разделиться на новые фрагменты при поступлении в другой дивайдер. Для корректного объединения фрагментов $\{\sigma_1, \dots, \sigma_{d_k}\}$ они все должны в итоге поступить в некоторый интегратор. Маршрутизация фрагментов задаётся так, что все фрагменты σ_i , $i = 1, \dots, d_k$, полученные в дивайдере F_k , должны впоследствии поступить и быть объединены только в соответствующем интеграторе J_k . Переход данных фрагментов в интеграторы J_l , $l \neq k$, не допускается.

При поступлении фрагмента $\sigma_i \in \{\sigma_1, \dots, \sigma_{d_k}\}$ в интегратор J_k он находится в этом интеграторе до тех пор, пока не поступят все оставшиеся фрагменты из данного множества. Как только это произошло, фрагменты $\{\sigma_1, \dots, \sigma_{d_k}\}$ освобождают обслуживающие их приборы, мгновенно объединяются в исходный фрагмент σ , который сразу же покидает интегратор, переходя в другую систему обслуживания или источник.

Для удобства фрагмент σ , характеризуемый сигнатурой (σ^*, k, i) , будем называть k -фрагментом. Фрагмент может вернуться в источник требований N_0 только в случае, когда он является требованием (0-фрагментом).

2.1.3 Маршрутизация фрагментов

Для каждого фрагмента в сети обслуживания введём в рассмотрение вектор перемещений \mathbf{v} .

Определение 3. Пусть произвольный фрагмент σ характеризуется сигнатурой (σ_m, k_m, i_m) , и при этом $\sigma_n \sim (\sigma_{n-1}, k_{n-1}, i_{n-1})$, $n = m, m-1, \dots, 2$, $\sigma_1 \sim (\sigma_1, 0, 1)$. Вектор $\mathbf{v} = (k_0, k_1, \dots, k_m) = (0, k_1, \dots, k_m)$ назовём *вектором перемещений* для k_m -фрагмента σ , где k_m — *ведущий элемент* вектора перемещений.

Предполагается, что маршрутизация фрагментов такова, что вектор перемещений каждого фрагмента не может содержать одинаковых элементов, то есть $k_i \neq k_j$, $i \neq j$, $i, j = 1, \dots, m$. Это условие запрещает повторное деление фрагментов, полученных ранее в результате деления в дивайдере F_k , снова в этом дивайдере без их предварительного объединения в единый фрагмент в интеграторе J_k .

Все векторы перемещений тогда образуют *множество допустимых векторов перемещений* \mathcal{V} , которое является конечным множеством, зависящим от топологии сети обслуживания. Пусть $\mathbf{v} = (0, k_1, \dots, k_m)$, тогда через (\mathbf{v}, k_{m+1}) обозначим вектор $(0, k_1, \dots, k_m, k_{m+1})$.

Возможные переходы фрагментов в сети обслуживания \mathcal{N} заданы при помощи *набора матриц передач* $\Theta = (\Theta^{(0)}, \Theta^{(1)}, \dots, \Theta^{(K)})$, где $\Theta^{(k)} = (\theta_{i,j}^{(k)})$, $i, j = 0, 1, \dots, L$, $k = 0, 1, \dots, K$.

Элемент $\theta_{i,j}^{(k)}$ определяет

- вероятность перехода k -фрагмента из N_i в N_j , если $N_i \notin \mathcal{F}$;
- число k -фрагментов, полученных при делении фрагмента в дивайдере N_i , которые перейдут в N_j , если $N_i \in \mathcal{F}$.

Очевидно, что тогда

$$\theta_{x_i^{\mathcal{F}},j}^{(k)} = 0, \quad i = 1, \dots, K, i \neq k, j = 0, 1, \dots, L;$$

$$\sum_{j=0}^L \theta_{x_k^{\mathcal{F}},j}^{(k)} = d_k.$$

Предполагается, что набор матриц передач Θ обеспечивает корректную маршрутизацию фрагментов, то есть маршрутизация удовлетворяет условиям указанным выше.

2.2 Анализ сети обслуживания

2.2.1 Потоки в сети обслуживания

Обозначим через $\lambda_{in}^{(v)}(N_i)$, $\lambda_{out}^{(v)}(N_i)$ интенсивности входящего и выходящего потоков фрагментов с вектором перемещений $\mathbf{v} \in \mathcal{V}$ для системы обслуживания N_i , $i = 1, \dots, L$. Тогда суммарные интенсивности входящего $\Lambda_{in}(N_i)$ и выходящего $\Lambda_{out}(N_i)$ потоков фрагментов для систем обслуживания N_i , $i = 1, \dots, L$,

$$\Lambda_{in}(N_i) = \sum_{\mathbf{v} \in \mathcal{V}} \lambda_{in}^{(\mathbf{v})}(N_i), \quad (2.1)$$

$$\Lambda_{out}(N_i) = \sum_{\mathbf{v} \in \mathcal{V}} \lambda_{out}^{(\mathbf{v})}(N_i). \quad (2.2)$$

Пусть k — ведущий элемент вектора перемещений $\mathbf{v} \in \mathcal{V}$, в стационарном режиме функционирования сети обслуживания интенсивности входящих и выходящих потоков связаны следующими соотношениями:

$$\lambda_{in}^{(0)}(N_j) = \Lambda_0 \theta_{0,j}^{(0)} + \sum_{i \in X_S \cup X_{\mathcal{J}}} \lambda_{out}^{(0)}(N_i) \theta_{i,j}^{(0)}, \quad k = 0; \quad (2.3)$$

$$\lambda_{in}^{(v)}(N_j) = \sum_{i \in X_S \cup X_{\mathcal{J}}} \lambda_{out}^{(v)}(N_i) \theta_{i,j}^{(k)} + \frac{1}{d_k} \lambda_{out}^{(v)}(F_k) \theta_{x_k^{\mathcal{F}},j}^{(k)}, \quad k > 0; \quad (2.4)$$

где $j = 1, \dots, L$.

Справедливо, что

$$\lambda_{out}^{(v)}(S_i) = \lambda_{in}^{(v)}(S_i), \quad i = 1, \dots, M, \quad (2.5)$$

$$\lambda_{out}^{(\mathbf{v},k)}(F_k) = d_k \lambda_{in}^{(\mathbf{v})}(F_k), \quad \lambda_{out}^{(\mathbf{v})}(J_k) = \frac{1}{d_k} \lambda_{in}^{(\mathbf{v},k)}(J_k), \quad k = 1, \dots, K. \quad (2.6)$$

$$\lambda_{in}^{(\mathbf{v})}(F_k) = \lambda_{out}^{(\mathbf{v})}(J_k), \quad k = 1, \dots, K. \quad (2.7)$$

Интенсивности входящих и выходящих потоков фрагментов для всех $\mathbf{v} \in \mathcal{V}$ можно найти как решение уравнений (2.3), (2.4) с использованием равенств (2.5) и (2.6). Однако для вычисления интенсивностей потоков фрагментов удобнее использовать следующий алгоритм, который основан на равенстве (2.7).

Алгоритм 1. Вычисление интенсивностей потоков фрагментов

1. Из уравнений потоков для 0-фрагментов найти $\lambda_{in}^{(0)}(N_j)$, для всех $j \in X_S \cup X_{\mathcal{F}}$:

$$\lambda_{in}^{(0)}(N_j) = \Lambda_0 \theta_{0,j}^{(0)} + \sum_{i=1}^M \lambda_{in}^{(0)}(S_i) \theta_{x_i^S, j}^{(0)} + \sum_{i=1}^K \lambda_{in}^{(0)}(F_i) \theta_{x_i^{\mathcal{F}}, j}^{(0)}. \quad (2.8)$$

2. Для всех векторов перемещений $\mathbf{v} \in \mathcal{V}$, для которых найдены $\lambda_{in}^{(\mathbf{v})}(N_i)$, где $i \in X_S \cup X_{\mathcal{F}}$, выполнить шаг 3.

Затем перейти к шагу 4.

3. Для всех k таких, что $\lambda_{in}^{(\mathbf{v})}(F_k) > 0$, найти интенсивности потоков фрагментов $\lambda_{in}^{(\mathbf{v},k)}(N_j)$, где $j \in X_S \cup X_{\mathcal{F}}$, если они не были найдены ранее, используя уравнения:

$$\lambda_{in}^{(\mathbf{v},k)}(N_j) = \lambda_{in}^{(\mathbf{v})}(F_k) \theta_{x_k^{\mathcal{F}}, j}^{(k)} + \sum_{i=1}^M \lambda_{in}^{(\mathbf{v},k)}(S_i) \theta_{x_i^S, j}^{(k)} + \sum_{i=1}^K \lambda_{in}^{(\mathbf{v},k)}(F_i) \theta_{x_i^{\mathcal{F}}, j}^{(k)}. \quad (2.9)$$

Возврат к шагу 2.

4. Вычислить суммарные входящие интенсивности потоков фрагментов во все системы обслуживания

$$\Lambda_{in}(N_i) = \sum_{\mathbf{v} \in \mathcal{V}} \lambda_{in}^{(\mathbf{v})}(N_i), \quad i \in X_S \cup X_{\mathcal{F}}, \quad (2.10)$$

$$\Lambda_{in}(J_k) = d_k \Lambda_{in}(F_k), \quad k = 1, \dots, K. \quad (2.11)$$

Конец алгоритма.

Из формулы Литтла получим следующее утверждение.

Утверждение 1. Математическое ожидание $\mathbb{E}[x_i]$ числа фрагментов x_i в системе S_i определяется следующим выражением

$$\mathbb{E}[x_i] = \frac{\Lambda_{in}(S_i)}{\mu_i}, \quad i = 1, \dots, M. \quad (2.12)$$

2.2.2 Длительность пребывания требований в сети обслуживания

Определение 4. Множество $\mathcal{H}_k \subset \{N_1, \dots, N_L\}$, $k \in \{1, \dots, K\}$, таких систем обслуживания, в которые могут поступить фрагменты с вектором перемещений, содержащим k , будем называть *подсетью порождённой дивайдером F_k* . Положим дополнительно, что $F_k \in \mathcal{H}_k$.

Определение 5. Пусть сеть обслуживания \mathcal{N} функционирует в стационарном режиме. Обозначим через τ_k , $k = 1, \dots, K$, случайную величину, равную длительности интервала времени от момента разделения в дивайдере F_k произвольного фрагмента до момента объединения всех его частей в интеграторе J_k . Будем называть τ_k *длительностью реакции подсети \mathcal{H}_k* .

Определение 6. Матрицу $\Theta^{(k)}$ из набора матриц передач Θ назовём *элементарной матрицей*, если столбцы, соответствующие переходам во все дивайдеры, состоят из нулей, то есть

$$\theta_{i, x_j^{\mathcal{F}}}^{(k)} = 0, \quad i = 0, 1, \dots, L, \quad j = 1, \dots, K.$$

Конечность множества допустимых векторов перемещений \mathcal{V} влечёт следующее утверждение.

Утверждение 2. Среди матриц передачи $\{\Theta^{(1)}, \dots, \Theta^{(K)}\}$ существует, по крайней мере, одна элементарная матрица.

Определение 7. Пусть для некоторого $k \in \{1, \dots, K\}$, $\Theta^{(k)}$ — элементарная матрица, тогда будем называть подсеть \mathcal{H}_k *элементарной подсетью*.

Пусть \mathcal{H}_k , $k \in \{1, \dots, K\}$, есть элементарная подсеть. Через $\mathcal{A}(k)$ обозначим множество номеров базовых систем обслуживания, в которые переходят фрагменты непосредственно после деления в дивайдере F_k , то есть

$$\mathcal{A}(k) = \left\{ j \in \{1, \dots, M\} : \theta_{x_k^{\mathcal{F}}, x_j^{\mathcal{S}}}^{(k)} > 0 \right\}. \quad (2.13)$$

Пусть $\theta_{x_k^{\mathcal{F}}, x_i^{\mathcal{S}}}^{(k)} > 0$, тогда определим (k, i) -*ветвь подсети* как упорядоченную последовательность $\mathcal{B}(k, S_i)$ номеров всех базовых систем S_{b_j} , $j = 2, \dots, c_b$, в которые возможно поступление k -фрагментов из системы S_i ,

$$\mathcal{B}(k, S_i) = (b_1, b_2, \dots, b_{c_b}),$$

$$b_1 = i, \quad b_2 < b_3 < \dots < b_{c_b}.$$

Приведём некоторые сведения, связанные с фазовым распределением. Пусть поглощающая цепь Маркова $\{\nu = \nu(t), t \geq 0\}$ с непрерывным временем и пространством состояний $\{0, 1, 2, \dots, c_\nu\}$, где 0 — единственное поглощающее состояние, задана начальным распределением $(\alpha_0, \boldsymbol{\alpha})$ и генератором

$$\mathbf{Q} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{A}^* & \mathbf{A} \end{pmatrix}, \quad (2.14)$$

где $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{c_\nu})$ — вектор-строка размерности c_ν , $\mathbf{A} = (a_{i,j})$ — квадратная матрица порядка c_ν — субгенератор, определяющий переходы в множестве невозвратных состояний, $\mathbf{A}^* = (a_1^*, \dots, a_{c_\nu}^*)^T$ — вектор-столбец размерности c_ν , определяющий переходы в поглощающее состояние,

$$a_{i,i} < 0, \quad a_i^* \geq 0, \quad a_{i,j} \geq 0, \quad 1 \leq i \neq j \leq c_\nu,$$

$$\mathbf{A}^* = -\mathbf{A}\mathbf{1}, \quad \alpha_0 + \boldsymbol{\alpha}\mathbf{1} = 1,$$

где $\mathbf{1}$ есть вектор-столбец из единиц [112].

Определение 8. Пусть случайная величина (с.в.) ξ — время до поглощения цепи Маркова ν . Тогда говорят, что ξ имеет фазовое распределение с параметрами $(\boldsymbol{\alpha}, \mathbf{A})$ и c_ν фазами, и обозначают $\xi \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{A})$ [109; 112–114].

Функция распределения $G(t)$ и плотность распределения $g(t)$ с.в. $\xi \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{A})$ имеют вид

$$G(t) = 1 - \boldsymbol{\alpha} \exp(\mathbf{A}t)\mathbf{1}, \quad t \geq 0,$$

$$g(t) = \boldsymbol{\alpha} \exp(\mathbf{A}t) \mathbf{A}^*, \quad t > 0,$$

а n -ый начальный момент

$$\mathbb{E}[\xi^n] = (-1)^n n! \boldsymbol{\alpha} (\mathbf{A}^{-1})^n \mathbf{1}, \quad n \geq 1. \quad (2.15)$$

Одним из свойств случайных величин с фазовым распределением является их замкнутость относительно различных преобразований.

Пусть ξ есть непрерывная случайная величина с фазовым распределением и параметрами $(\boldsymbol{\alpha}, \mathbf{A})$, обозначим это как $\xi \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{A})$ [109; 112–114]. Известно [109], что если $\xi \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{A})$ и $\eta \sim \text{PH}(\boldsymbol{\beta}, \mathbf{B})$ суть независимые случайные величины, то случайная величина $\max\{\xi, \eta\}$ имеет фазовое распределение $\text{PH}(\boldsymbol{\gamma}, \mathbf{C})$, где $(\boldsymbol{\gamma}, \mathbf{C}) = (\boldsymbol{\alpha}, \mathbf{A}) \vee (\boldsymbol{\beta}, \mathbf{B})$, а операция “ \vee ” определена следующим образом

$$\boldsymbol{\gamma} = (\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, (1 - \boldsymbol{\beta}\mathbf{1})\boldsymbol{\alpha}, (1 - \boldsymbol{\alpha}\mathbf{1})\boldsymbol{\beta}), \quad (2.16)$$

$$\mathbf{C} = \begin{pmatrix} \mathbf{A} \oplus \mathbf{B} & \mathbf{I} \otimes (-\mathbf{B}\mathbf{1}) & (-\mathbf{A}\mathbf{1}) \otimes \mathbf{I} \\ \mathbf{0} & \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B} \end{pmatrix}, \quad (2.17)$$

здесь $\mathbf{1}$ — вектор-столбец из единиц, \oplus , \otimes — символы операций суммы и произведения Кронекера, \mathbf{I} — единичная матрица.

Теорема 1. *Длительность реакции τ_k элементарной подсети \mathcal{H}_k , $k \in \{1, \dots, K\}$, имеет фазовое распределение $\text{PH}(\boldsymbol{\alpha}^{(k)}, \mathbf{A}^{(k)})$, где*

$$(\boldsymbol{\alpha}^{(k)}, \mathbf{A}^{(k)}) = \bigvee_{j \in \mathcal{A}(k)} \left(\bigvee_{l=1}^{\theta_{x_k^{\mathcal{F}}, x_j^{\mathcal{S}}}^{(k)}} (\hat{\boldsymbol{\alpha}}^{(j)}, \hat{\mathbf{A}}^{(j)}) \right), \quad (2.18)$$

$$\begin{aligned}\hat{\boldsymbol{\alpha}}^{(j)} &= \left(\hat{\alpha}_n^{(j)} \right), \quad n = 1, \dots, c_{\mathbf{b}}, \\ \hat{\alpha}_1^{(j)} &= 1,\end{aligned}\tag{2.19}$$

$$\begin{aligned}\hat{\mathbf{A}}^{(j)} &= \left(\hat{a}_{n,m}^{(j)} \right), \quad n, m = 1, \dots, c_{\mathbf{b}}, \\ \hat{a}_{n,m}^{(j)} &= \mu_{b_n} \theta_{n^*, m^*}^{(k)}, \quad n \neq m,\end{aligned}\tag{2.20}$$

$$\hat{a}_{n,n}^{(j)} = -\mu_{b_n}.\tag{2.21}$$

Здесь $(b_1, \dots, b_{c_{\mathbf{b}}}) = \mathcal{B}(k, S_j)$,

$$n^* = x_{b_n}^{\mathcal{S}}, \quad m^* = x_{b_m}^{\mathcal{S}}.$$

Доказательство. Рассмотрим произвольный фрагмент σ , поступающий в дивайдер F_k , в котором происходит его деление на d_k фрагментов $\{\sigma_1, \dots, \sigma_{d_k}\}$. Фрагменты переходят по элементарной подсети обслуживания \mathcal{H}_k так, что поступают в интегратор J_k для объединения. В процессе перехода фрагментов от F_k до J_k они не делятся, так как \mathcal{H}_k является элементарной подсетью.

Пусть фрагмент σ_i , $i \in \{1, \dots, d_k\}$, переходит из дивайдера F_k непосредственно в базовую систему S_{q_i} , $q_i \in \mathcal{A}(k)$. Пусть $(b_1, \dots, b_{c_{\mathbf{b}}}) = \mathcal{B}(k, S_{q_i})$. Для удобства введём следующие обозначения:

$$n^* = x_{b_n}^{\mathcal{S}}, \quad m^* = x_{b_m}^{\mathcal{S}}.$$

Процесс перехода фрагмента σ_i , начиная от момента его поступления в базовую систему S_{q_i} и до момента его поступления в интегратор J_k , описывается поглощающей цепью Маркова с непрерывным временем [115]. Данная цепь Маркова имеет множество состояний $\{0, 1, \dots, c_{\mathbf{b}}\}$. Переход цепи в состояние $n \in \{1, \dots, c_{\mathbf{b}}\}$ соответствует поступлению фрагмента σ_i в базовую систему S_{b_n} . Длительность пребывания в состоянии n есть случайная величина с экспоненциальным распределением с параметром μ_{b_n} . Цепь переходит из состояния n в состояние $m \in \{1, \dots, c_{\mathbf{b}}\}$ с вероятностью $\theta_{n^*, m^*}^{(k)}$, которая является вероятностью

перехода фрагмента σ_i из базовой системы S_{b_n} в систему S_{b_m} после окончания обслуживания в базовой системе S_{b_n} .

Состояние цепи $m = 0$ является поглощающим состоянием. Переход цепи в это состояние отображает поступление фрагмента σ_i в интегратор J_k . Цепь переходит из состояния $n \in \{1, \dots, c_b\}$ в состояние $m = 0$ с вероятностью $\theta_{n^*, x_k^*}^{(k)}$.

Таким образом, процесс перехода фрагмента σ_i представлен в виде процесса эволюции цепи Маркова с начальным распределением $(\hat{\alpha}_0^{(q_i)}, \hat{\alpha}^{(q_i)})$, где $\hat{\alpha}_1^{(q_i)} = 1$.

Генератор $\hat{Q}^{(q_i)}$ цепи, описывающей переходы фрагмента σ_i , имеет вид

$$\hat{Q}^{(q_i)} = \begin{pmatrix} 0 & \mathbf{0} \\ -\hat{\mathbf{A}}^{(q_i)} \mathbf{1} & \hat{\mathbf{A}}^{(q_i)} \end{pmatrix}, \quad (2.22)$$

где

$$\hat{\mathbf{A}}^{(q_i)} = \left(\hat{a}_{n,m}^{(q_i)} \right), \quad n, m = 1, \dots, c_b, \quad (2.23)$$

$$\hat{a}_{n,m}^{(q_i)} = \mu_{b_n} \theta_{n^*, m^*}^{(k)}, \quad n \neq m,$$

$$\hat{a}_{n,n}^{(q_i)} = -\mu_{b_n}. \quad (2.24)$$

Обозначим через ξ_i случайную величину, равную длительности интервала времени до поглощения цепи Маркова, тогда, исходя из вероятностной интерпретации фазового распределения, $\xi_i \sim \text{PH}(\hat{\alpha}^{(q_i)}, \hat{\mathbf{A}}^{(q_i)})$.

Длительность реакции τ_k элементарной подсети \mathcal{H}_k тогда определяется как наибольшая из всех случайных величин ξ_i , $i = 1, \dots, d_k$. Таким образом, $\tau_k = \max \{ \xi_1, \dots, \xi_{d_k} \}$ имеет фазовое распределение с параметрами $(\alpha^{(k)}, \mathbf{A}^{(k)})$, где

$$(\alpha^{(k)}, \mathbf{A}^{(k)}) = \bigvee_{i=1}^{d_k} \left(\hat{\alpha}^{(q_i)}, \hat{\mathbf{A}}^{(q_i)} \right).$$

Упрощая данное выражение выделением одинаковых параметров фазовых распределений, приходим к утверждению теоремы. \square

Замечание 1. Аналогичным образом может быть получено распределение длительности пребывания требований в сети массового обслуживания, которая не

содержит дивайдеров и интеграторов. Поглощение в соответствующей модельной цепи Маркова отображает возвращение требования в источник требований. Начальное распределение определяется исходя из тех базовых систем, в которые требование может поступить непосредственно из источника.

Замечание 2. Элементарная подсеть \mathcal{H}_k может быть представлена как система типа $\cdot / \text{PH}(\boldsymbol{\alpha}^{(k)}, \mathbf{A}^{(k)}) / \infty$.

Теорема 1 позволяет рассмотреть процесс обслуживания фрагмента в элементарной подсети как эволюцию некоторой поглощающей цепи Маркова. Данную цепь Маркова можно рассматривать также как модель процесса обслуживания фрагмента в подсети обслуживания, которая не содержит дивайдеров и интеграторов (то есть все системы обслуживания — базовые системы).

Рассмотрим процедуру модификации исходной сети обслуживания \mathcal{N} , при которой элементарная подсеть \mathcal{H}_k замещается на другую подсеть, не содержащую дивайдеров и интеграторов, с таким же распределением длительности пребывания фрагментов в ней. Указанную процедуру будем называть *редукцией сети обслуживания \mathcal{N} относительно элементарной подсети \mathcal{H}_k* .

Пусть $\text{PH}(\boldsymbol{\alpha}^{(k)}, \mathbf{A}^{(k)})$ есть распределение длительности реакции для элементарной подсети \mathcal{H}_k с Y фазами,

$$\boldsymbol{\alpha}^{(k)} = \left(\alpha_i^{(k)} \right),$$

$$\mathbf{A}^{(k)} = \left(a_{i,j}^{(k)} \right), \quad i, j = 1, \dots, Y.$$

Интенсивности перехода a_i^* модельной цепи Маркова данного фазового распределения из состояния $i = 1, \dots, Y$, в поглощающее состояние имеют вид

$$a_i^* = - \sum_{j=1}^Y a_{i,j}^{(k)}.$$

В результате редукции будет получена новая сеть обслуживания $\tilde{\mathcal{N}}$, построенная по следующему алгоритму.

Обозначение параметров сети $\tilde{\mathcal{N}}$ отличается от аналогичных параметров сети \mathcal{N} наличием знака \sim над обозначением параметра.

Алгоритм 2. Редукция сети обслуживания относительно подсети \mathcal{H}_k

1. Положить параметры сети $\tilde{\mathcal{N}}$ равными параметрам сети \mathcal{N} .
2. Определить множество новых базовых систем $\mathcal{S}' = \{S'_1, \dots, S'_Y\}$. Интенсивность обслуживания μ'_i базовой системы S'_i , $i = 1, \dots, Y$, полагается равной

$$\mu'_i := -a_{i,i}^{(k)}. \quad (2.25)$$

3. Добавить базовые системы из множества \mathcal{S}' в множество базовых систем сети $\tilde{\mathcal{N}}$. В каждую из матриц $\tilde{\Theta}^{(l)}$, $l = 0, 1, \dots, K$, ввести Y новых строк и столбцов, отображающих переходы фрагментов для базовых систем из множества \mathcal{S}' , таким образом, что элементы строки и столбца с номером $L + i$ соответствуют вероятностям переходов для базовой системы S'_i , $i = 1, \dots, Y$.

Определить элементы матрицы передачи следующим образом:

$$\tilde{\theta}_{L+i,L+j}^{(l)} := \frac{a_{i,j}^{(k)}}{\mu'_i}, \quad i \neq j, \quad (2.26)$$

$$\tilde{\theta}_{L+i,L+i}^{(l)} := 0, \quad (2.27)$$

$$i, j = 1, \dots, Y, \quad l = 0, \dots, K.$$

4. Для $l = 0, \dots, K$, положить:

$$\tilde{\theta}_{i,L+j}^{(l)} := \alpha_j^{(k)} \theta_{i,x_k^{\mathcal{F}}}^{(l)}, \quad i = 0, \dots, L, \quad j = 1, \dots, Y, \quad (2.28)$$

$$\tilde{\theta}_{L+i,j}^{(l)} := \frac{a_i^*}{\mu_i} \theta_{x_k^{\mathcal{J}},j}^{(l)}, \quad i = 1, \dots, Y, \quad j = 0, \dots, L. \quad (2.29)$$

5. Исключить матрицу $\tilde{\Theta}^{(k)}$ из набора матриц передач $\tilde{\Theta}$.
6. Исключить из множества систем обслуживания сети $\tilde{\mathcal{N}}$ дивайдер F_k и интегратор J_k , а также соответствующие им строки и столбцы из набора матриц передач $\tilde{\Theta}$.
7. Все базовые системы $\{S_i : i \in X_{\mathcal{S}}\}$ сети \mathcal{N} , в которые поступал поток только k -фрагментов, исключить из сети $\tilde{\mathcal{N}}$. Из набора матриц передач $\tilde{\Theta}$ исключить соответствующие строки и столбцы.

Конец алгоритма.

Теорема 2. *Длительность пребывания требований в сети массового обслуживания \mathcal{N} с делением и слиянием требований имеет фазовое распределение.*

Доказательство. Пусть для некоторого $k \in \{1, \dots, K\}$, \mathcal{H}_k является элементарной подсетью в сети обслуживания \mathcal{N} . Выполним процедуру редукции относительно \mathcal{H}_k , в результате получим новую сеть массового обслуживания $\tilde{\mathcal{N}}_1$. Если в сети обслуживания $\tilde{\mathcal{N}}_1$ будет существовать элементарная подсеть, то выполним редукцию снова относительно найденной элементарной подсети. Таким образом, получим последовательность $\tilde{\mathcal{N}}_1, \tilde{\mathcal{N}}_2, \dots, \tilde{\mathcal{N}}_K$ сетей массового обслуживания. Сеть $\tilde{\mathcal{N}}_K$ не будет содержать дивайдеров и интеграторов, а распределение длительности пребывания требований в ней может быть получено исходя из следствия теоремы 1. \square

2.3 Пример анализа сети обслуживания с делением и слиянием требований

Рассмотрим сеть массового обслуживания с делением и слиянием требований, топология которой представлена ориентированным графом на рисунке 2.3, вершины которого обозначают системы обслуживания, а дуги соответствуют переходам фрагментов между системами сети.

Сеть обслуживания состоит из $L = 10$ систем, $\mathcal{S} = \{S_1, \dots, S_6\}$, $\mathcal{F} = \{F_1, F_2\}$, $\mathcal{J} = \{J_1, J_2\}$, N_0 — источник требований.

Дуга из N_i в N_j имеет метку $(\theta_{i,j}^{(0)}, \theta_{i,j}^{(1)}, \theta_{i,j}^{(2)})$. Например, дуга с меткой $(0; 0.5; 1)$, ведущая из S_2 в S_4 , означает, что 1-фрагменты переходят из S_2 в S_4 с вероятностью 0,5, для 2-фрагментов этот же переход осуществляется с вероятностью 1. Вероятность поступления требования из источника N_0 в F_1 равна 0,5. Требование, поступившее в F_1 , делится на два 1-фрагмента, один из которых поступает в S_1 , а другой — в S_2 .

Пусть $\Lambda_0 = 1$, $\mu_i = 2$, $i = 1, \dots, 6$.

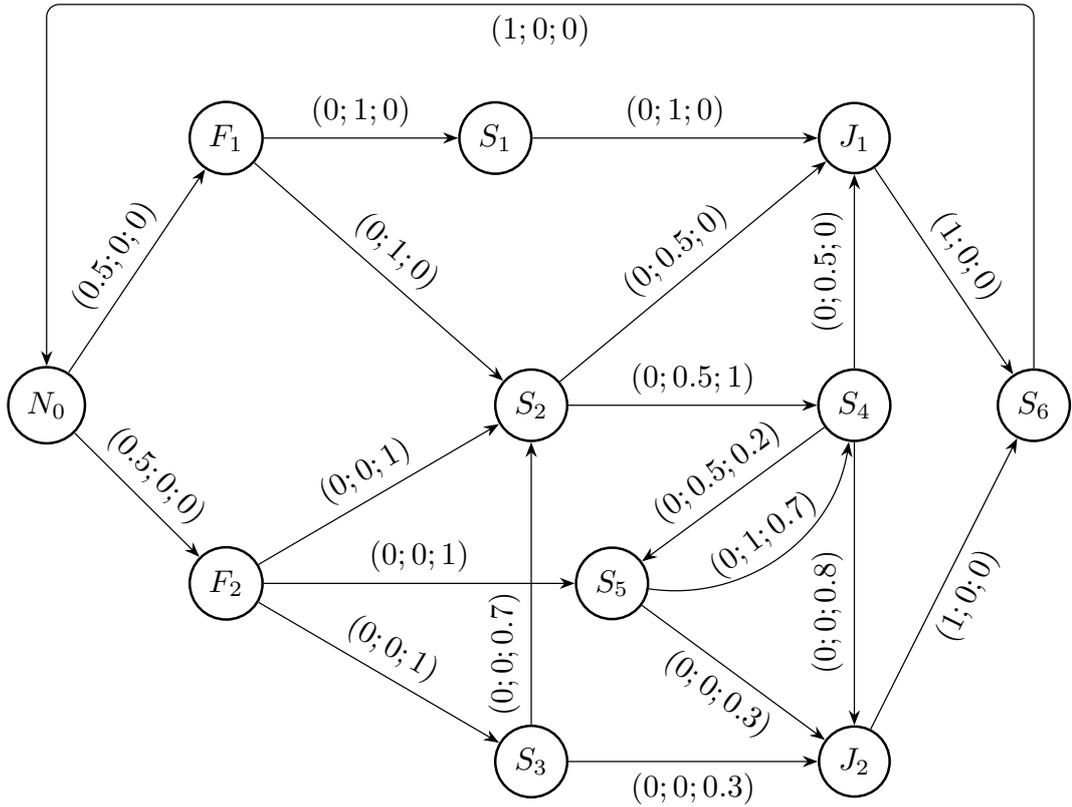


Рисунок 2.3 — Пример бескончноприборной сети обслуживания с делением и слиянием требований

Суммарные интенсивности входящих потоков фрагментов для систем обслуживания сети:

$$\begin{aligned} \Lambda_{in}(S_1) &= 0.5, & \Lambda_{in}(S_2) &= 1.35, \\ \Lambda_{in}(S_3) &= 0.5, & \Lambda_{in}(S_4) &= 1.8953, \\ \Lambda_{in}(S_5) &= 1.0291, & \Lambda_{in}(S_6) &= 1, \\ \Lambda_{in}(F_1) &= 0.5, & \Lambda_{in}(F_2) &= 0.5, \\ \Lambda_{in}(J_1) &= 1, & \Lambda_{in}(J_2) &= 1.5. \end{aligned}$$

Рассмотрим распределение длительности реакции τ_1 для элементарной подсети $\mathcal{H}_1 = \{F_1, S_1, S_2, S_4, S_5, J_1\}$.

Из теоремы 1 получаем:

$$\mathcal{A}(1) = \{1, 2\},$$

$$\mathcal{B}(1, S_1) = (1), \quad \mathcal{B}(1, S_2) = (2, 4, 5);$$

параметры фазового распределения

$$\hat{\alpha}^{(1)} = (1), \quad \hat{A}^{(1)} = (-2);$$

$$\hat{\boldsymbol{\alpha}}^{(2)} = (1, 0, 0), \quad \hat{\mathbf{A}}^{(2)} = \begin{pmatrix} -2 & 1 & 0 \\ 0 & -2 & 1 \\ 0 & 2 & -2 \end{pmatrix};$$

тогда параметры распределения длительности реакции подсети \mathcal{H}_1

$$\boldsymbol{\alpha}^{(1)} = (1, 0, 0, 0, 0, 0, 0),$$

$$\mathbf{A}^{(1)} = \begin{pmatrix} -4 & 1 & 0 & 1 & 2 & 0 & 0 \\ 0 & -4 & 1 & 1 & 0 & 2 & 0 \\ 0 & 2 & -4 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{pmatrix}.$$

После выполнения редукций получаем, что математическое ожидание $\mathbf{E}[\tau]$ длительности пребывания требований в сети обслуживания \mathcal{N} и дисперсия $\mathbf{Var}[\tau]$ длительности пребывания требований равны

$$\mathbf{E}[\tau] = 2.2027, \quad \mathbf{Var}[\tau] = 1.8736.$$

2.4 Стационарные характеристики для элементарной сети с делением и слиянием требований

Пусть \mathcal{N} — сеть массового обслуживания с делением и слиянием требований, состоящая из дивайдера F , интегратора J , базовых систем S_1, \dots, S_M . Предполагается, что каждое требование из источника непосредственно поступает в дивайдер F , где делится на d фрагментов, объединение фрагментов происходит затем в интеграторе J , а полученное требование возвращается в источник. Структура сети изображена на рисунке 2.4.

Будем называть такую сеть *элементарной* сетью обслуживания. Рассмотрим задачу нахождения стационарного распределения такой сети обслуживания.

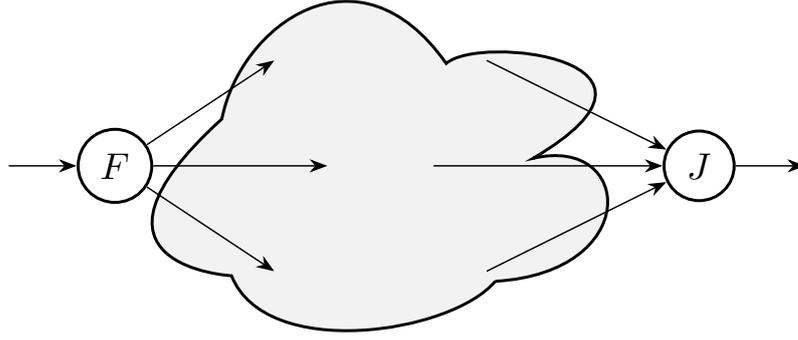


Рисунок 2.4 — Структура элементарной сети с делением и слиянием требований

Состояние сети \mathcal{N} определяется вектором $\mathbf{x} = (x_1, \dots, x_M)$, где x_i — число фрагментов в системе S_i .

Очевидно, что процесс $\{\mathbf{x}(t), t > 0\}$ является цепью Маркова с непрерывным временем и счётным множеством состояний, стационарное распределение которого может быть найдено из уравнений глобального равновесия. Далее рассмотрим новый метод определения стационарного распределения вероятностей состояний элементарной сети обслуживания.

Определение 9. Пусть фрагмент σ делится в дивайдере F на фрагменты $\{\sigma_1, \dots, \sigma_d\}$, тогда под *способом размещения* требования в сети обслуживания будем понимать мультимножество \mathbf{l} , состоящее из d элементов и определяемое следующим образом

$$\mathbf{l} = \{N(\sigma_1), \dots, N(\sigma_d)\}, \quad (2.30)$$

где $N(\sigma_i)$ обозначает систему, в которой находится фрагмент σ_i .

Процесс $\{\mathbf{l}(t), t > 0\}$, описывающий переход фрагментов некоторого требования от момента деления в дивайдере до момента объединения в интеграторе, является поглощающей цепью Маркова на множестве $\mathbf{L} = \{\mathbf{l}_0, \dots, \mathbf{l}_X\}$ всех размещений с поглощающим состоянием $\{J, \dots, J\}$. Для удобства положим, что \mathbf{l}_1 является начальным состоянием, определяемым мультимножеством тех базо-

вых систем $\mathcal{A}(F)$, в которые фрагменты требования поступают непосредственно из дивайдера, а $\mathbf{l}_0 = \{J, \dots, J\}$ — поглощающим.

В качестве примера рассмотрим сеть обслуживания, состоящую из параллельных систем обслуживания, где каждое требование при поступлении делится на M фрагментов.

Цепь Маркова $\{\mathbf{l}(t), t > 0\}$, описывающая эволюцию рассматриваемой сети, начинает свою эволюцию в состоянии $\{S_1, \dots, S_M\}$, интенсивность $a(\mathbf{l}, \mathbf{l}')$ перехода цепи из состояния \mathbf{l} в состояние \mathbf{l}' имеет вид

$$a(\{N_{i_1}, \dots, S_{i_j}, \dots, N_{i_d}\}, \{N_{i_1}, \dots, J, \dots, N_{i_d}\}) = \mu_{i_j}.$$

В том случае, когда число базовых систем $M = 3$, диаграмма состояний цепи Маркова $\{\mathbf{l}(t), t > 0\}$ изображена на рисунке 2.5.

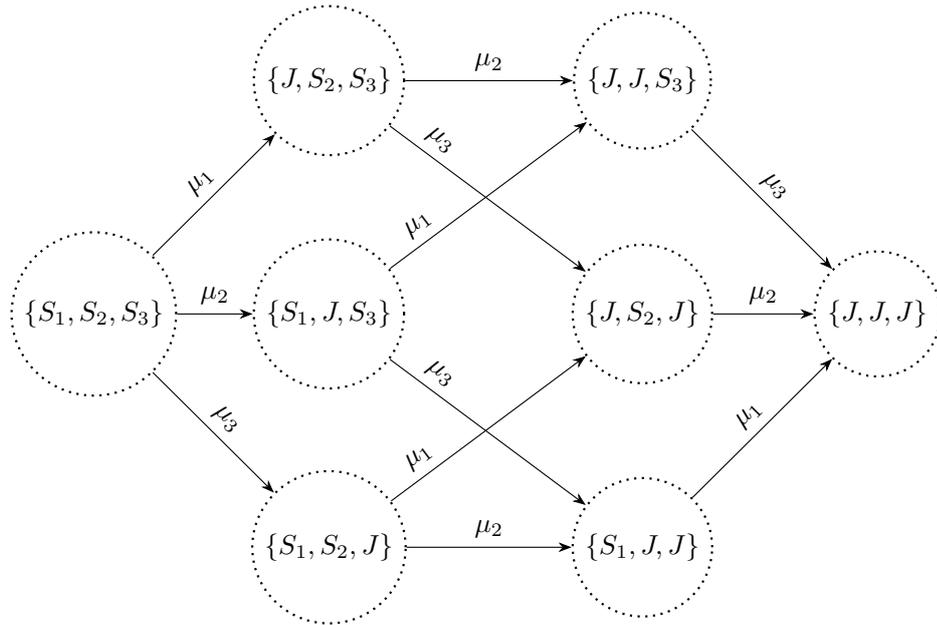


Рисунок 2.5 — Диаграмма состояний цепи Маркова для $M = 3$.

Поглощающей цепи Маркова можно поставить в соответствие бесконечно-приборную сеть Джексона таким образом, что процесс эволюции цепи может быть рассмотрен как процесс перехода некоторого требования от момента его поступления из источника и до момента возвращения его в источник. В силу того, что процессы перехода фрагментов различных требований по сети \mathcal{N} независимы в совокупности, каждому требованию в сети \mathcal{N} (процессу изменения

способа размещения \mathbf{l} для этого требования) можно поставить в соответствие некоторое требование в введённой сети Джексона.

Определение 10. Сопоставим цепи Маркова $\{\mathbf{l}(t), t > 0\}$ сеть Джексона $\mathcal{N}^{\mathbf{L}}$ с бесконечным числом обслуживающих приборов в системах обслуживания так, что источнику требований будет соответствовать поглощающее состояние \mathbf{l}_0 , а каждой системе в сети $\mathcal{N}^{\mathbf{L}}$ поставим во взаимно-однозначное соответствие некоторое непоглощающее состояние из \mathbf{L} . Будем называть сеть $\mathcal{N}^{\mathbf{L}}$ *сетью размещения*. Далее будем использовать элементы из \mathbf{L} для обозначения соответствующих им систем обслуживания или источника в сети $\mathcal{N}^{\mathbf{L}}$.

Сеть обслуживания $\mathcal{N}^{\mathbf{L}}$ содержит X систем. Обозначим через y_j число требований в системе обслуживания сети $\mathcal{N}^{\mathbf{L}}$, соответствующей способу размещения \mathbf{l}_j . Пусть сеть $\mathcal{N}^{\mathbf{L}}$ находится в состоянии $\mathbf{y} = (y_1, \dots, y_X)$, тогда число фрагментов x_i в системе S_i сети обслуживания \mathcal{N}

$$x_i = \sum_{j=1}^X y_j \varphi(\mathbf{l}_j, S_i), \quad i = 1, \dots, M, \quad (2.31)$$

где $\varphi(\mathbf{l}_j, S_i)$ обозначает число вхождений элемента S_i в мультимножество \mathbf{l}_j .

Известно, что для сети $\mathcal{N}^{\mathbf{L}}$, стационарное распределение вероятностей состояний имеет мультипликативную форму [81],

$$\pi(y_1, \dots, y_X) = \prod_{i=1}^X \pi_i(y_i), \quad (2.32)$$

где $\pi_i(y_i)$ — стационарная вероятность пребывания y_i требований в системе с номером i . Для системы $M/GI/\infty$ стационарная вероятность $\pi(q)$ нахождения в ней q требований равна

$$\pi(q) = \frac{(\lambda/\mu)^q e^{-\lambda/\mu}}{q!}, \quad q \geq 0, \quad (2.33)$$

здесь λ задаёт интенсивность входящего потока требований, μ — интенсивность обслуживания требований одним прибором.

Поскольку способ размещения требования точно определяет размещение всех его фрагментов по системам сети \mathcal{N} , стационарное распределение сети $\mathcal{N}^{\mathbf{L}}$ позволяет найти стационарное распределение для сети \mathcal{N} .

Рассмотрим стационарную вероятность $p_i(x_i)$ того, что в системе S_i сети \mathcal{N} обслуживаются x_i фрагментов. Фрагменты каждого требования размещены по сети в соответствии с некоторым способом из множества \mathbf{L} . Если \mathbf{l} есть способ размещения некоторого требования, то число фрагментов этого требования в системе S_i будет равно $\varphi(\mathbf{l}, S_i)$. Тогда для стационарной вероятности $p_i(x_i)$ справедливо

$$p_i(x_i) = \sum_{u_i(\mathbf{y})=x_i} \pi(y_1, \dots, y_X) = \sum_{u_i(\mathbf{y})=x_i} \pi_1(y_1) \dots \pi_X(y_X), \quad (2.34)$$

где $u_i(\mathbf{y}) = \sum_{j=1}^X y_j \varphi(\mathbf{l}_j, S_i)$.

Элементы множества $\mathbf{L}_i \subset \mathbf{L}$ будем называть значимыми способами размещения для системы S_i , если для любого $\mathbf{l} \in \mathbf{L}_i$, выполняется $\varphi(\mathbf{l}, S_i) > 0$. Тогда выражение для $p_i(x_i)$ может быть записано в виде суммы с конечным числом слагаемых, таким образом, получим утверждение следующей теоремы.

Теорема 3. *Для сети \mathcal{N} стационарная вероятность $p_i(x_i)$ того, что система S_i содержит x_i фрагментов*

$$p_i(x_i) = \sum_{u_i((y_{r_1}, \dots, y_{r_m}))=x_i} \pi_{r_1}(y_{r_1}) \dots \pi_{r_m}(y_{r_m}), \quad (2.35)$$

где $r_j, j = 1, \dots, m$, есть множество номеров значимых состояний для системы S_i .

Доказательство. Является следствием процесса построения сети размещений. □

Из указанных выше выражений и теоремы приходим непосредственно к следующему утверждению.

Следствие 1. Для сети \mathcal{N} стационарная вероятность $p(x_1, \dots, x_M)$ определяется как

$$p(x_1, \dots, x_M) = \sum_{\substack{u_1(\mathbf{y})=x_1, \\ \dots \\ u_M(\mathbf{y})=x_M}} \pi_1(y_1) \dots \pi_X(y_X). \quad (2.36)$$

Таким образом задача нахождения стационарного распределения элементарной сети с делением и слиянием требований сводится к определению соответствующей сети размещений.

Для построения сети размещений требуется определить все способы размещения фрагментов некоторого требования по сети обслуживания. Для этого необходимо рассмотреть все размещения начиная с начального состояния $\mathbf{l}_1 = \mathcal{A}(F)$ соответствующей цепи Маркова. На каждом шаге определяются смежные размещения для всех полученных ранее размещений, вычисляются интенсивности переходов в смежные размещения. Таким образом происходит перебор всех допустимых размещений.

Рассмотрим здесь алгоритм построения сети размещений для произвольной элементарной сети обслуживания. Обозначим через $\theta_{i,j}$ элемент маршрутной матрицы, соответствующий вероятности перехода фрагментов из системы N_i в систему N_j без указания типа фрагмента.

Для сети размещений \mathcal{N}^L введём следующие обозначения:

- $\theta^L(\mathbf{l}_i, \mathbf{l}_j)$ — вероятность перехода требований из системы обслуживания \mathbf{l}_i в систему \mathbf{l}_j ,
- $\mu^L(\mathbf{l})$ — интенсивность обслуживания требований при пребывании в системе обслуживания \mathbf{l}

Алгоритм 3. Построение сети размещений \mathcal{N}^L

1. Положить:

$$\mathbf{l}_0 := \{J, \dots, J\},$$

$$\mathbf{L} := \{\mathbf{l}_0\},$$

$$\mathbf{L}_{next} := \{\mathcal{A}(F)\},$$

$$\theta^{\mathbf{L}}(\mathbf{l}_0, \mathcal{A}(F)) := 1.$$

2. Пока \mathbf{L}_{next} не пусто, для всех элементов $\mathbf{l} \in \mathbf{L}_{next}$ выполнить:

а) определить интенсивность обслуживания $\mu^{\mathbf{L}}(\mathbf{l})$ для \mathbf{l}

$$\mu^{\mathbf{L}}(\mathbf{l}) := \sum_{i=1}^M \mu_i \varphi(\mathbf{l}, S_i),$$

б) определить все смежные системы обслуживания и интенсивности переходов в них для системы \mathbf{l} , для этого выполнить для всех $S_i \in \mathbf{l}$:

i. сформировать размещение \mathbf{l}' путём одной замены $S_i \in \mathbf{l}$ на N_j в том случае, когда

$$\theta_{x_i^S, j} > 0,$$

ii. определить

$$\theta^{\mathbf{L}}(\mathbf{l}, \mathbf{l}') := \theta^{\mathbf{L}}(\mathbf{l}, \mathbf{l}') + \theta_{x_i^S, j} \frac{\mu_i}{\mu^{\mathbf{L}}(\mathbf{l})}.$$

iii. добавить \mathbf{l}' во множество \mathbf{L}_{next} , если $\mathbf{l}' \notin \mathbf{L} \cup \mathbf{L}_{next}$ и $\mathbf{l}' \neq \{J, \dots, J\}$.

с) удалить \mathbf{l} из множества \mathbf{L}_{next} .

3. Определить интенсивность источника \mathbf{l}_0 равной интенсивности входящего потока в исходную сеть обслуживания \mathcal{N} .

Конец алгоритма.

Примеры нахождения стационарного распределения с использованием описанного алгоритма представлены в главе 4.

Глава 3. Сети обслуживания, зависящие от нагрузки

В первом разделе главы рассматриваются открытые сети массового обслуживания произвольной топологии с делением и слиянием требований, в которых все базовые системы являются бесконечноприборными. Предполагается, что интенсивность обслуживания фрагментов каждым прибором базовой системы является экспоненциально распределённой случайной величиной, зависящей от интенсивности потока фрагментов в эту систему. Показано, что сеть обслуживания данного класса является хорошим приближением сети массового обслуживания, в которой все базовые системы содержат по одному прибору. Для анализа этих сетей обслуживания может быть использован метод, предложенный в предыдущей главе. Во втором разделе главы рассматривается сеть массового обслуживания с делением и слиянием требований, в которой все подсети порождаемые дивайдерами являются элементарными, а требования из источника поступают непосредственно в дивайдеры и после объединения соответствующих фрагментов в интеграторах сразу возвращаются в источник. Вводится понятие веса фрагмента, определяющего долю фрагмента в целой части требования. В базовых системах устанавливается линейная зависимость интенсивности обслуживания фрагментов от значений интенсивности поступающих фрагментов и их веса. Обсуждается задача нахождения для каждого дивайдера сети обслуживания такого распределения весов, которое минимизирует математическое ожидание длительности пребывания требований в сети обслуживания. В третьем разделе главы демонстрируется применение сети массового обслуживания с делением и слиянием требований в качестве модели сети передачи данных с многопутевой маршрутизацией.

3.1 Метод анализа сети обслуживания, состоящей из одноприборных базовых систем

Рассмотрим открытую бесконечноприборную сеть обслуживания \mathcal{N} с делением и слиянием требований, описанную в предыдущей главе, в которой ин-

тенсивность обслуживания фрагментов одним прибором базовой системы сети *зависит от интенсивности поступающего в эту систему потока* [116].

Пусть для базовых систем известны интенсивности входящего потока $\Lambda_{in}(S_i)$, $i = 1, \dots, M$. В таком случае интенсивность обслуживания $\hat{\mu}_i$ фрагментов прибором базовой системы S_i определяется следующей зависимостью

$$\hat{\mu}_i = \mu_i - \Lambda_{in}(S_i), \quad i = 1, \dots, M, \quad (3.1)$$

где μ_i есть заданные константы. Длительность пребывания фрагментов в базовой системе S_i имеет экспоненциальное распределение с параметром $\hat{\mu}_i$.

Для анализа представленной сети обслуживания \mathcal{N} можно использовать методы, изложенные в предыдущей главе.

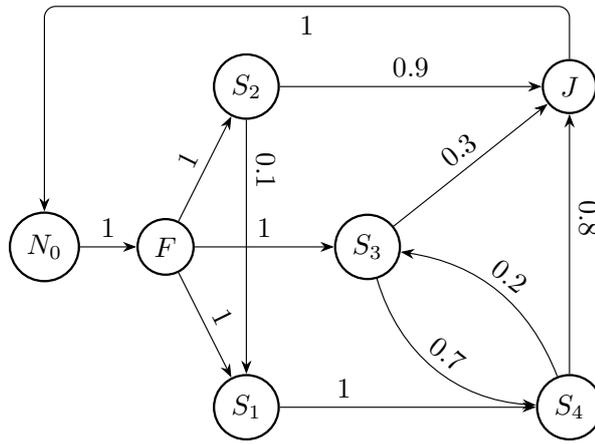
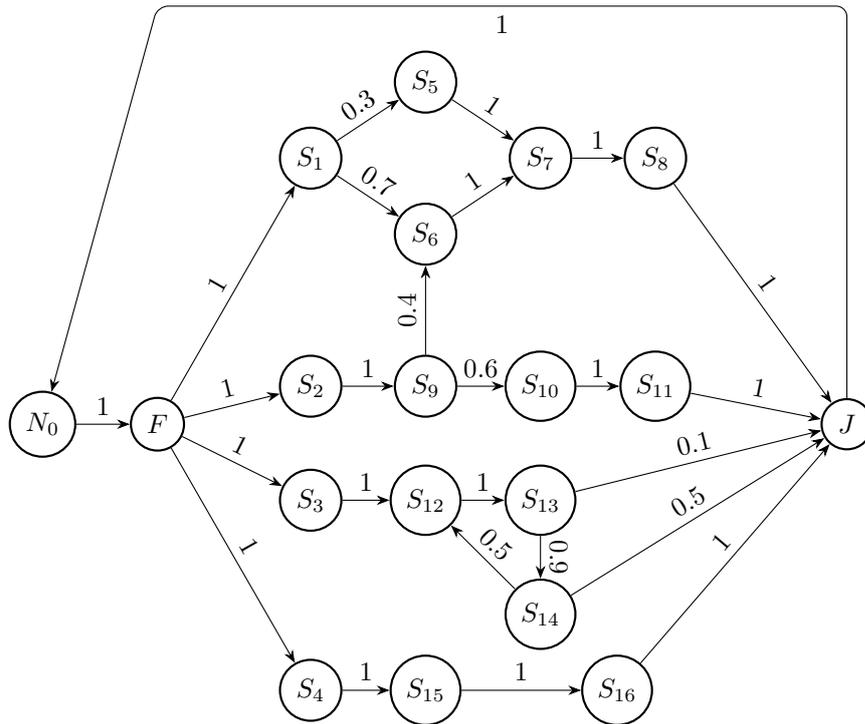
Рассмотрим теперь сеть обслуживания \mathcal{N}^f , топология и маршрутизация в которой совпадают с сетью \mathcal{N} , кроме того, что *все базовые системы являются одноприборными*, а интенсивность обслуживания фрагментов прибором базовой системы S_i не зависит от интенсивности поступающего в неё потока и равна μ_i .

Отметим, что длительность пребывания требований в системе массового обслуживания типа $M/M/1$ с интенсивностью входящего потока λ и интенсивностью обслуживания μ есть экспоненциально распределённая случайная величина с параметром $(\mu - \lambda)$ [104].

Задачу анализа сети обслуживания \mathcal{N}^f с одноприборными базовыми системами можно свести к известной задаче анализа зависящей от нагрузки сети обслуживания \mathcal{N} с бесконечноприборными базовыми системами. Данный метод анализа является приближённым. Рассмотрим применение этого подхода и его точность на примере двух сетей обслуживания \mathcal{N}_1 и \mathcal{N}_2 , изображённых на рисунках 3.1 и 3.2, со следующими интенсивностями обслуживания:

- для сети \mathcal{N}_1 , $\boldsymbol{\mu} = (2, 2, 2, 3)$;
- для сети \mathcal{N}_2 , $\boldsymbol{\mu} = (8, 6, 6, 6, 7, 8, 8, 8, 8, 8, 8, 8, 8, 6, 6, 6)$.

Метки на соответствующих дугах задают вероятности перехода между системами сети обслуживания или число фрагментов, поступающих в базовые системы из дивайдера F .

Рисунок 3.1 — Сеть обслуживания \mathcal{N}_1 Рисунок 3.2 — Сеть обслуживания \mathcal{N}_2

Для бесконечноприборных сетей обслуживания \mathcal{N}_1 и \mathcal{N}_2 точные значения $E[\tau(\mathcal{N}_1)]$ и $E[\tau(\mathcal{N}_2)]$ для математического ожидания длительности пребывания требований в сети обслуживания соответственно вычислены на основе результатов главы 2, оценки $E[\tau^{sim}(\mathcal{N}_1^f)]$ и $E[\tau^{sim}(\mathcal{N}_2^f)]$ для математического ожидания длительности пребывания в сетях обслуживания \mathcal{N}_1^f и \mathcal{N}_2^f получены с использованием разработанной программы имитационного моделирования сетей обслуживания с делением и слиянием требований, описание которой представлено в главе 4.

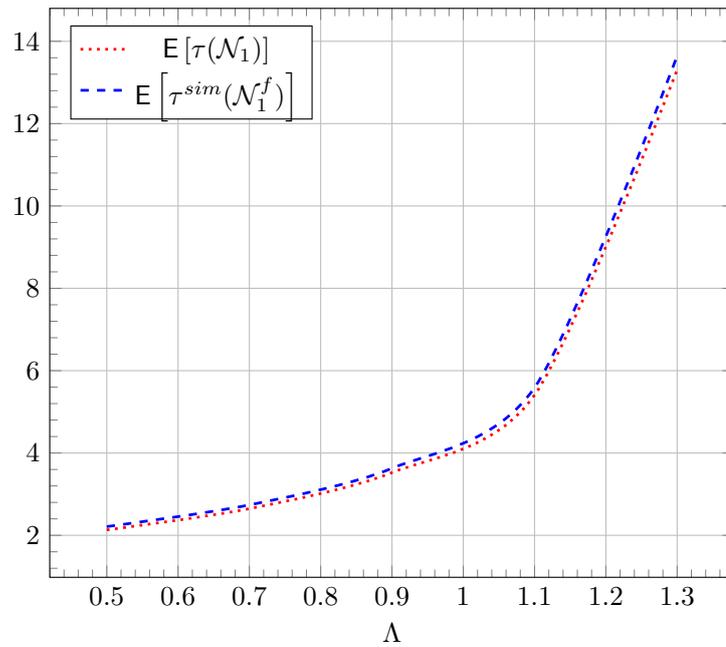


Рисунок 3.3 — Сравнение результатов для сетей \mathcal{N}_1 и \mathcal{N}_1^f

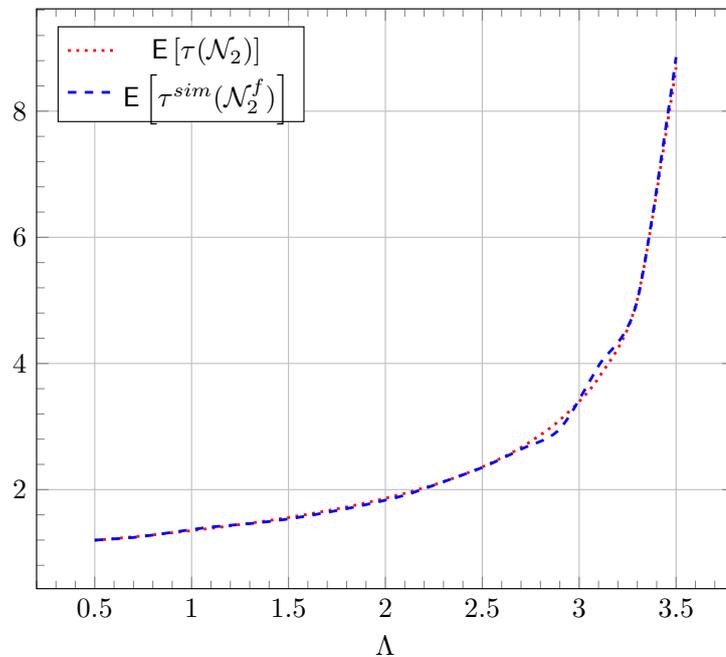


Рисунок 3.4 — Сравнение результатов для сетей \mathcal{N}_2 и \mathcal{N}_2^f

На рисунке 3.3 изображены графики зависимости математического ожидания длительности пребывания требований в сети обслуживания \mathcal{N}_1 и его оценки для сети \mathcal{N}_1^f в зависимости от интенсивности Λ поступающего в сеть потока. Аналогичные результаты для сетей обслуживания \mathcal{N}_2 и \mathcal{N}_2^f изображены на рисунке 3.4.

Более подробные численные результаты представлены в таблицах 3.1 и 3.2, здесь также приведены значения коэффициента использования ρ каждой

из базовых систем для сетей обслуживания \mathcal{N}_1^f и \mathcal{N}_2^f ,

$$\rho = \frac{\lambda}{\mu},$$

где λ и μ есть интенсивности поступающего потока в базовую систему и интенсивность обслуживания фрагментов прибором базовой системы обслуживания соответственно.

Полученное таким образом приближение для математического ожидания длительности пребывания требований в сети обслуживания является допустимым для реальных технических задач при большом выборе значений коэффициентов использования в базовых системах сети обслуживания, что отражено в таблицах 3.1 и 3.2.

Максимальная относительная погрешность при этом составила не более 4% для экспериментов с сетью \mathcal{N}_1^f , и не более 5% для сети \mathcal{N}_2^f . Средние относительные погрешности составили 3.13% и 1.49% соответственно.

Подробное описание экспериментов содержится в разделе 4.5.

Таблица 3.1 — Таблица с результатами для сетей \mathcal{N}_1 и \mathcal{N}_1^f

Λ	$E[\tau(\mathcal{N}_1)]$	$E[\tau^{sim}(\mathcal{N}_1^f)]$	ρ_1	ρ_2	ρ_3	ρ_4
0.5	2.13	2.21	0.28	0.25	0.35	0.35
0.7	2.65	2.74	0.39	0.35	0.5	0.49
0.9	3.52	3.63	0.5	0.45	0.64	0.63
1.1	5.41	5.59	0.61	0.55	0.78	0.77
1.3	13.31	13.65	0.72	0.65	0.92	0.91

Таблица 3.2 — Таблица с результатами для сетей \mathcal{N}_2 и \mathcal{N}_2^f

Λ	$E[\tau(\mathcal{N}_2)]$	$E[\tau^{sim}(\mathcal{N}_2^f)]$	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	ρ_{10}	ρ_{11}	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{15}	ρ_{16}
0.5	1.2	1.2	0.06	0.08	0.08	0.08	0.02	0.07	0.07	0.07	0.06	0.04	0.04	0.11	0.11	0.14	0.08	0.08
0.7	1.25	1.24	0.09	0.12	0.12	0.12	0.03	0.1	0.1	0.1	0.09	0.05	0.05	0.16	0.16	0.19	0.12	0.12
0.9	1.32	1.32	0.11	0.15	0.15	0.15	0.04	0.12	0.12	0.12	0.11	0.07	0.07	0.2	0.2	0.25	0.15	0.15
1.1	1.39	1.41	0.14	0.18	0.18	0.18	0.05	0.15	0.15	0.15	0.14	0.08	0.08	0.25	0.25	0.3	0.18	0.18
1.3	1.47	1.46	0.16	0.22	0.22	0.22	0.06	0.18	0.18	0.18	0.16	0.1	0.1	0.3	0.3	0.35	0.22	0.22
1.5	1.56	1.53	0.19	0.25	0.25	0.25	0.06	0.21	0.21	0.21	0.19	0.11	0.11	0.34	0.34	0.41	0.25	0.25
1.7	1.67	1.64	0.21	0.28	0.28	0.28	0.07	0.23	0.23	0.23	0.21	0.13	0.13	0.39	0.39	0.46	0.28	0.28
1.9	1.79	1.76	0.24	0.32	0.32	0.32	0.08	0.26	0.26	0.26	0.24	0.14	0.14	0.43	0.43	0.52	0.32	0.32
2.1	1.94	1.91	0.26	0.35	0.35	0.35	0.09	0.29	0.29	0.29	0.26	0.16	0.16	0.48	0.48	0.57	0.35	0.35
2.3	2.13	2.12	0.29	0.38	0.38	0.38	0.1	0.32	0.32	0.32	0.29	0.17	0.17	0.52	0.52	0.63	0.38	0.38
2.5	2.36	2.35	0.31	0.42	0.42	0.42	0.11	0.34	0.34	0.34	0.31	0.19	0.19	0.57	0.57	0.68	0.42	0.42
2.7	2.67	2.64	0.34	0.45	0.45	0.45	0.12	0.37	0.37	0.37	0.34	0.2	0.2	0.61	0.61	0.74	0.45	0.45
2.9	3.1	2.96	0.36	0.48	0.48	0.48	0.12	0.4	0.4	0.4	0.36	0.22	0.22	0.66	0.66	0.79	0.48	0.48
3.1	3.77	3.96	0.39	0.52	0.52	0.52	0.13	0.43	0.43	0.43	0.39	0.23	0.23	0.7	0.7	0.85	0.52	0.52
3.3	5.02	5	0.41	0.55	0.55	0.55	0.14	0.45	0.45	0.45	0.41	0.25	0.25	0.75	0.75	0.9	0.55	0.55
3.5	8.69	8.86	0.44	0.58	0.58	0.58	0.15	0.48	0.48	0.48	0.44	0.26	0.26	0.8	0.8	0.95	0.58	0.58

3.2 Задача оптимизации сети обслуживания с делением и слиянием требований

Рассмотрим сеть массового обслуживания с делением и слиянием требований, в которой все подсети порождаемые дивайдерами являются элементарными, а требования из источника поступают непосредственно в дивайдеры и после объединения соответствующих фрагментов в интеграторах сразу возвращаются в источник. Будем называть такие сети обслуживания *квазиэлементарными* сетями обслуживания.

Относительно рассматриваемой квазиэлементарной сети сделаем следующие предположения. Пусть каждое поступающее в сеть обслуживания требование имеет *вес* $w = 1$. Требование, поступившее на дивайдер F_k , $k = 1, \dots, K$, мгновенно делится на фрагменты, которые распределяются по базовым системам в соответствии с распределением весов \mathcal{W} ,

$$\mathcal{W} = \{w_k(l) : k = 1, \dots, K, l \in \mathcal{A}(k)\}.$$

Фрагмент с весом $w_k(l) \in (0, 1]$ поступает в базовую систему S_l , $l \in \mathcal{A}(k)$, где множество $\mathcal{A}(k)$ определяется выражением (2.13).

Потребуем, чтобы выполнялся *закон сохранения веса*

$$\sum_{l \in \mathcal{A}(k)} w_k(l) = 1, \quad k = 1, \dots, K.$$

Фрагменты, полученные при делении в дивайдере F_k , будем так же называть k -фрагментами, если же дополнительно известно, что k -фрагмент поступил из F_k *непосредственно* в S_l , то будем называть его (k, l) -фрагментом.

Длительность пребывания фрагмента в базовой системе S_i , $i = 1, \dots, M$, имеет *экспоненциальное распределение с параметром* $\hat{\mu}_i$,

$$\hat{\mu}_i = \mu_i - \Lambda_{in}^*(S_i),$$

где $\Lambda_{in}^*(S_i)$ есть *суммарная взвешенная интенсивность* входящего в S_i потока.

Предполагается, что выполнено условие

$$\mu_i - \Lambda_{in}^*(S_i) > 0, \quad i = 1, \dots, M.$$

Зависимость параметра $\hat{\mu}_i$ длительности обслуживания может быть выбрана исходя из технических особенностей функционирования реальной системы.

Для систем обслуживания N_i , $i = 1, \dots, L$, обозначим через

- $\lambda_{in}^{(k,l)}(N_i)$, $\lambda_{out}^{(k,l)}(N_i)$ интенсивности входящего и выходящего потоков (k,l) -фрагментов;
- $\lambda_{in}^{(0)}(N_i)$, $\lambda_{out}^{(0)}(N_i)$ интенсивности входящего и выходящего потоков (0) -фрагментов.

Суммарная взвешенная интенсивность $\Lambda_{in}^*(N_i)$ входящего в N_i потока фрагментов определяется выражением

$$\Lambda_{in}^*(N_i) = \lambda_{in}^{(0)}(N_i) + \sum_{k=1}^K \sum_{l \in \mathcal{A}(k)} \lambda_{in}^{(k,l)}(N_i) w_k(l). \quad (3.2)$$

Тогда *средний вес* фрагментов $w(N_i)$, поступающих в N_i ,

$$w(N_i) = \frac{\Lambda_{in}^*(N_i)}{\Lambda_{in}(N_i)},$$

где $\Lambda_{in}(N_i)$ определяет суммарную интенсивность входящего потока фрагментов в N_i ,

$$\Lambda_{in}(N_i) = \lambda_{in}^{(0)}(N_i) + \sum_{k=1}^K \sum_{l \in \mathcal{A}(k)} \lambda_{in}^{(k,l)}(N_i). \quad (3.3)$$

Интенсивности входящих и выходящих потоков фрагментов для квазиэлементарной сети можно найти в два этапа:

1. Найти $\lambda_{in}^{(0)}(F_k)$, для всех $k = 1, \dots, K$:

$$\lambda_{in}^{(0)}(F_k) = \Lambda_0 \theta_{0, x_k^F}^{(0)}. \quad (3.4)$$

2. Найти интенсивности потоков фрагментов $\lambda_{in}^{(k,l)}(N_j)$, $k = 1, \dots, K$, $l \in \mathcal{A}(k)$:

$$\lambda_{in}^{(k,l)}(N_j) = \lambda_{in}^{(0)}(F_k) \theta_{x_k^F, x_i^S}^{(k)} + \sum_{i=1}^M \lambda_{in}^{(k,l)}(S_i) \theta_{x_i^S, j}^{(k)}. \quad (3.5)$$

Рассмотрим задачу нахождения для каждого дивайдера F_k , $k = 1, \dots, K$, такого *распределения весов* $\{w_k^*(l) : l \in \mathcal{A}(k)\}$, которое *минимизирует мате-*

математическое ожидание длительности пребывания требований в сети обслуживания. Предполагается, что найденное распределение может иметь нулевые компоненты, что будет означать запрет поступления фрагментов в соответствующие базовые системы.

Для начала рассмотрим зависимость среднего веса фрагментов, поступающих в базовые системы сети обслуживания, от распределения весов \mathcal{W} .

Из (3.4) и (3.5) следует, что средний вес фрагментов, поступающих в системы S_i , определяется следующей зависимостью

$$w(S_i) = \sum_{k=1}^K \sum_{l \in \mathcal{A}(k)} \varepsilon_{k,l}^{(i)} w_k(l), \quad i = 1, \dots, M. \quad (3.6)$$

где коэффициенты $\varepsilon_{k,l}^{(i)}$ определяются только набором матриц передач Θ сети обслуживания. Тогда средний вес фрагментов, поступающих в каждую из систем, является выпуклой функцией распределения весов \mathcal{W} . При этом множество \mathcal{W} также является выпуклым множеством в силу линейности всех ограничений, наложенных на это множество, а именно

$$\begin{cases} \sum_{l \in \mathcal{A}(k)} w_k(l) = 1, & k = 1, \dots, K, \\ \mu_i > 0. \end{cases} \quad (3.7)$$

Будем рассматривать длительность реакции τ_k для каждой элементарной подсети \mathcal{H}_k . Пусть γ_k обозначает вероятность поступления требования из источника в дивайдер F_k , тогда для математического ожидания $\mathbf{E}[\tau]$ длительности пребывания требований в сети обслуживания справедливо

$$\mathbf{E}[\tau] = \sum_{k=1}^K \gamma_k \mathbf{E}[\tau_k]. \quad (3.8)$$

Распределение для длительности реакции элементарной сети может быть найдено с использованием теоремы 1. Напомним, что математическое ожидание $\mathbf{E}[\xi]$ для случайной величины $\xi \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{A})$ определяется как

$$\mathbf{E}[\xi] = -\boldsymbol{\alpha} \mathbf{A}^{-1} \mathbf{1}. \quad (3.9)$$

С другой стороны,

$$\mathbf{E} [\xi] = -\mathbf{z}\mathbf{1}, \quad (3.10)$$

где \mathbf{z} есть решение системы линейных уравнений

$$\mathbf{z}\mathbf{A} = \boldsymbol{\alpha}. \quad (3.11)$$

Для решения поставленной задачи минимизации приемлемым способом является использование численных методов нелинейной оптимизации.

Выражение для целевой функции с использованием формулы Литтла можно переписать в следующем виде

$$\mathbf{E} [\tau] = \sum_{k=1}^K \gamma_k \mathbf{E} [\tau_k] = \frac{1}{\Lambda_0} \sum_{k=1}^K q(\mathcal{H}_k), \quad (3.12)$$

здесь $q(\mathcal{H}_k)$ обозначает математическое ожидание числа требований в подсети \mathcal{H}_k ,

$$q(\mathcal{H}_k) = \lambda_{in}^{(0)}(F_k) \mathbf{E} [\tau_k]. \quad (3.13)$$

Получаем, что необходимо минимизировать следующую целевую функцию

$$\sum_{k=1}^K q(\mathcal{H}_k) = \sum_{k=1}^K \lambda_{in}^{(0)}(F_k) \mathbf{E} [\tau_k] \rightarrow \min. \quad (3.14)$$

Из неравенства Йенсена следует, что увеличение математического ожидания числа требований в любой подсети будет приводить к увеличению математического ожидания длительности пребывания требований в сети обслуживания,

$$\mathbf{E} [\max \{\xi_1, \dots, \xi_n\}] \geq \max \{\mathbf{E} [\xi_1], \dots, \mathbf{E} [\xi_n]\}. \quad (3.15)$$

Для первоначальной минимизации математического ожидания длительности пребывания требований в сети обслуживания можно произвести последовательное устранение «узких мест» в сети.

Алгоритм 4. Эвристическая оптимизация распределения весов в сети обслуживания

Пусть задано начальное распределение $\{w_k(l) : l \in \mathcal{A}(k)\}$, $k = 1, \dots, K$.

1. Определить математическое ожидание $q(\mathcal{H}_k)$ числа требований в каждой подсети \mathcal{H}_k ;
2. Найти «узкое место» сети обслуживания— подсеть \mathcal{H}_{k^*} с максимальным математическим ожиданием числа требований в ней;
3. Выполнить процедуру оптимизации подсети \mathcal{H}_{k^*} :
 - Найти ветви $\mathcal{B}(k^*, S_{i^*})$, $\mathcal{B}(k^*, S_{j^*})$ в подсети \mathcal{H}_{k^*} с максимальным и минимальным математическим ожиданием длительностей прохождения фрагментов соответственно.
 - Перенести δ единиц веса из $\mathcal{B}(k^*, S_{i^*})$ в $\mathcal{B}(k^*, S_{j^*})$,

$$\delta = \min \left\{ \frac{\bar{w}_k(j^*) - w_k(j^*)}{2}, \frac{w_k(i^*)}{2} \right\}, \quad (3.16)$$

здесь $\bar{w}_k(j^*)$ обозначает максимально допустимое значение веса для фрагментов, поступающих в S_{j^*} .

4. Если после выполнения перераспределения веса, ветви $\mathcal{B}(k^*, S_{i^*})$, $\mathcal{B}(k^*, S_{j^*})$ в подсети \mathcal{H}_{k^*} имеют максимальное и минимальное математическое ожидание длительностей прохождения фрагментов, то выполнить процедуру оптимизации подсети снова.
 5. Пересчитать характеристики сети обслуживания и перейти к шагу 1.
- Конец алгоритма.

Рассмотрим задачу оптимизации для сети обслуживания, представленной на рисунке 3.5, для которой $\boldsymbol{\mu} = (1, 2, 2, 3, 2, 5, 6)$.

Задача состоит в нахождении распределения весов

$$\mathcal{W}^* = \{w_1^*(1), w_1^*(2), w_2^*(2), w_2^*(3), w_2^*(5)\}.$$

Таблица 3.3 содержит значения оптимальных распределений при различных значениях интенсивности входящего потока Λ .

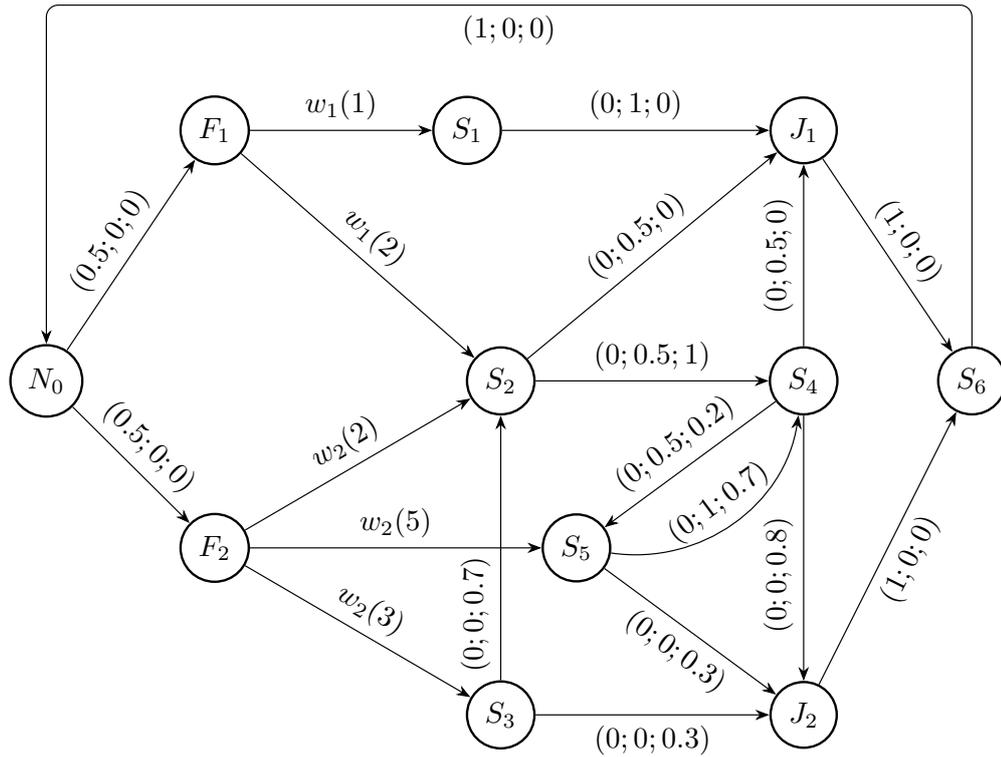


Рисунок 3.5 — Пример сети обслуживания для задачи оптимизации

Таблица 3.3 — Результаты оптимизации распределения весов в сети массового обслуживания

Λ	τ	$w_1^*(1)$	$w_1^*(2)$	$w_2^*(2)$	$w_2^*(3)$	$w_2^*(5)$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\mu}_5$	$\hat{\mu}_6$
1	1.7	0	1	0	0	1	1	1.5	2	2.09	1.17	4
1.2	1.9	0	1	0	0	1	1	1.4	2	1.91	1	3.8
1.4	2.2	0	1	0	0	1	1	1.3	2	1.73	0.84	3.6
1.6	2.5	0.45	0.55	0	0	1	0.64	1.56	2	1.91	0.85	3.4
1.8	2.8	0.48	0.52	0	0	1	0.57	1.53	2	1.8	0.72	3.2
2	3.1	0.45	0.55	0	0.55	0.45	0.55	1.07	1.45	1.64	1.11	3
2.2	3.3	0.46	0.54	0	0.56	0.44	0.5	0.97	1.39	1.51	1.04	2.8
2.4	3.6	0.45	0.55	0	0.54	0.46	0.45	0.89	1.35	1.37	0.93	2.6
2.6	4	0.46	0.54	0	0.56	0.44	0.41	0.79	1.28	1.23	0.86	2.4
2.8	4.5	0.46	0.54	0	0.56	0.44	0.36	0.69	1.22	1.1	0.77	2.2
3	5.1	0.45	0.55	0	0.55	0.45	0.32	0.6	1.18	0.96	0.67	2
3.2	5.8	0.46	0.54	0	0.55	0.45	0.27	0.51	1.11	0.83	0.59	1.8
3.4	6.9	0.45	0.55	0	0.55	0.45	0.24	0.4	1.06	0.68	0.49	1.6
3.6	8.5	0.45	0.55	0	0.55	0.45	0.18	0.33	1.01	0.55	0.4	1.4

Из таблицы видно, что при малых интенсивностях потока, использование медленной базовой системы S_1 для фрагментов требований, поступающих в дивайдер F_1 , неоправданно. Вместо этого используется маршрут, состоящий из базовых систем множества $\{S_2, S_4, S_5\}$. Однако при увеличении интенсивности входящего потока длительность прохождения фрагментов через этот маршрут увеличивается, а использование системы S_1 приносит выигрыш.

Аналогичная ситуация происходит при делении требований на дивайдере F_2 .

3.3 Моделирование сетей с многопутевой маршрутизацией

Для современных сетей передачи данных со сложной структурой и огромным количеством пользователей характерно увеличение числа информационных потоков, циркулирующих в сети, при этом, возрастают требования к уровню надёжности и пропускной способности. В таком случае удобным и эффективным способом решения возникающих проблем является распараллеливание информационных потоков. При данном подходе поток от источника к получателю разделяется на несколько субпотоков, которые передаются по различным маршрутам, что позволяет эффективно распределить нагрузку в сети и тем самым увеличить пропускную способность. Однако при этом возникает необходимость осуществления синхронизации всех субпотоков. В настоящее время одним из методов организации такого типа маршрутизации является протокол транспортного уровня МРТСП (Multipath Transmission Control Protocol) [117], который представлен как набор расширений однопутевого ТСП. Если соединение использует протокол МРТСП, то возможен обмен пакетами с несколькими адресами/интерфейсами одновременно, в рамках одного соединения. Таким образом, при использовании многопутевой маршрутизации передача данных происходит одновременно по нескольким маршрутам, которые могут быть не полностью изолированными (совместно использовать несколько маршрутизаторов).

В качестве типичных примеров использования протокола МРТСП его разработчики приводят следующие ситуации [118]:

- увеличение скорости доступа за счёт подключения к нескольким провайдерам (подключение устройства по нескольким интерфейсам),
- улучшение качества мобильной связи за счёт подключения к нескольким точкам доступа одновременно,
- одновременное использование сетей Wi-Fi и мобильной связи (4G).

Для моделирования процессов, которые происходят с субпотокami в сетях передачи данных с многопутевой маршрутизацией, воспользуемся сетями массового обслуживания с делением и слиянием требований [119].

Пусть рассматриваемая сеть передачи данных с многопутевой маршрутизацией состоит из

- множества терминалов (пользователей)

$$\mathcal{U} = \{U_1, U_2, \dots, U_{c_U}\},$$

- множества маршрутизаторов

$$\mathcal{R} = \{R_1, R_2, \dots, R_{c_R}\}.$$

Терминалы и маршрутизаторы соединены в сеть передачи данных посредством каналов связи. Терминал $U \in \mathcal{U}$ может одновременно посылать и принимать данные, то есть терминал поддерживает несколько соединений, каждое из которых осуществляет приём, либо передачу данных. На выходе из терминала потоки данных делятся на несколько субпотокoв, а при поступлении в терминал объединяются.

Пусть между двумя терминалами установлено соединение, в котором отправителем является U_i , а получателем U_j , в таком случае будем обозначать это соединение упорядоченной парой (U_i, U_j) . Для каждой пары зададим интенсивность потока $\lambda(U_i, U_j)$ от источника U_i к получателю U_j , предполагается, что при передаче поток от U_i делится на $d_{i,j}$ субпотокoв, которые затем объединяются у получателя U_j .

Предполагается, что поток (субпоток) состоит из однотипных пакетов, который в модели представим потоком требований (фрагментов).

На рисунке 3.6 изображена часть модели соединения (U_i, U_j) , состоящая из

- дивайдера $F_{i,j}$, осуществляющего деление требований на $d_{i,j}$ фрагментов,
- интегратора $J_{i,j}$, осуществляющего объединение фрагментов, полученных на дивайдере $F_{i,j}$,
- пуассоновского потока требований, поступающего с интенсивностью $\Lambda(F_{i,j})$, на дивайдер $F_{i,j}$.

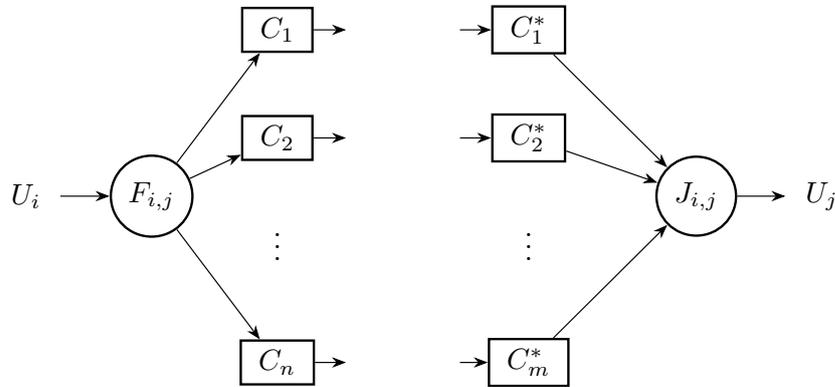


Рисунок 3.6 — Пример модели соединения между пользователями

Маршрутизатор $R \in \mathcal{R}$ принимает поступающие в него потоки данных, выполняет обработку и перенаправляет их на выходные порты (каналы). Будем рассматривать маршрутизатор как бесконечноприборную систему массового обслуживания с экспоненциальной длительностью обслуживания. Другие более сложные модели маршрутизатора, например, в виде поллинговых систем обслуживания, предложены в [81]. Каналы будем моделировать также бесконечноприборными системами массового обслуживания с экспоненциальной длительностью обслуживания.

Будем предполагать, что *интенсивность обслуживания зависит от нагрузки*.

На рисунке 3.7 представлен пример маршрутизатора, соединённого с каналами C_1, \dots, C_k .

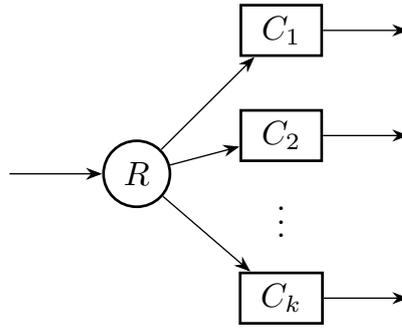


Рисунок 3.7 — Пример модели маршрутизатора, соединённого с несколькими каналами

Рассмотрим сеть передачи данных, структура которой представлена на рисунке 3.8.

Положим, что в сети имеются следующие соединения (U_1, U_2) , (U_2, U_1) , (U_3, U_1) . Каждое из этих соединений использует следующие множества детерминированных маршрутов соответственно:

1. $\{(U_1, R_1, U_2), (U_1, R_2, R_3, R_4, U_2)\}$,
2. $\{(U_2, R_1, U_1), (U_2, R_4, R_3, R_2, U_1)\}$,
3. $\{(U_3, R_3, R_1, U_1), (U_3, R_2, U_1)\}$.

Таким образом, каждому соединению предоставлено для использования два маршрута для передачи субпотокa.

Модельная сеть обслуживания для представленной сети передачи изображена на рисунке 3.9. Здесь опущен источник требований N_0 , предполагается, что требования поступают из источника непосредственно в дивайдеры.

В построенной сети обслуживания фрагменты передаются по следующим маршрутам:

1. $\{(F_{1,2}, S_5, S_1, S_6, J_{1,2}), (F_{1,2}, S_{12}, S_2, S_9, S_3, S_8, S_4, S_7, J_{1,2})\}$,
2. $\{(F_{2,1}, S_6, S_1, S_5, J_{2,1}), (F_{2,1}, S_7, S_4, S_8, S_3, S_9, S_2, S_{12}, J_{2,1})\}$,
3. $\{(F_{3,1}, S_{10}, S_3, S_1, S_5, J_{3,1}), (F_{3,1}, S_{11}, S_2, S_{12}, J_{3,1})\}$.

Здесь базовые системы S_1, S_2, S_3, S_4 , соответствуют маршрутизаторам R_1, R_2, R_3, R_4 соответственно. Остальные базовые системы моделируют каналы передачи данных.

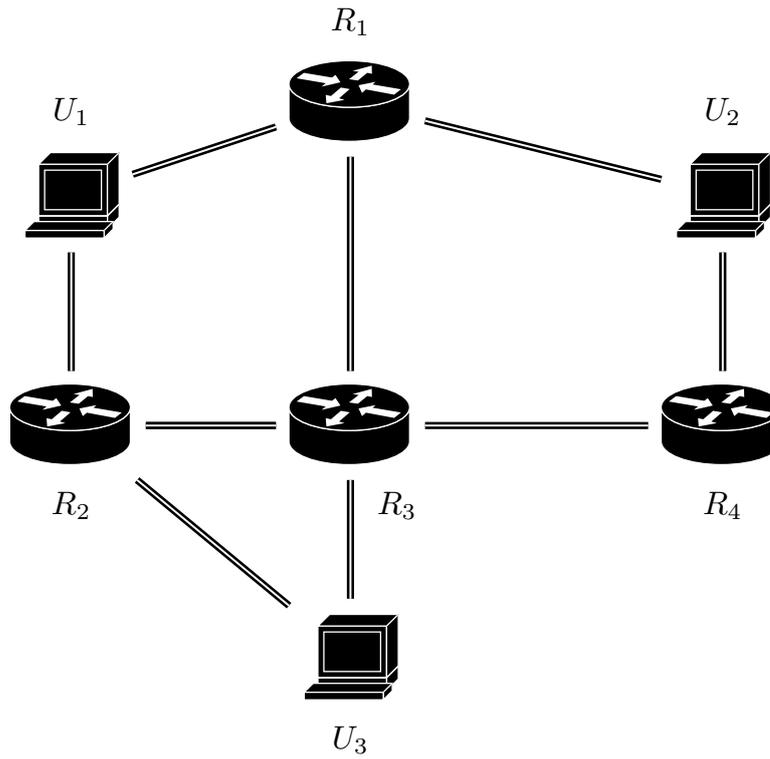


Рисунок 3.8 — Пример сети передачи данных

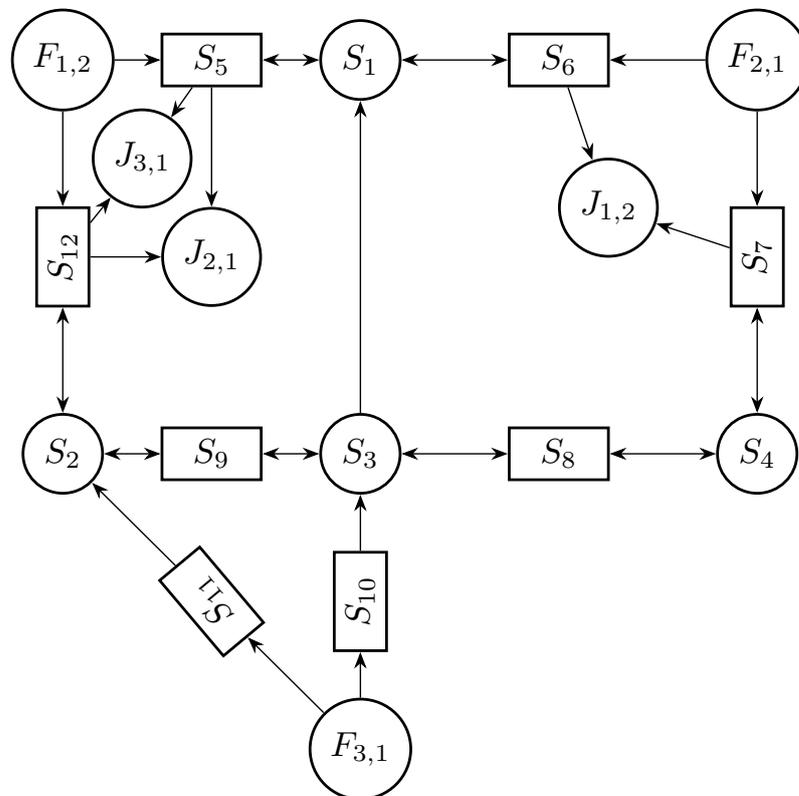


Рисунок 3.9 — Модель сети передачи данных в виде сети обслуживания с делением и слиянием требований

Глава 4. Комплекс программ имитационного моделирования и численного анализа сетей массового обслуживания с делением и слиянием требований

В данной главе описывается комплекс программ, разработанный для анализа сетей обслуживания произвольной топологии с делением и слиянием требований. Приводится краткое описание структуры программ, примеры их использования, обсуждаются полученные результаты.

Представленный проблемно-ориентированный комплекс состоит из программы имитационного моделирования сетей обслуживания, а также программы для численных расчётов вероятностных характеристик сетей обслуживания на основе полученных в работе результатов. Интерфейс пользователя позволяет удобно описывать исследуемые модели, проводить моделирование, выполнять анализ результатов.

С использованием программного комплекса исследуется возможность использования бесконечноприборных моделей с зависимой от нагрузки интенсивностью обслуживания для приближённого анализа аналогичных сетей с одноприборными базовыми системами.

4.1 Структура имитационной модели

Описание архитектуры имитационной модели [120] представлено в [121]. В разработанной имитационной модели сети обслуживания каждому типу систем соответствует описание на языке программирования, оформленное в виде некоторого класса. Требования и фрагменты в имитационной модели представлены также объектами некоторого класса, наделённого существенными для данной модели значениями.

Представление элементов сети обслуживания объектами позволяет добиться независимости при разработке модели — каждый класс описывается отдельно, однако классы проектируются таким образом, что они могут взаимодействовать друг с другом посредством событий. К тому же использование

объектно-ориентированной методологии программирования позволяет использовать такие её преимущества как наследование, инкапсуляция, полиморфизм.

Имитационная модель является дискретно-событийной [122], то есть процесс функционирования сети обслуживания в имитационной модели представлен логически связанной последовательностью событий, характеризуемой интервалами времени между событиями и типом событий. Управление работой модели производится ведущей программой.

Кратко опишем ниже наиболее важные классы, из которых состоит разработанная программа имитационного моделирования.

Класс `NetworkModel` является основным классом для моделирования сети обслуживания и содержит следующие поля:

- `Nodes` — массив узлов сети обслуживания,
- `InfoNode` — информационный узел сети.

Здесь под узлом сети понимается объект класса `Node`, моделирующий процессы, возникающие в системе обслуживания или источнике. В процессе работы имитационной модели узлы обмениваются фрагментами друг с другом. Информационный узел обеспечивает хранение и передачу информации, необходимой всем остальным узлам в процессе функционирования имитационной модели, например, значение модельного времени.

Класс `Node` содержит следующие поля:

- `ID` — уникальный идентификатор узла в сети обслуживания;
- `Buffer` — буффер для хранения фрагментов в узле;
- `NodesOut` — массив узлов, в которые могут поступать фрагменты из текущего узла;
- `InfoNode` — информационный узел сети;
- `NextEventTime` — момент времени возникновения следующего события в этом узле.

В классе `Node` определены следующие методы:

- `Activate()` — обработка ближайшего события в текущем узле;

- **Route(f)** — отправка фрагмент по сети обслуживания согласно заданной в сети обслуживания маршрутизации;
- **Send(f, N)** — отправка фрагмент **f** в узел-получатель **N**;
- **Receive(f)** — реализация процесса получения фрагмента **f** текущим узлом.

Таким образом, в момент времени возникновения события в некотором узле, управление передаётся методу **Activate()** данного узла, который в зависимости от типа события и состояния узла выполняет обработку этого события. Процедура обработки может включать в себя, например, получение узлом фрагмента от другого узла или отправку фрагмента текущим узлом в некоторый узел.

Класс **Node** является родительским для следующих классов:

- **ServerNode** — класс, реализующий процесс функционирования базовой системы;
- **ForkNode** — класс, реализующий процесс функционирования дивайдера;
- **JoinNode** — класс, реализующий процесс функционирования интегратора;
- **SourceNode** — отображает источник требований.

Каждый из этих классов реализует специфические алгоритмы функционирования для соответствующих объектов сети обслуживания.

Фрагмент в имитационной модели описан классом **Fragment**, который содержит:

- **Sigma** — содержит описание для сигнатуры фрагмента в виде набора (**ParentFragment**, **ForkNodeID**, **SubID**), где каждая компонента соответствует компонентам сигнатуры, определённым ранее;
- **CreationTime** — определяет момент времени создания фрагмента. Для требований, поступающих из источника, соответствует моменту поступления, а для фрагментов, полученных в результате деления в дивайдере, соответствует моменту деления;

- **ArrivalTime** — момент поступления фрагмента в систему обслуживания;
- **StartServiceTime** — момент начала обслуживания фрагмента в системе обслуживания;
- **LeaveTime** — момент завершения обслуживания фрагмента в системе обслуживания;
- **TotalTime** — длительность времени пребывания в сети обслуживания.

Запуск имитационной модели производится посредством вызова метода `Run()` экземпляра класса `NetworkModel`. Алгоритм работы метода `Run()` состоит из следующих шагов:

1. Установить значение для модельного времени `CurrentTime:=0` для информационного узла.
2. Выбор узла из массива `Nodes []` с наименьшим значением момента времени возникновения следующего события `NextEventTime`.
3. Передвинуть текущее значение модельного времени до момента времени возникновения ближайшего события, значение которого было определено на предыдущем шаге.
4. Для выбранного узла вызвать метод `Activate()` обработки возникшего события.
5. Переход к шагу 2.

В качестве условия останова работы алгоритма может быть выбрано ограничение на максимальное значение модельного времени или факт накопления достаточной выборки, необходимой для последующего статистического анализа.

После того, как имитационное моделирование завершено, выполняется обработка результатов — статистический анализ полученных выборочных значений. А именно определяются оценки для следующих стационарных характеристик:

- стационарное распределение числа фрагментов в базовых системах сети,

- стационарное распределение числа фрагментов для каждой базовой системы,
- стационарное распределение числа требований в сети обслуживания,
- математическое ожидание числа фрагментов в сети обслуживания,
- функция и плотность распределения длительности пребывания требований в сети обслуживания,
- математическое ожидание и дисперсия длительности пребывания требований в сети обслуживания.

Вопросы, связанные со статистическим анализом результатов имитационного моделирования, обсуждаются в [122].

4.2 Программа численных расчётов

Программа для численного анализа сети обслуживания реализует предложенные в работе методы.

Вспомогательной частью программы является разработанная библиотека для работы со случайными величинами с фазовым распределением. Случайная величина с РН-распределением описана в классе `PhaseTypeVariable`, который содержит следующие свойства и методы:

- `SubGenerator` — субгенератор модельной поглощающей цепи Маркова,
- `InitialDistribution` — начальное распределение во множестве невозвратных состояний,
- `NumberOfPhases` — число фаз,
- `ExpectedValue()` — метод, возвращающий математическое ожидание,
- `Variance()` — метод, возвращающий дисперсию.

Также реализована операция `Max(PhaseTypeVariable[] RVs)`, результатом которой является объект `PhaseTypeVariable`, с параметрами соответствующими случайной величине — максимуму из величин в массиве `RVs`.

Для операций с матрицами используется библиотека `MathNet.Numerics` [123].

4.3 Вычислительные аспекты нахождения параметров длительности пребывания требований в сети обслуживания

Поскольку получение распределения длительности пребывания требований в сети обслуживания связано с задачей поиска параметров соответствующего фазового распределения, это приводит к большой вычислительной сложности. В данном разделе обсуждаются некоторые вычислительные аспекты реализации методов.

Отметим, что если $\xi \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{A})$ порядка c_ξ , $\eta \sim \text{PH}(\boldsymbol{\beta}, \mathbf{B})$ порядка c_η , то результатом операции « \vee » станет случайная величина ζ с параметрами $(\boldsymbol{\gamma}, \mathbf{C}) = (\boldsymbol{\alpha}, \mathbf{A}) \vee (\boldsymbol{\beta}, \mathbf{B})$ порядка $c_\xi c_\eta + c_\xi + c_\eta$. В результате этих манипуляций получают преимущественно разреженные матрицы.

Поскольку для n -го начального момента случайной величины ζ справедливо

$$\mathbb{E}[\zeta^n] = (-1)^n n! \boldsymbol{\gamma} (\mathbf{C}^{-1})^n \mathbf{1}, \quad n \geq 1,$$

рассмотрим подробнее задачу обращения блочной матрицы \mathbf{C} , образованной в результате операции « \vee » над матрицами, представленными в блочном виде.

$$\mathbf{C} = \begin{pmatrix} \mathbf{A} \oplus \mathbf{B} & \mathbf{I} \otimes (-\mathbf{B}\mathbf{1}) & (-\mathbf{A}\mathbf{1}) \otimes \mathbf{I} \\ \mathbf{0} & \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B} \end{pmatrix}. \quad (4.1)$$

Переобозначим блоки в матрице \mathbf{C}

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{1,1} & \mathbf{C}_{1,2} & \mathbf{C}_{1,3} \\ \mathbf{0} & \mathbf{C}_{2,2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{3,3} \end{pmatrix}. \quad (4.2)$$

Будем искать обратную матрицу непосредственно из матричного уравнения $\mathbf{C}\mathbf{C}^{-1} = \mathbf{I}$,

$$\begin{pmatrix} \mathbf{C}_{1,1} & \mathbf{C}_{1,2} & \mathbf{C}_{1,3} \\ \mathbf{0} & \mathbf{C}_{1,2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{1,2} \end{pmatrix} \begin{pmatrix} \mathbf{C}_{1,1}^* & \mathbf{C}_{1,2}^* & \mathbf{C}_{1,3}^* \\ \mathbf{C}_{2,1}^* & \mathbf{C}_{2,2}^* & \mathbf{C}_{2,3}^* \\ \mathbf{C}_{3,1}^* & \mathbf{C}_{3,2}^* & \mathbf{C}_{3,3}^* \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix},$$

предполагается, что каждая из диагональных единичных матриц имеет соответствующий порядок.

Из представленного уравнения следует, что

$$\mathbf{C}^{-1} = \begin{pmatrix} \mathbf{C}_{1,1}^{-1} & -\mathbf{C}_{1,1}^{-1}\mathbf{C}_{1,2}\mathbf{C}_{2,2}^{-1} & -\mathbf{C}_{1,1}^{-1}\mathbf{C}_{1,3}\mathbf{C}_{3,3}^{-1} \\ \mathbf{0} & \mathbf{C}_{2,2}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{3,3}^{-1} \end{pmatrix}. \quad (4.3)$$

Таким образом, данное выражение показывает, что можно свести обращение квадратной матрицы порядка $c_\xi c_\eta + c_\xi + c_\eta$ к задаче обращения трёх квадратных матриц меньшего порядка.

4.4 Уменьшение вычислительной сложности на основе уменьшения числа фаз

В качестве одного из возможных способов уменьшения вычислительной сложности задачи, рассмотрим способ аппроксимации исходных распределений случайных величин распределениями с меньшим числом фаз.

В качестве такого распределения может быть использовано гиперэкспоненциальное или обобщённое эрланговское распределение, которые принадлежат классу PH-распределений [109].

В [81] рассматривается метод аппроксимации функции распределения случайной величины ξ по двум моментам $\mathbf{E}[\xi]$ и $\mathbf{Var}[\xi]$.

В том случае, когда коэффициент вариации $c_v[\xi] = \frac{\mathbf{Var}[\xi]}{\mathbf{E}[\xi]} > 1$ для аппроксимации удобно использовать гиперэрланговское распределение, например, с числом этапов $r = 2$, функция распределения $f(t)$ которого имеет вид

$$f(t) = be^{-a_1 t} + (1 - b)e^{-a_2 t}, \quad t \geq 0.$$

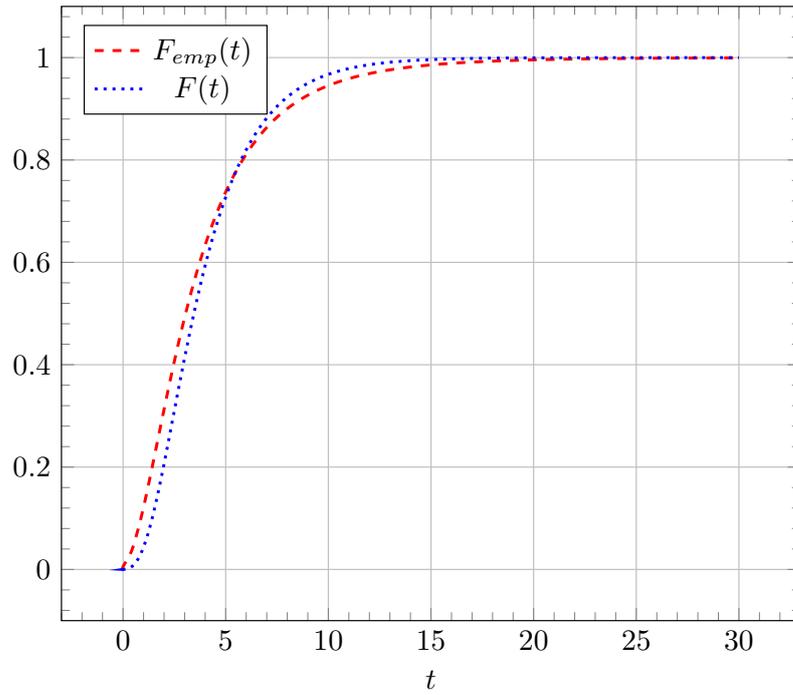


Рисунок 4.1 — Сравнение приближённой $F(t)$ и эмпирической $F_{emp}(t)$ функций распределения длительности пребывания требований в сети \mathcal{N}_1^f

Тогда

$$a_1 = 2b/E[\xi], \quad a_2 = 2(1 - b)/E[\xi].$$

$$b = \frac{1}{2} \left(1 \pm \sqrt{\frac{c_v[\xi]^2 - 1}{c_v[\xi]^2 + 1}} \right),$$

где в качестве значения b можно использовать любой корень.

Для случая $c_v[\xi] < 1$, используют обобщённое эрланговское распределение.

4.5 Аспекты практического использования комплекса

Для сетей обслуживания \mathcal{N}_1^f и \mathcal{N}_2^f с одноприборными базовыми системами, определённых в разделе 3.1, построим эмпирические функции и плотности распределения длительности пребывания требований в сети обслуживания и сравним их с полученным приближённым методом.

Предполагается, что интенсивность входящего потока равна 1.

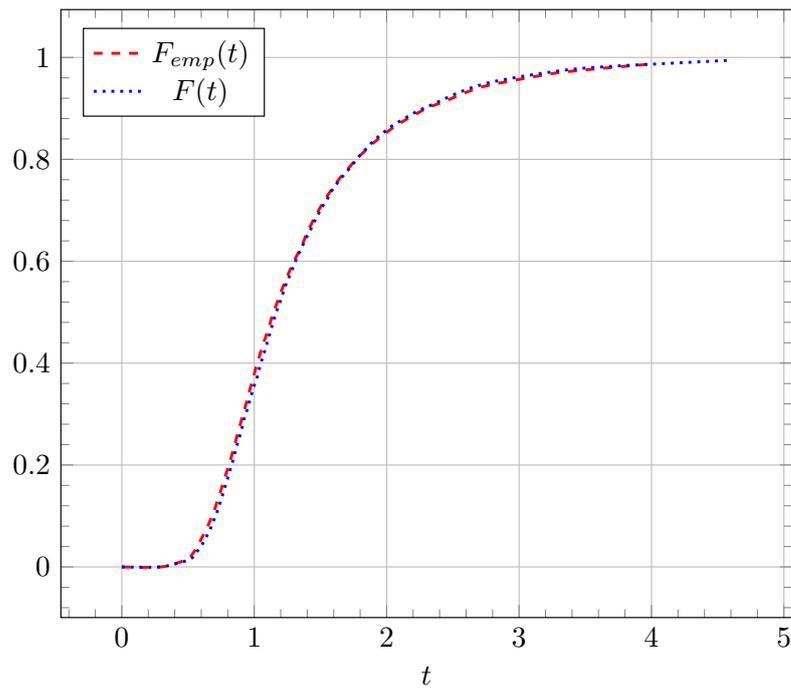


Рисунок 4.2 — Сравнение приближённой $F(t)$ и эмпирической $F_{emp}(t)$ функций распределения длительности пребывания требований в сети \mathcal{N}_2^f

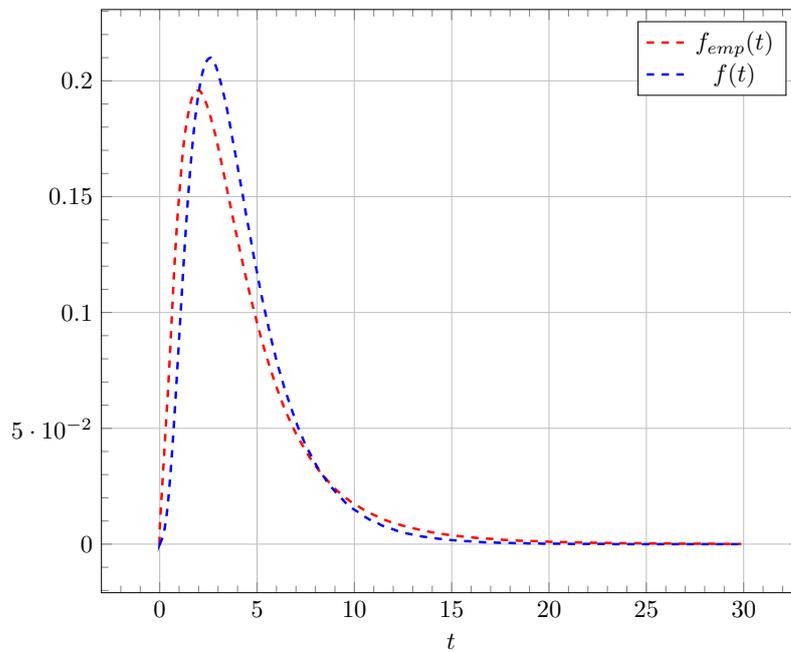


Рисунок 4.3 — Сравнение приближённой $f(t)$ и эмпирической $f_{emp}(t)$ плотностей распределения длительности пребывания требований в сети \mathcal{N}_1^f

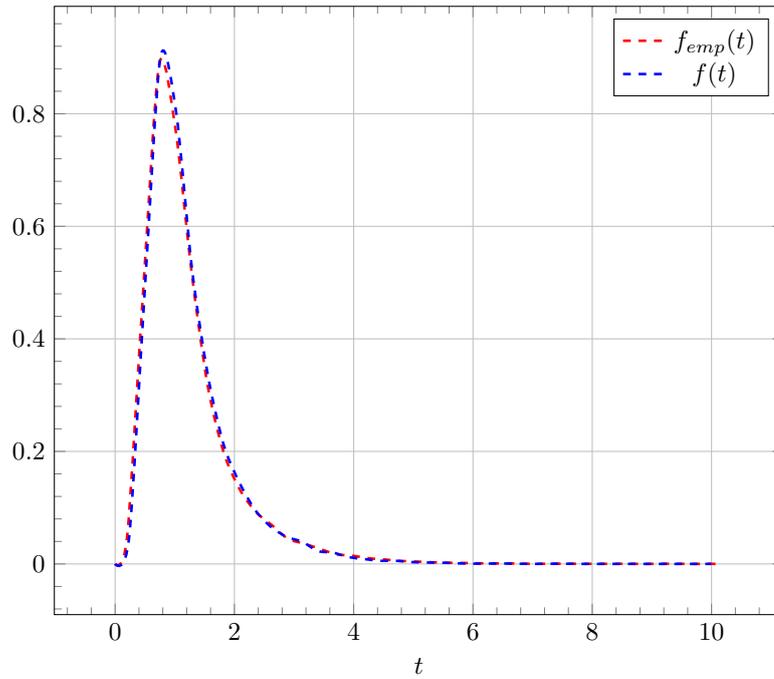


Рисунок 4.4 — Сравнение приближённой $f(t)$ и эмпирической $f_{emp}(t)$ плотностей распределения длительности пребывания требований в сети \mathcal{N}_2^f

Найдём расстояние Δ Колмогорова в каждом из случаев при различных значениях интенсивности входящего потока,

$$\Delta = \sup_t |F_{emp}(t) - F(t)|,$$

где F_{emp} есть эмпирическая функция распределения, построенная по выборке, F — некоторая заданная функция распределения.

Таблицы 4.1, 4.2 содержат полученные результаты вычисления статистики Δ . Объём выборки для получения статистических оценок составил 50000.

Таблица 4.1 — Расстояние Колмогорова для сети обслуживания \mathcal{N}_1^f

Λ	0.5	0.7	0.9	1.1	1.3
Δ	0.0246	0.1049	0.1248	0.0899	0.0780

Таблица 4.2 — Расстояние Колмогорова для сети обслуживания \mathcal{N}_2^f

Λ	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9
Δ	0.2067	0.0255	0.0406	0.0509	0.0497	0.0686	0.1025	0.0787
Λ	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5
Δ	0.0687	0.0590	0.0590	0.1315	0.0547	0.0625	0.0893	0.0752

Отметим, что были проведены эксперименты на других сетях обслуживания с делением и слиянием требований, которые дали аналогичные результаты, что позволяет говорить о возможности использования бесконечноприборных сетей обслуживания для приближённого анализа сетей с одноприборными базовыми системами.

Анализ полученных результатов показывает, что точность приближения, которая оценивалась относительной погрешностью, увеличивается с увеличением числа базовых систем в сети обслуживания. Данное наблюдение связано прежде всего с уменьшением зависимости между длительностями пребывания фрагментов, порождённых одним требованием, в ветвях сети, которая возникает вследствие общего момента поступления указанных фрагментов.

Таблица 4.3 содержит стационарные вероятности числа фрагментов в системах сети обслуживания \mathcal{N}_1 с бесконечноприборными базовыми системами в случае, когда интенсивность входящего потока требований $\Lambda = 1$.

Таблица 4.3 — Стационарные вероятности числа фрагментов для сети \mathcal{N}_1

Число фрагментов	S_1	S_2	S_3	S_4
0	0.5842	0.6065	0.5022	0.5301
1	0.3067	0.3032	0.3355	0.3034
2	0.0878	0.0758	0.1223	0.1193
3	0.0179	0.0126	0.0318	0.0356
4	0.0029	0.0016	0.0066	0.009
5	0.0004	0.0002	0.0011	0.002

Полученная для сети обслуживания \mathcal{N}_1 сеть размещений состоит из 27 систем массового обслуживания, основные параметры которой представлены в таблице 4.4.

Таблица 4.4 — Параметры сети размещений

СМО	Расшифровка	Интенсивность	Смежные СМО	Вероятности
l_1	$\{S_1, S_2, S_3\}$	6	l_2	0.3333
			l_3	0.0333
			l_4	0.3000
			l_5	0.2333
			l_6	0.1000
l_2	$\{S_2, S_3, S_4\}$	7	l_7	0.0286
			l_8	0.2571
			l_9	0.2000
			l_{10}	0.0857
			l_{11}	0.0857
			l_{12}	0.3429
l_3	$\{S_1, S_1, S_3\}$	6	l_7	0.6667
			l_{13}	0.2333
			l_{14}	0.1000
l_4	$\{S_1, S_3, J\}$	4	l_8	0.5000
			l_{15}	0.3500
			l_{16}	0.1500
l_5	$\{S_1, S_2, S_4\}$	7	l_1	0.0857
			l_6	0.3429
			l_9	0.2857
			l_{13}	0.0286
			l_{15}	0.2571
l_6	$\{S_1, S_2, J\}$	4	l_{10}	0.5000
			l_{14}	0.0500
			l_{16}	0.4500

Продолжение таблицы 4.4

СМО	Расшифровка	Интенсивность	Смежные СМО	Вероятности
l_7	$\{S_2, S_3, S_4\}$	7	l_4	0.3429
			l_{15}	0.0857
			l_{17}	0.2857
			l_{18}	0.2000
			l_{19}	0.0857
l_8	$\{S_3, S_4, J\}$	5	l_{20}	0.2800
			l_{21}	0.1200
			l_{22}	0.1200
			l_{23}	0.4800
l_9	$\{S_2, S_4, S_4\}$	8	l_2	0.1500
			l_{10}	0.6000
			l_{18}	0.0250
			l_{20}	0.2250
l_{10}	$\{S_2, S_4, J\}$	5	l_{12}	0.1200
			l_{15}	0.0400
			l_{21}	0.3600
			l_{24}	0.4800
l_{11}	$\{S_3, S_3, S_3\}$	6	l_2	0.4667
			l_{12}	0.2000
			l_{19}	0.0333
			l_{22}	0.300
l_{12}	$\{S_2, S_3, J\}$	4	l_4	0.0500
			l_{10}	0.3500
			l_{23}	0.4500
			l_{24}	0.1500

Продолжение таблицы 4.4

СМО	Расшифровка	Интенсивность	Смежные СМО	Вероятности
l_{13}	$\{S_1, S_1, S_4\}$	7	l_3	0.0857
			l_{14}	0.3429
			l_{18}	0.5714
l_{14}	$\{S_1, S_1, J\}$	4	l_{15}	1.0000
l_{15}	$\{S_1, S_4, J\}$	5	l_4	0.1200
			l_{16}	0.4800
			l_{20}	0.4000
l_{16}	$\{S_1, J, J\}$	2	l_{21}	1.0000
l_{17}	$\{S_3, S_4, S_4\}$	8	l_8	0.6000
			l_{20}	0.0750
			l_{25}	0.1750
			l_{26}	0.1500
l_{18}	$\{S_1, S_4, S_4\}$	8	l_7	0.1500
			l_{15}	0.6000
			l_{25}	0.2500
l_{19}	$\{S_1, S_3, S_3\}$	6	l_4	0.2000
			l_7	0.4667
			l_{26}	0.3333
l_{20}	$\{S_4, S_4, J\}$	6	l_8	0.2000
			l_{21}	0.8000
l_{21}	$\{S_4, J, J\}$	3	l_0	0.8000
			l_{23}	0.2000
l_{22}	$\{S_3, S_3, J\}$	4	l_8	0.7000
			l_{23}	0.3000

Окончание таблицы 4.4

СМО	Расшифровка	Интенсивность	Смежные СМО	Вероятности
l_{23}	$\{S_3, J, J\}$	2	l_0	0.3000
			l_{21}	0.7000
l_{24}	$\{S_2, J, J\}$	2	l_0	0.9000
			l_{16}	0.1000
l_{25}	$\{S_4, S_4, S_4\}$	9	l_{17}	0.2000
			l_{20}	0.8000
l_{26}	$\{S_3, S_3, S_4\}$	7	l_8	0.1714
			l_{17}	0.4000
			l_{22}	0.3429
			l_{27}	0.0875
l_{27}	$\{S_3, S_3, S_3\}$	6	l_{22}	0.3000
			l_{26}	0.7000

Заключение

Основные результаты работы заключаются в следующем.

1. Дано формальное описание сетей массового обслуживания произвольной топологии с делением и слиянием требований.
2. Разработаны методы анализа сетей массового обслуживания произвольной топологии с делением и слиянием требований в случае бесконечноприборных базовых систем обслуживания. Доказано, что длительность пребывания требований в сети массового обслуживания произвольной топологии с делением и слиянием требований имеет фазовое распределение и определены его параметры.
3. Для элементарной сети обслуживания с бесконечноприборными базовыми системами получен метод нахождения стационарного распределения вероятностей состояний сети обслуживания.
4. Предложен приближённый метод для сетей обслуживания с одноприборными базовыми системами, исследована его точность посредством имитационного моделирования.
5. Рассмотрена задача оптимального распределения весов в квазиэлементарной сети обслуживания с делением и слиянием требований.
6. Разработан комплекс программ имитационного моделирования и численного анализа сетей массового обслуживания с делением и слиянием требований.

Полученные результаты являются новыми, позволяют существенно расширить круг решаемых задач в теории массового обслуживания и могут быть использованы в задачах анализа, оптимизации и синтеза дискретных систем с сетевой структурой и стохастическим характером функционирования.

Дальнейшие исследования могут быть посвящены задачам синтеза и оптимизации сетей массового обслуживания произвольной топологии с делением и слиянием требований.

Список литературы

1. Models for Parallel and Distributed Computation / ed. by R. Corrêa [et al.]. – New York : Springer, 2002. – 323 p.
2. Voxma O. J., Koole G., Liu Z. Queueing-Theoretic Solution Methods for Models of Parallel and Distributed Systems : tech. rep. / Centrum voor Wiskunde en Informatica. – 1994. – 25 p.
3. Plateau B., Fourneau J.-M. A methodology for solving Markov models of parallel systems // Journal of parallel and distributed computing. – 1991. – Vol. 12, № 4. – P. 370–387.
4. Боровков А. А. Вероятностные процессы в теории массового обслуживания. – Москва : Наука, 1971. – 368 с.
5. Боровков А. А. Асимптотические методы в теории массового обслуживания. – Москва : Наука, 1980. – 382 с.
6. Башарин Г. П., Толмачев А. Л. Теория сетей массового обслуживания и ее приложения к анализу информационно-вычислительных систем // Итоги науки и техн. Сер. Теор. вероятн. Мат. стат. Теор. кибернет. – 1983. – Т. 21. – С. 1–119.
7. Бочаров А. А., Печинкин А. В. Теория массового обслуживания: Учебник. – Москва : Изд-во РУДН, 1995. – 529 с.
8. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. – Москва : Наука, 1966. – 432 с.
9. Кёниг Д., Рыков В. В., Шмидт Ф. Стационарные системы массового обслуживания с зависимостями // Итоги науки и техн. Сер. Теор. вероятн. Мат. стат. Теор. кибернет. – 1981. – Т. 18. – С. 95–186.
10. Рыков В. В. Управляемые системы массового обслуживания // Итоги науки и техн. Сер. Теор. вероятн. Мат. стат. Теор. кибернет. – 1975. – Т. 12. – С. 43–153.

11. Назаров А. А., Терпугов А. Ф. Теория массового обслуживания. – Томск : Изд-во НТЛ, 2010. – 238 с.
12. Вишнеvский В. М., Семенова О. В. Системы Поллинга: Теория и применение в широкополосных беспроводных сетях. – Москва : Техносфера, 2007. – 312 с.
13. Митрофанов Ю. И. Анализ сетей массового обслуживания. – Саратов : Изд-во Научная книга, 2004. – 176 с.
14. Митрофанов Ю. И. Синтез сетей массового обслуживания. – Саратов : Изд-во Саратов. ун-та, 1995. – 164 с.
15. Уолрэнд Д. Введение в теорию сетей массового обслуживания. – Москва : Мир, 1993. – 336 с.
16. Ивницкий В. А. Теория сетей массового обслуживания. – Москва : Изд-во физ.-мат. лит., 2004. – 772 с.
17. Queueing Networks and Markov Chains: Modeling and Performance Evaluation With Computer Science Applications: Second Edition / G. Bolch [et al.]. – New Jersey : John Wiley & Sons, Inc., 2006. – 878 p.
18. Кениг Д., Штоян Д. Методы теории массового обслуживания. – Москва : Радио и связь, 1981. – 128 с.
19. Jackson J. R. Networks of Waiting Lines // Operations Research. – 1957. – Vol. 5, № 4. – P. 518–521.
20. Jackson J. R. Jobshop-Like Queueing Systems // Management Science. – 2004. – Vol. 50, 12_supplement. – P. 1796–1802.
21. Buzen J. P. Computational Algorithms for Closed Queueing Networks with Exponential Servers // Commun. ACM. – New York, 1973. – Vol. 16, № 9. – P. 527–531.

22. Open, Closed, and Mixed Networks of Queues with Different Classes of Customers / F. Baskett [et al.] // J. ACM. – New York, 1975. – Vol. 22, № 2. – P. 248–260.
23. Chandy K. M., Howard Jr. J. H., Towsley D. F. Product Form and Local Balance in Queueing Networks // J. ACM. – New York, 1977. – Vol. 24, № 2. – P. 250–263.
24. Harrison J. M. Brownian Models of Queueing Networks with Heterogeneous Customer Populations // Stochastic Differential Systems, Stochastic Control Theory and Applications / ed. by W. Fleming, P.-L. Lions. – New York : Springer, 1988. – P. 147–186.
25. Gelenbe E. Product-Form Queueing Networks with Negative and Positive Customers // Journal of Applied Probability. – 1991. – Vol. 28, № 3. – P. 656–663.
26. Kumar P. R., Meyn S. P. Stability of queueing networks and scheduling policies // IEEE Transactions on Automatic Control. – 1995. – Vol. 40, № 2. – P. 251–260.
27. Kelly F. P. Reversibility and Stochastic Networks. – New York : Cambridge University Press, 2011. – 238 p.
28. Применение многоканальных систем массового обслуживания с отказами к конструированию телекоммуникационных сетей / Г. В. Цициашвили [и др.] // Дальневосточный математический журнал. – 2018. – Т. 18, вып. 1. – С. 123–126.
29. О распределении времени первого обрыва связи в беспроводных сетях с кэшированием / Ю. Н. Орлов [и др.] // Математическое моделирование. – 2018. – Т. 30, вып. 8. – С. 131–142.

30. Моделирование отношения сигнал/интерференция в мобильной сети со случайным блужданием взаимодействующих устройств / Ю. В. Гайдамака [и др.] // Информатика и её применения. – 2017. – Т. 11, вып. 2. – С. 50–58.
31. К анализу показателей качества обслуживания в современных беспроводных сетях / О. Г. Вихрова [и др.] // Информатика и её применения. – 2015. – Т. 9, вып. 4. – С. 48–55.
32. Study of the $MMPP/GI/\infty$ queueing system with random customers capacities / E. Lisovskaya [и др.] // Информатика и её применения. – 2017. – Т. 11, вып. 4. – С. 109–117.
33. Дудин С. А., Дудина О. С. Модель функционирования центра информационной и технической поддержки как двухфазная система массового обслуживания // Проблемы передачи информации. – 2013. – Т. 49, вып. 1. – С. 66–82.
34. Wu H., Wolter K. Stochastic Analysis of Delayed Mobile Offloading in Heterogeneous Networks // IEEE Transactions on Mobile Computing. – 2018. – Vol. 17, № 2. – P. 461–474.
35. Xu L., Zhuang W. Energy-Efficient Cross-Layer Resource Allocation for Heterogeneous Wireless Access // IEEE Transactions on Wireless Communications. – 2018. – Vol. 17, № 7. – P. 4819–4829.
36. Delay and Physical Layer Security Tradeoff in Large Wireless Networks / Y. Zhong [et al.] // 2018 IEEE International Conference on Communications (ICC). – 2018. – P. 1–7.
37. Novel trends in automotive networks: Performance evaluations based on Network queueing theory / A. Hasnaoui [et al.] // 2017 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC). – 2017. – P. 195–200.

38. Link-state routing protocol for flow optimization in delay-constrained queueing networks / M. D. Kartik [et al.] // 2015 10th International Conference on Information, Communications and Signal Processing (ICICS). – 2015. – P. 1–5.
39. Leahu H., Mandjes M., Oprescu A. A Numerical Approach to Stability of Multiclass Queueing Networks // IEEE Transactions on Automatic Control. – 2017. – Vol. 62, № 10. – P. 5478–5484.
40. Bandi C., Bertsimas D., Youssef N. Robust transient analysis of multi-server queueing systems and feed-forward networks // Queueing Systems. – 2018. – Vol. 89, № 3/4. – P. 351–413.
41. Xia L., Chen S. Dynamic Pricing Control for Open Queueing Networks // IEEE Transactions on Automatic Control. – 2018. – Vol. 63, № 10. – P. 3290–3300.
42. Thomasian A. Analysis of Fork/Join and Related Queueing Systems // ACM Computing Surveys. – New York, 2014. – Vol. 47, № 2. – 17:1–17:71.
43. Обзор систем параллельной обработки заявок / А. В. Горбунова [и др.] // Вестник Российского университета дружбы народов. Серия: Математика. Информатика. Физика. – 2017. – Т. 25, № 4. – С. 350–362.
44. Горбунова А. В., Зарядов И. С., Самуйлов К. Е. Обзор систем параллельной обработки заявок. Часть II // Вестник Российского университета дружбы народов. Серия: Математика. Информатика. Физика. – 2018. – Т. 26, № 1. – С. 13–27.
45. Flatto L., Hahn S. Two Parallel Queues Created by Arrivals with Two Demands I // SIAM Journal on Applied Mathematics. – 1984. – Vol. 44, № 5. – P. 1041–1053.
46. Narahari Y., Sundarrajan P. Performability Analysis of Fork-join Queueing Systems // Journal of the Operational Research Society. – 1995. – Vol. 46, № 10. – P. 1237–1249.

47. Ko S.-S., Serfozo R. F. Response times in $M/M/s$ fork-join networks // *Advances in Applied Probability*. – 2004. – Vol. 36, № 3. – P. 854–871.
48. Nelson R., Tantawi A. N. Approximate analysis of fork/join synchronization in parallel queues // *IEEE Transactions on Computers*. – 1988. – Vol. 37, № 6. – P. 739–743.
49. Rizk A., Poloczek F., Ciucu F. Computable Bounds in Fork-Join Queueing Systems // *ACM SIGMETRICS Performance Evaluation Review*. – New York, 2015. – Vol. 43, № 1. – P. 335–346.
50. Pinotsi D., Zazanis M. A. Synchronized queues with deterministic arrivals // *Operations Research Letters*. – 2005. – Vol. 33, № 6. – P. 560–566.
51. Varki E., Dowdy L. W. Analysis of balanced fork-join queueing networks // *ACM SIGMETRICS Performance Evaluation Review*. – 1996. – Vol. 24, № 1. – P. 232–241.
52. Varki E. Mean value technique for closed fork-join networks // *ACM SIGMETRICS Performance Evaluation Review*. – 1999. – Vol. 27, № 1. – P. 103–112.
53. Kemper B., Mandjes M. Mean sojourn times in two-queue fork-join systems: bounds and approximations // *OR Spectrum*. – 2012. – Vol. 34, № 3. – P. 723–742.
54. Fidler M., Walker B., Jiang Y. Non-Asymptotic Delay Bounds for Multi-Server Systems with Synchronization Constraints // *IEEE Transactions on Parallel and Distributed Systems*. – 2018. – Vol. 29, № 7. – P. 1545–1559.
55. Marin A., Rossi S., Sottana M. Biased Processor Sharing in Fork-Join Queues // *Quantitative Evaluation of Systems* / ed. by A. McIver, A. Horvath. – Cham : Springer International Publishing, 2018. – P. 273–288.
56. Baccelli F., Massey W. A., Towsley D. Acyclic fork-join queueing networks // *Journal of the ACM*. – 1989. – Vol. 36, № 3. – P. 615–642.

57. Baccelli F., Makowski A. M. Queueing models for systems with synchronization constraints // Proceedings of the IEEE. – 1989. – Vol. 77, № 1. – P. 138–161.
58. Baccelli F., Makowski A. M., Shwartz A. The fork-join queue and related systems with synchronization constraints: stochastic ordering and computable bounds // Advances in Applied Probability. – 1989. – Vol. 21, № 3. – P. 629–660.
59. Ko S.-S. Cycle Times in a Serial Fork-Join Network // Computational Science and Its Applications – ICCSA 2007 / ed. by O. Gervasi, M. L. Gavrilova. – Berlin, Heidelberg : Springer, 2007. – P. 758–766.
60. Green L. A Queueing System in Which Customers Require a Random Number of Servers // Operations Research. – 1980. – Vol. 28, № 6. – P. 1335–1346.
61. Rumyantsev A., Morozov E. Stability criterion of a multiserver model with simultaneous service // Annals of Operations Research. – 2015. – Vol. 252, № 1. – P. 29–39.
62. Omahen K., Schrage L. A queueing analysis of a multiprocessor system with shared memory // Proceedings of the Symposium on Computer Communication Networks and Teletraffic. – 1972. – P. 77–88.
63. Kumar A., Shorey R. Performance Analysis and Scheduling of Stochastic Fork-Join Jobs in a Multicomputer System // IEEE Transactions on Parallel and Distributed Systems. – 1993. – Vol. 10, № 4. – P. 1147–1164.
64. Аппроксимация времени отклика системы облачных вычислений / А. В. Горбунова [и др.] // Информатика и её применения. – 2015. – Т. 9, вып. 3. – С. 31–38.
65. Generalized parallel-server fork-join queues with dynamic task scheduling / M. S. Squillante [et al.] // Annals of Operations Research. – New York, 2008. – Vol. 160, № 1. – P. 227–255.

66. Вышенский С. В., Григорьев П. В., Дубенская Ю. Ю. Идеальный синхронизатор маркированных пар в сети разветвление-объединение // Обозрение прикладной и промышленной математики. – 2008. – Т. 15, № 3. – С. 385–399.
67. Baccelli F., Makowski A. Simple computable bounds for the fork-join queue : tech. rep. / INRIA. – 1985.
68. Flatto L. Two Parallel Queues Created by Arrivals with Two Demands II // SIAM Journal on Applied Mathematics. – 1985. – Vol. 45, № 5. – P. 861–878.
69. Thomasian A., Tantawi A. N. Approximate Solutions for $M/G/1$ Fork/Join Synchronization // Proceedings of the 26th Conference on Winter Simulation. – Orlando : Society for Computer Simulation International, 1994. – P. 361–368.
70. Varma S., Makowski A. M. Interpolation approximations for symmetric Fork-Join queues // Performance Evaluation. – 1994. – Vol. 20, № 1–3. – P. 245–265.
71. Kim C., Agrawala A. K. Analysis of the fork-join queue // IEEE Transactions on Computers. – 1989. – Vol. 38, № 2. – P. 250–255.
72. Knessl C. On the Diffusion Approximation to a Fork and Join Queueing Model // SIAM Journal on Applied Mathematics. – 1991. – Vol. 51, № 1. – P. 160–171.
73. Tan X., Knessl C. A fork-join queueing model: Diffusion approximation, integral representations and asymptotics // Queueing Systems. – 1996. – Vol. 22, № 3/4. – P. 287–322.
74. Determinism minimizes waiting time in queues : tech. rep. / P. A. Humblet [et al.]. – 1982. – 6 p.
75. Kleinrock L. Queueing systems, vol. 1. – New York : Wiley, 1975.

76. Stewart W. J. Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling. – Princeton : Princeton University Press, 2009. – 776 p.
77. Liu Y., Perros H. Approximate analysis of a closed fork/join model // European Journal of Operational Research. – 1991. – Vol. 53, № 3. – P. 382–392.
78. Liu Y., Perros H. A decomposition procedure for the analysis of a closed fork/join queueing system // IEEE Transactions on Computers. – 1991. – Vol. 40, № 3. – P. 365–370.
79. El-Taha M., Jafar M. J. Characterization of the departure process in a closed fork–join synchronization network // Applied Mathematics and Computation. – 2006. – Vol. 181, № 1. – P. 214–219.
80. Osman R., Harrison P. G. Approximating closed fork-join queueing networks using product-form stochastic Petri-nets // Journal of Systems and Software. – 2015. – Vol. 110. – P. 264–278.
81. Вишнеvский В. М. Теоретические основы проектирования компьютерных сетей. – Москва : Техносфера, 2003. – 512 с.
82. Duda A., Czachórski T. Performance evaluation of fork and join synchronization primitives // Acta Informatica. – 1987. – Т. 24, № 5. – С. 525–553.
83. Balsamo S., Donatiello L., Dijk N. V. Bound performance models of heterogeneous parallel processing systems // IEEE Transactions on Parallel and Distributed Systems. – 1998. – Vol. 9, № 10. – P. 1041–1056.
84. Осипов О. А. Система обслуживания с делением и слиянием требований, в которой требование занимает все свободные обслуживающие приборы // Вестник Российского университета дружбы народов. Серия: Математика. Информатика. Физика. – 2018. – Т. 26, № 1. – С. 28–38.

85. Осипов О. А. Анализ RQ -сети массового обслуживания с делением и слиянием требований // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2018. – № 43. – С. 49–55.
86. Qiu Z., Pérez J. F., Harrison P. G. Beyond the mean in fork-join queues: Efficient approximation for response-time tails // Performance Evaluation. – 2015. – Vol. 91. – P. 99–116. – Special Issue: Performance 2015.
87. Cochran J. J., Cox L. A., Kharoufeh J. P. Wiley encyclopedia of operations research and management science. – Hoboken : Wiley, 2011. – 2012 p.
88. Lu H., Pang G. Heavy-traffic limits for an infinite-server fork-join queueing system with dependent and disruptive services // Queueing Systems. – 2017. – Vol. 85, № 1. – P. 67–115.
89. Tsimashenka I., Knottenbelt W. J. Reduction of Subtask Dispersion in Fork-Join Systems // Computer Performance Engineering / ed. by M. S. Balsamo, W. J. Knottenbelt, A. Marin. – Berlin, Heidelberg : Springer, 2013. – P. 325–336.
90. Tsimashenka I., Knottenbelt W. J., Harrison P. G. Controlling variability in split-merge systems and its impact on performance // Annals of Operations Research. – 2016. – Vol. 239, № 2. – P. 569–588.
91. Atar R., Mandelbaum A., Zviran A. Control of Fork-Join Networks in heavy traffic // 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton). – 2012. – P. 823–830.
92. Ward A. On the Control of Fork-Join Networks. – 2015. – URL: <https://open.library.ubc.ca/cIRcle/collections/48630/items/1.0221662> (access date: 02/22/2017).
93. Dallery Y., Liu Z., Towsley D. Equivalence, Reversibility, Symmetry and Concavity Properties in Fork-join Queueing Networks with Blocking // J. ACM. – New York, 1994. – Vol. 41, № 5. – P. 903–942.

94. A Necessary and Sufficient Condition for Throughput Scalability of Fork and Join Networks with Blocking / Y. Zeng [et al.] // SIGMETRICS Perform. Eval. Rev. – New York, 2016. – Vol. 44, № 1. – P. 25–36.
95. Lebrecht A. S., Dingle N. J., Knottenbelt W. J. Modelling Zoned RAID Systems Using Fork-Join Queueing Simulation // Computer Performance Engineering / ed. by J. T. Bradley. – Berlin, Heidelberg : Springer Berlin Heidelberg, 2009. – P. 16–29.
96. Joshi G., Liu Y., Soljanin E. On the Delay-Storage Trade-Off in Content Download from Coded Distributed Storage Systems // IEEE Journal on Selected Areas in Communications. – 2014. – Vol. 32, № 5. – P. 989–997.
97. Heidelberger P., Trivedi K. S. Analytic Queueing Models for Programs with Internal Concurrency // IEEE Transactions on Computers. – 1983. – Vol. C–32, № 1. – P. 73–82.
98. Cremonesi P., Turrin R. Performance models for hierarchical grid architectures // 2006 7th IEEE/ACM International Conference on Grid Computing. – Institute of Electrical, Electronics Engineers (IEEE), 2006.
99. Performance Prediction of Cloud-Based Big Data Applications / D. Ardagna [et al.] // Proceedings of the 2018 ACM/SPEC International Conference on Performance Engineering. – Berlin : ACM, 2018. – P. 192–199.
100. Осипов О. А. Построение модели системы распределённых вычислений в виде системы массового обслуживания с делением и слиянием требований // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем. – РУДН. 2017. – С. 135–136.
101. Anderson D. P. BOINC: A System for Public-Resource Computing and Storage // Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing. – Washington : IEEE Computer Society, 2004. – P. 4–10.

102. Kurochkin I., Saevskiy A. BOINC Forks, Issues and Directions of Development // *Procedia Computer Science*. – 2016. – Vol. 101. – P. 369–378.
103. Serfozo R. *Basics of Applied Stochastic Processes*. – Berlin : Springer Berlin Heidelberg, 2009. – 443 p.
104. Boucherie R. J., Van Dijk N. M. *Queueing networks: a fundamental approach*. Vol. 154. – New York : Springer Science & Business Media, 2010. – 824 p.
105. Osogami T., Harchol-Balter M. Closed Form Solutions for Mapping General Distributions to Quasi-minimal PH Distributions // *Perform. Eval.* – Amsterdam, 2006. – Vol. 63, № 6. – P. 524–552.
106. Boxma O., Daduna H. Sojourn times in queueing networks // *Stochastic Analysis of Computer and Communication Systems* / ed. by H. Takagi. – Netherlands : North-Holland Publishing Company, 1990. – P. 401–450.
107. Harrison P. G. Approximate analysis and prediction of time delay distributions in networks of queues // *Int. CMG Conference*. – 1981. – P. 70–80.
108. Woolet S. P. *Performance Analysis of Computer Networks* : PhD thesis / Woolet Steven P. – Durham, 1993.
109. He Q.-M. *Fundamentals of Matrix-Analytic Methods*. – New York : Springer, 2014. – 349 p.
110. Осипов О. А. Исследование сетей массового обслуживания с делением и слиянием требований // *Международная научная конференция студентов, аспирантов и молодых учёных Ломоносов-2015*. – МАКС Пресс. 2015.
111. Осипов О. А., Тананко И. Е. Сети массового обслуживания произвольной топологии с делением и слиянием требований: случай бесконечноприборных систем обслуживания // *Вестник Тверского государственного университета. Серия: Прикладная математика*. – 2017. – № 4. – С. 43–58.

112. Latouche G., Ramaswami V. Introduction to Matrix Analytic Methods in Stochastic Modeling. – Philadelphia : Society for Industrial & Applied Mathematics (SIAM), 1999. – 349 p.
113. Neuts M. Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. – Baltimore : The Johns Hopkins University Press, 1981.
114. Lipsky L. Queueing Theory : A Linear Algebraic Approach. – New York : Springer, 2009. – 548 p.
115. Башарин Г. П., Бочаров П. П., Коган Я. А. Анализ очередей в вычислительных сетях. Теория и методы расчёта. – Москва : Наука, 1989. – 336 с.
116. Осипов О. А. О сети массового обслуживания с делением и слиянием требований с ограничением на фрагментацию // Шестнадцатый Всероссийский симпозиум по прикладной и промышленной математике (осенняя сессия). Т. 22. – Обзорение прикладной и промышленной математики. 2015. – С. 490–491.
117. MultiPath TCP — Linux Kernel implementation. – URL: <https://www.multipath-tcp.org/> (дата обращения: 22.03.2018).
118. How Hard Can It Be? Designing and Implementing a Deployable Multipath TCP / C. Raiciu [et al.] // 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12). – San Jose : USENIX Association, 2012. – P. 399–412.
119. Осипов О. А., Тананко И. Е. Моделирование сетей передачи данных с многопутевой маршрутизацией сетями массового обслуживания с делением и слиянием требований // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем. – РУДН. 2016. – С. 110–112.

120. Анализ сетей массового обслуживания с разделением и слиянием требований и управлением потоками : 2014662807 / О. А. Осипов ; правообладатель: Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Саратовский государственный университет имени Н.Г. Чернышевского (RU)». – Заявл. 2014.
121. Тананко И. Е., Осипов О. А. Методология имитационного моделирования открытых сетей массового обслуживания произвольной топологии с делением и слиянием требований // Компьютерные науки и информационные технологии. Материалы Международной научной конференции. – 2016. – С. 408–411.
122. Law A. M., Kelton W. D., Kelton W. D. Simulation modeling and analysis. – New York : McGraw-Hill, 2015. – 804 p.
123. MathNet.Numerics. – URL: <https://www.nuget.org/packages/MathNet.Numerics/> (access date: 03/22/2018).