

Министерство образования и науки Российской Федерации

Санкт-Петербургский политехнический
Университет Петра Великого

А.Г. Сиднеев, С.С. Сабонис, В.Н. Цыган

**Массовое обслуживание для
исследования и оптимизации
систем**

Учебное пособие

Издание второе, исправленное и дополненное

Санкт-Петербург
2021

УДК 303.732

Сиднев А. Г. **Массовое обслуживание для исследования и оптимизации систем** : Учебное пособие, изд. 2-е, испр. и дополненное / А.Г. Сиднев, С.С. Сабонис, В.Н. Цыган. — СПб.: Санкт-Петербургский политехнический университет, 2021. – 201 с.

В пособии рассматриваются основные принципы и подходы к построению формальных моделей систем с использованием теории массового обслуживания, а также аналитические и численные методы анализа их характеристик. Уделено внимание решению задач оптимизации на основе моделей массового обслуживания. Пособие ориентировано как на изучение теоретической части в рамках курсов «Системный анализ и принятие решений» и «Методы оптимизации», так и на практическое использование при проведении упражнений.

Пособие может быть рекомендовано студентам института компьютерных наук и технологий, обучающимся по направлению «Информатика и вычислительная техника», а также студентам других направлений при изучении курсов, связанных с применением методов теории массового обслуживания для моделирования систем.

Ключевые слова и выражения: системы массового обслуживания (СМО), сети СМО, модели массового обслуживания, аналитические расчеты, численные методы.

© Сиднев А.Г., Сабонис С.С., Цыган В.Н., 2021

© Санкт-Петербургский политехнический
университет Петра Великого, 2021

ОГЛАВЛЕНИЕ

Предисловие ко второму изданию	6
Введение к первому изданию	7
1. Теория массового обслуживания. Понятия и определения.....	10
1.1. Основные обозначения.....	10
1.2. Классификация систем массового обслуживания	11
1.3. Потоки событий и их свойства	14
1.3.1. Классификация потоков	14
1.3.2. Простейший поток и его свойства	17
1.3.3. Потоки Пальма. Рекуррентные потоки.....	21
1.4. Общие результаты теории СМО	23
1.5. Примеры формализации задач в терминах СМО	25
Задачи	28
2. Марковские системы массового обслуживания	31
2.1. Марковские случайные процессы и их классификация	31
2.2. Дискретные марковские последовательности.....	32
2.3. Общие уравнения непрерывного марковского процесса	36
2.4. Агрегирование состояний марковского процесса	40
2.5. Процесс гибели и размножения	50
2.6. Основные соотношения для простейших СМО	54
3. Немарковские системы массового обслуживания	64
3.1. Классификация методов исследования немарковских СМО ...	64
3.1.1. Метод фаз	64
3.1.2. Использование фазовых аппроксимаций законов распределения интервалов поступления и обслуживания заявок для расчета СМО типа G/G/m	66

3.1.3. Метод введения избыточной переменной	70
3.1.4. Метод вложенных марковских цепей	71
3.1.5. Интегральный метод	72
3.2. Характеристики простейших СМО $M/D/1$, $M/E_K/1$, $M/G/1$	74
4. Стохастические сетевые модели СМО	78
4.1. Предпосылки и цели применения аппарата сетей СМО	78
4.2. Однородные экспоненциальные сети СМО	82
4.2.1. Разомкнутые сети СМО. Анализ их характеристик.....	82
4.2.2. Замкнутые сети СМО. Анализ их характеристик	90
4.2.3. Определение основных показателей качества обслуживания в узлах замкнутой сети СМО	98
4.2.4. Расчет нормирующей константы однородной замкнутой сети СМО	101
4.2.5. Расчетные соотношения, используемые для определения характеристик узлов однородной замкнутой сети СМО	106
4.2.6. Метод анализа средних значений характеристик узлов однородной замкнутой сети СМО	107
4.2.7. Приближенный расчет замкнутых сетей с использованием метода баланса заявок	110
4.3. Неоднородные сети СМО	114
4.3.1. Общие сведения о неоднородных сетях СМО	114
4.3.2. Разомкнутые неоднородные сети СМО	116
4.3.3. Замкнутые неоднородные сети СМО	118
4.4. Приближенные методы расчета немарковских сетей СМО ...	127
4.4.1. Расчет показателей разомкнутых однородных однока-	

нальных немарковских сетей СМО	129
4.4.2. Расчет показателей разомкнутых однородных много- канальных немарковских сетей СМО	131
4.4.3. Расчет показателей разомкнутых неоднородных много- канальных немарковских сетей СМО	132
Задачи	135
5. Оптимизация сетей СМО	145
5.1. Оптимизация разомкнутых сетей СМО	145
5.1.1. Постановка задачи оптимизации разомкнутой сети СМО	145
5.1.2. Классификация задач оптимизации разомкнутых сетей СМО.....	150
5.1.3. Оптимизация сетей Джексона.....	151
5.1.3.1. Точные методы решения задач SP.1.1/J/S/R и SP.2.1/J/S/R	151
5.1.3.2. Алгоритмы поиска оптимальных решений задач SP.1.1/J/M/N и SP.2.1/J/M/N.....	153
5.1.4. Оптимизация разомкнутых немарковских сетей СМО	157
5.1.4.1. Оптимизация разомкнутых однородных немар- ковских сетей СМО	157
5.1.4.2. Оптимизация разомкнутых неоднородных не- марковских сетей СМО	162
5.2. Оптимизация замкнутых сетей СМО	170
Задачи	182
Приложение. Построение матриц A_i , F_i , B_i , D_i для марковского графа состояний.....	196
Библиографический список.....	200

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

Первое издание данного пособия вышло в 2017 году и хотя оно по-прежнему соответствует программе обучения по дисциплине «Системный анализ и принятие решений», все же к настоящему времени стала ощущаться потребность в некоторой переработке пособия и его расширении. При работе над вторым изданием авторы только первый раздел пособия оставили почти без изменений, все остальные разделы ощутимо переработаны. Так, например, во второй раздел добавлены новый подраздел «Агрегирование состояний марковского процесса» и несколько примеров, в третий раздел включен подраздел «Использование фазовых аппроксимаций законов распределения интервалов поступления и обслуживания заявок для расчета СМО типа G/G/m». Расширено содержание четвертого и пятого разделов за счет включения подразделов, посвященных новым методам и алгоритмам расчета и оптимизации сетей СМО. Написание новых подразделов выполнено преимущественно А. Г. Сидневым, редактирование текста осуществлял В. Н. Цыган, а новый соавтор – С. С. Сабонис внес важный вклад в работу над пособием путем тщательного экспериментального испытания обсуждаемых в пособии алгоритмов расчета сетевых моделей. При работе над вторым изданием авторы постарались исправить опечатки и некоторые ошибки, допущенные в тексте первого издания.

Пользуясь случаем, авторы выражают свою признательность коллегам, высказавшим свои замечания и предложения по содержанию пособия. Также авторы благодарны студентам, оказавшим помощь в подготовке нового издания.

ВВЕДЕНИЕ К ПЕРВОМУ ИЗДАНИЮ

Модели и методы теории массового обслуживания в настоящее время приобрели весьма широкое распространение во многих прикладных областях. Круг практических задач, решаемых методами теории массового обслуживания, непрерывно расширяется. Если раньше в качестве основных областей применения этой теории указывались задачи телефонии, бытового обслуживания, торговой сети, здравоохранения и т. д., то в последнее время она находит применение в таких областях, как:

организация и автоматизация производства и транспорта;

военное дело;

обработка информации в сложных управляющих и вычислительных системах;

совершенствование аппаратного и программного обеспечения вычислительных систем и сетей;

управление обменом в компьютерных сетях и др.

Предлагаемое читателю пособие непосредственно связано с лекционным курсом “Системный анализ и принятие решений”, который читается доцентом кафедры компьютерных систем и программных технологий СПбПУ А. Г. Сидневым для студентов Института компьютерных наук и технологий. Авторы считают своим долгом еще раз вспомнить самым добрым словом ныне покойного профессора Дмитрия Николаевича Колесникова, который длительное время был основным лектором учебных курсов, посвященных системному анализу и оптимизации. При его непосредственном участии и под его редакцией издавались учебные пособия.

Настоящее издание подготовлено на основе предыдущих путем их переработки и существенных дополнений. Данное учебное пособие не

ставит целью дать абсолютно полное и строгое изложение всех вопросов, которые сегодня принято относить к теории массового обслуживания.

Главное внимание уделено той части вопросов, которые позволяют ознакомить студентов с аппаратом теории массового обслуживания и наиболее распространенными способами его практического применения.

В пособии изложены формулировки основных задач, вывод соответствующих систем дифференциальных или алгебраических уравнений, получение главных формул в результате решения этих уравнений, а также приведены примеры, иллюстрирующие использование теории массового обслуживания для решения прикладных задач.

Особенностью данного пособия является стремление авторов проиллюстрировать принципы формализации и математического описания процессов поведения систем различного назначения в терминах и понятиях массового обслуживания, а также развить навыки в использовании аналитического аппарата для оценки показателей качества обслуживания.

Основное отличие настоящего издания от предыдущих состоит в том, что в нем существенное внимание уделено вопросам оптимизации систем, представленных в виде сетевых моделей массового обслуживания. В соответствии с этим определено и содержание пособия: в начальных разделах излагается аппарат теории массового обслуживания, который затем используется при решении задач оптимизации.

Первый раздел посвящен основным понятиям и определениям, а также классификации систем массового обслуживания.

Во втором разделе рассмотрено применение случайных марковских процессов для анализа показателей качества систем массового обслуживания.

Третий раздел посвящен немарковским системам массового обслуживания и сравнительному анализу методов их исследования.

В четвертом разделе представлены основные результаты практического применения теории сетей массового обслуживания в решении прикладных задач.

Пятый раздел посвящен оптимизации сетей массового обслуживания. Подробно рассмотрены вопросы оптимизации как разомкнутых, так и замкнутых сетевых моделей. Проанализированы возможности использования аналитических и численных методов для решения оптимизационных задач.

Работа над пособием распределялась следующим образом:

Разд. 1—3 представляют собой материалы, которые были опубликованы в ранее изданных пособиях, но для данного издания заново отредактированы А. Г. Сидневым;

разд. 4 заново отредактирован А. Г. Сидневым и В. Н. Цыганом, а также дополнен новым подразд. 4.4 (А. Г. Сиднев);

разд. 5 написан А. Г. Сидневым (подразд. 5.1 и 5.3) и В. Н. Цыганом (подразд. 5.2).

Авторы выражают благодарность студентам, оказавшим содействие в подготовке материалов пособия: Е. Д. Хандыго, Н. В. Тышковцу, А. А. Цыганову, А. М. Васильеву.

Данное учебное пособие является продолжением и развитием учебных пособий [15, 17, 18], содержит много примеров и задач на развитие навыков формализации процессов функционирования систем и их оптимизации на основе применения теории массового обслуживания.

1. ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ.

ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

1.1. Основные обозначения

λ — интенсивность потока

μ — интенсивность обслуживания

$\rho = \lambda/\mu$ — коэффициент загрузки в системе **G/G/1**

x — длительность интервала обслуживания (в общем случае случайная величина)

$B(t) = P(x \leq t)$ — интегральный закон распределения длительности обслуживания

$b(t)$ — плотность распределения длительности обслуживания

$v^{(1)} = \bar{x}$, $v^{(2)} = \bar{x^2}$ — первый и второй начальные моменты длительности обслуживания

$A(t) = P(y \leq t)$ — интегральный закон распределения длительности интервалов между требованиями во входном потоке

$a(t)$ — плотность распределения длительности интервалов между требованиями

\bar{y} — средняя длительность интервалов между требованиями

K — число каналов обслуживания в системе

$\rho_c = \frac{\lambda}{K\mu}$ — суммарный коэффициент загрузки в системе **G/G/K**

m — емкость накопителя

n — число требований в конечном источнике

\bar{j} — среднее число требований в системе

\bar{n}_o — среднее число требований в очереди

P_j — вероятность того, что в системе находится j требований

($j = 0, 1, 2, \dots$)

P_0 — вероятность простого системы

\bar{t}_c — среднее время пребывания требования в системе

σ_j^2 — дисперсия числа требований в системе

$\bar{t}_{ож}$ — среднее время ожидания требования в очереди

\bar{K}_3 — среднее число занятых каналов в системе

η_3 — коэффициент загрузки системы

$P_{отк}$ — вероятность отказа в обслуживании

$P(n_o > k)$ — вероятность того, что число требований в очереди больше k

$P(t_{ож} < t)$ — вероятность того, что время ожидания в очереди меньше t

$\Delta = \{\lambda_{ij}\}$ — матрица интенсивностей передач на графе состояний

$P = \{P_{ij}\}$ — матрица вероятностей передач на графе состояний

1.2. Классификация систем массового обслуживания

Модели функционирования объектов, представленных в виде систем массового обслуживания (СМО), имеют структуру, показанную на рис. 1.1.

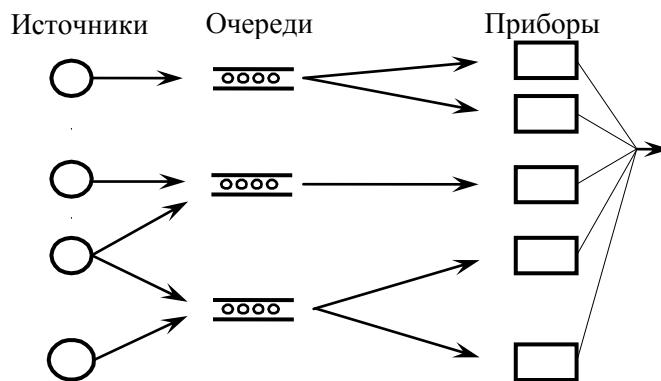


Рис. 1.1

Введем основные определения.

Последовательность событий будем называть потоком. Поток, состоящий из требований на обслуживание, назовем потоком требований.

Первопричину возникновения требований независимо от их физической природы будем называть источником.

Совокупность очередей и приборов обслуживания будем называть обслуживающей системой.

Алгоритм взаимодействия обслуживающих приборов с очередью назовем дисциплиной обслуживания.

Рассмотрим основные классификационные категории СМО.

По характеру источника требований различают источники с бесконечным и конечным числом требований.

По отсутствию или наличию возможностей ожидания для окончательного завершения обслуживания требований в системе выделяют: *системы с потерями* (по причине ограниченной емкости буферного накопителя или ограничения времени ожидания в очереди) и *системы без потерь*.

По числу приборов (каналов) различают: *одноканальные и многоканальные обслуживающие системы*.

По числу этапов (фаз) обслуживания в системе различают *однофазные и многофазные системы обслуживания*.

По правилу формирования очереди различают *системы массового обслуживания с общей очередью* и *системы массового обслуживания с несколькими очередями*.

По дисциплине обслуживания различают *системы массового обслуживания: с бесприоритетными дисциплинами и приоритетными* (относительный или абсолютный приоритет). При относительном приоритете

требование занимает обслуживающий прибор без прерывания фазы обслуживания, а при абсолютном — с прерыванием.

Представленная классификация СМО является далеко не полной, но достаточной для рассмотрения моделей, используемых в данном учебном пособии.

Для изложения дальнейшего содержания пособия целесообразно воспользоваться принятой в теории массового обслуживания системой обозначений [1].

В их основе лежит трехбуквенное обозначение вида **A/B/K**, где **A** и **B** описывают соответственно распределение промежутков времени между требованиями и распределение времени их обслуживания, а **K** — число обслуживающих приборов.

A и **B** принимают значения из следующего набора символов, интерпретация которых дается распределениями, указанными в круглых скобках: **M** (показательное или иначе экспоненциальное), **E_r** (распределение Эрланга порядка **r**), **D** (детерминированное), **G** (распределение общего вида). Иногда приходится указывать также емкость накопителя системы, которую обозначим **m**, и число источников нагрузки — **n**. В этом случае будем использовать пятибуквенное обозначение: **A/B/K/m/n**.

При отсутствии одного из двух последних индексов предполагается, что его значение сколь угодно велико, т. е. равно бесконечности.

В дальнейшем эта терминология будет использована при формализации и анализе технических характеристик систем, представляемых в виде СМО.

1.3. Потоки событий и их свойства

1.3.1. Классификация потоков

Поток — последовательность событий, происходящих одно за другим в случайные или детерминированные моменты времени. Физическая природа потоков может быть разнообразной: поступление заявок на решение задачи, выдача результата, посадка или взлет самолета, отказ аппаратуры, поступление вызовов на автоматические телефонные станции и пр.

Различают потоки:

однородные, в которых события характеризуются только моментом времени появления;

неоднородные, в которых события дополняются другими качественными и количественными характеристиками, такими как вид и уровень приоритета, время обработки и т. п.

В дальнейшем, пока будем рассматривать однородные потоки.

С потоком связаны две группы характеристик случайных величин:

плотность распределения длительности интервалов $f(\tau_j)$:

$$\tau_j = t_{j+1} - t_j, \quad j = 0, 1, 2, \dots;$$

закон $P_n(t, T)$ распределения числа событий $n(t, T)$ на интервале T , отсчитываемом от момента t , где $n = 0, 1, 2, \dots$

В первом случае — это последовательность непрерывных случайных величин; во втором — дискретных случайных величин. По характеру величины τ_j выделяют потоки:

случайные, если величины τ_j случайные;

детерминированные, если величина τ_j постоянная, т. е. $\tau_j = \tau_0$;

случайные с дискретным временем: $\tau_j = k_j T_0$, где T_0 — шаг дискретного времени; k_1, k_2, \dots, k_j — целочисленные случайные величины; регулярные, в которых события появляются обязательно в точках $\tau_j = jT_0$.

Определим основные модели распределений:

$$1) \quad f(\tau) = \lambda e^{-\lambda\tau}, \quad \tau \geq 0, \quad M — \text{показательное (Markovian);}$$

$$2) \quad f_r(\tau) = \frac{r\lambda(r\lambda\tau)^{r-1}}{(r-1)!} e^{-r\lambda\tau}, \quad \tau \geq 0, \quad E_r — \text{эрланговское (Erlangian);}$$

$$3) \quad f_r(\tau) = \sum_{i=1}^r \alpha_i \lambda_i e^{-\lambda_i \tau}, \quad \sum_{i=1}^r \alpha_i = 1, \quad H_r — \text{гиперэкспоненциальное (Hyperexponential);}$$

$$4) \quad f(\tau) = \delta(\tau - \tau_0), \quad \tau_j = \tau_0, \quad D — \text{постоянное (Deterministic);}$$

$$5) \quad f(\tau), \quad G — \text{произвольное (General).}$$

Статистические свойства случайных величин $n(t, T)$ определяют важнейшие свойства потоков.

1. Стационарность: $P_n(t, T) = P_n(T)$, $n = 0, 1, 2, \dots$

(распределение не зависит от положения интервала T на оси времени и зависит только от длительности T).

2. Отсутствие последействия:

$$P_{n_1 n_2 \dots n_i}(T_1 T_2, \dots, T_i) = \prod_{k=1}^i P_{n_k}(T_k) \quad (1.1)$$

(независимость числа событий в неперекрывающихся интервалах T_i). Данное свойство с большим трудом поддается экспериментальной проверке.

3. Ординарность: вероятность появления более одного события на бесконечно малом интервале имеет порядок малости выше, чем вероятность появления одного события на этом интервале, т. е.

$$\lim_{\Delta t \rightarrow 0} \frac{P_n(\Delta t)}{P_1(\Delta t)} = 0, \quad n = 2, 3, \dots$$

или

$$\lim_{\Delta t \rightarrow 0} \frac{P_{n>1}(\Delta t)}{\Delta t} = 0, \quad n = 2, 3, \dots \quad (1.2)$$

Можно показать, что свойство ординарности следует из свойств стационарности и отсутствия последействия.

Поток, характеризующийся стационарностью, отсутствием последействия и ординарностью, называется простейшим потоком.

Стационарные потоки с независимыми интервалами, распределенными по произвольным законам $F_1(t), F_2(t), \dots, F_i(t)$, где $F_i(t) = P(\tau_i < t)$ есть функция распределения i -го интервала в потоке, называются потоками Пальма.

Если все интервалы распределены одинаково, то такие потоки называются рекуррентными.

Поток Пальма с экспоненциальным распределением длительности интервала в потоке является простейшим.

Из отсутствия последействия следует независимость интервалов, но независимость интервалов не означает автоматически отсутствия последействия.

1.3.2. Простейший поток и его свойства

1. Закон распределения длительности интервала выводится из свойств стационарности и отсутствия последействия:

$$\begin{aligned} f(\tau) &= \lambda e^{-\lambda \tau}, \quad \tau > 0; \\ P(\tau < t) &= F(t) = 1 - e^{-\lambda t}, \end{aligned} \tag{1.3}$$

где λ — параметр потока.

Закон распределения числа событий на интервале T :

$$P_n(T) = \frac{(\lambda T)^n}{n!} e^{-\lambda T}.$$

2. Определение параметра потока λ .

Сначала найдем среднюю интенсивность потока на интервале $[t, t + T]$:

$$\lambda(t, T) = \frac{M[n(t, T)]}{T}, \tag{1.4}$$

где $M[n(t, T)]$ — среднее число событий на интервале $[t, t + T]$.

Мгновенная интенсивность потока определяется так:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{M[n(t, \Delta t)]}{\Delta t}, \tag{1.5}$$

причем $\lambda(t, T) = \frac{1}{T} \int_t^{t+T} \lambda(x) dx.$

Тогда для стационарного потока

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{M[n(t, \Delta t)]}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{M[n(\Delta t)]}{\Delta t} = \lambda = \text{const}, \quad (1.6)$$

а среднее число событий на интервале T

$$M[n(T)] = \lambda T. \quad (1.7)$$

Тот же результат можно получить с учетом подчинения событий простейшего потока закону Пуассона:

$$\begin{aligned} M[n(T)] &= \sum_{n=0}^{\infty} n \frac{(\lambda T)^n}{n!} e^{-\lambda T} = \sum_{n=1}^{\infty} n \frac{(\lambda T)^n}{n!} e^{-\lambda T} \\ &= \lambda T \sum_{n=1}^{\infty} \frac{(\lambda T)^{n-1}}{(n-1)!} e^{-\lambda T} = \lambda T \sum_{r=0}^{\infty} \frac{(\lambda T)^r}{r!} e^{-\lambda T} = \lambda T e^{\lambda T} e^{-\lambda T} = \lambda T, \end{aligned}$$

$$\text{т. е. } \lambda = \frac{M[n(T)]}{T}.$$

Отсюда следует, что параметр потока равен интенсивности потока (средняя интенсивность равна мгновенной).

3. Законы распределения, экспоненциальный (показательный)

$F(t) = 1 - e^{-\lambda t}$ и Пуассона $P_n(T) = \frac{(\lambda T)^n}{n!} e^{-\lambda T}$, связаны между собой:

$$P_0(T) = e^{-\lambda T} = P(\tau > T), \text{ тогда } P(\tau < T) = 1 - e^{-\lambda T},$$

а это и есть функция распределения $F(t) = 1 - e^{-\lambda t}$.

4. Численные характеристики экспоненциального закона:

$$M[\tau] = \int_0^{\infty} \tau \lambda e^{-\lambda \tau} d\tau = \frac{1}{\lambda}; \quad (1.8)$$

$$D[\tau] = M[\tau^2] - (M[\tau])^2 = \frac{1}{\lambda^2}; \quad (1.9)$$

$$M[\tau] = \sigma[\tau]. \quad (1.10)$$

Соотношение (1.10) — основное для выдвижения гипотезы о простейшем потоке при обработке экспериментальных данных о наблюдаемых потоках.

5. Закон распределения остатка интервала T_1 между событиями в простейшем потоке не зависит от прошедшего времени T (рис. 1.2):

$$\begin{aligned} P(\tau < t) &= 1 - e^{-\lambda t}; \\ F(T_1 < t) &= P[T_1 \leq t / \tau \geq T] = \frac{P[T_1 \leq t, \tau \geq T]}{P[\tau \geq T]} = \\ &= \frac{P[T \leq \tau \leq T+t]}{P[\tau \geq T]} = \frac{1 - e^{-\lambda(T+t)} - 1 + e^{-\lambda T}}{e^{-\lambda T}} = 1 - e^{-\lambda t}. \end{aligned}$$

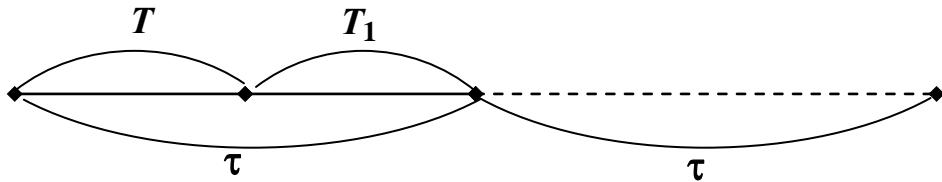


Рис. 1.2

Остаток интервала T_1 в простейшем потоке распределен так же, как и полный интервал τ . Этот результат, известный как парадокс остаточного времени [1], является наиболее полным выражением сущности свойства отсутствия последействия.

6. Закон распределения суммы интервалов $\tau = \sum_{i=1}^r \tau_i$, где каждый интервал имеет экспоненциальное распределение с параметром λ , равен

$$f_r(\tau) = \frac{\lambda(\lambda\tau)^{r-1}}{(r-1)!} e^{-\lambda\tau}, \quad \tau \geq 0,$$

а численные характеристики определяются как: $M[\tau] = \frac{r}{\lambda}$; $D[\tau] = \frac{r}{\lambda^2}$.

Данный закон описывает распределение между событиями в потоке Эрланга r -го порядка.

7. Преобразования простейшего потока (рис. 1.3–1.5).

A. Регулярное прореживание с параметром r :

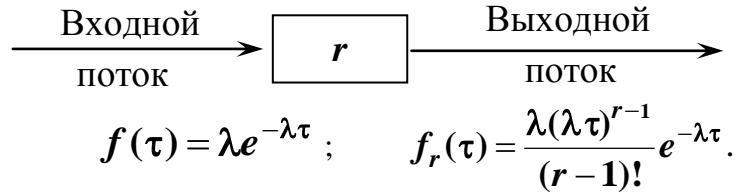


Рис. 1.3

B. Случайное просеивание с параметром p :

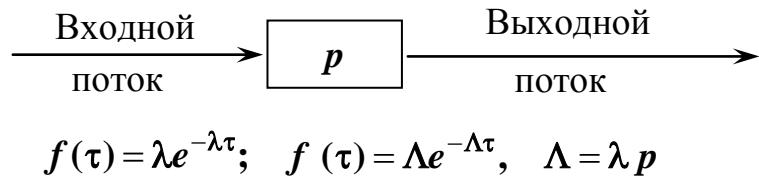


Рис. 1.4

B. Суммирование потоков:

$$f(\tau_i) = \lambda_i e^{-\lambda_i \tau_i} \quad \left| \begin{array}{l} \lambda_1 \rightarrow \\ \lambda_2 \rightarrow \\ \vdots \\ \lambda_i \rightarrow \\ \vdots \\ \lambda_n \rightarrow \end{array} \right. \quad \Lambda = \sum_i^n \lambda_i; \quad f(\tau) = \Lambda e^{-\Lambda\tau}.$$

Рис. 1.5

8. Моделирование влияния последействия в потоке.

Поток Эрланга имеет характеристики:

$$f_r(\tau) = \frac{\lambda(\lambda\tau)^{r-1}}{(r-1)!} e^{-\lambda\tau}; M[\tau] = \frac{r}{\lambda}, \quad D[\tau] = \frac{r}{\lambda^2}.$$

Пусть $M[\tau] = \frac{r}{\lambda} = \text{const} = c_0$. Если $\lambda = \frac{r}{c_0}$, то $D[\tau] = \frac{rc_0^2}{r^2} = \frac{c_0^2}{r}$.

Отсюда следует:

a) при $r \rightarrow \infty \quad D[\tau] = 0$;

b) при $r \rightarrow 1 \quad D[\tau] = c_0^2$.

В случае *a* справедливо максимальное последействие — приближающееся к регулярному потоку.

В случае *b* — отсутствие последействия, т. е. простейший поток.

1.3.3. Потоки Пальма. Рекуррентные потоки

Потоки Пальма — потоки с независимыми интервалами, распределенными по произвольным законам (рис. 1.6): $F_1(t), F_2(t), \dots, F_i(t), \dots$
Для рекуррентного потока справедливо $F_i(t) = F(t)$.

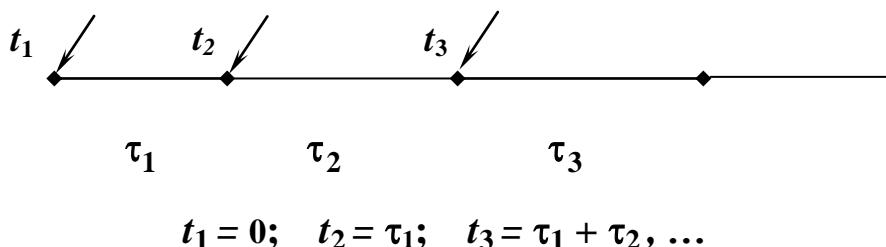


Рис. 1.6

Пусть $n(t)$ — число событий рекуррентного потока, появившихся до момента t , т. е. на интервале $[0, t]$. Математическое ожидание $M[n(t)]$ называется функцией восстановления $H(t)$:

$$H(t) = M[n(t)] = \sum_{n=1}^{\infty} n P[n(t) = n]. \quad (1.11)$$

Показано в [2], что: $H^*(s) = \frac{F^*(s)}{1 - F^*(s)}$, (1.12)

где $F^*(s)$ и $H^*(s)$ — преобразования Лапласа $F(t)$ и $H(t)$.

Для иллюстрации рассмотрим пример. Простейший поток:

$$F(t) = 1 - e^{-\lambda t}; \quad F^*(s) = \frac{\lambda}{\lambda + s}, \quad \text{т. е. } H^*(s) = \frac{\lambda}{s} \quad \text{и} \quad H(t) = \lambda t.$$

Формирование рекуррентного потока в задачах теории надежности поясняется на рис. 1.7.

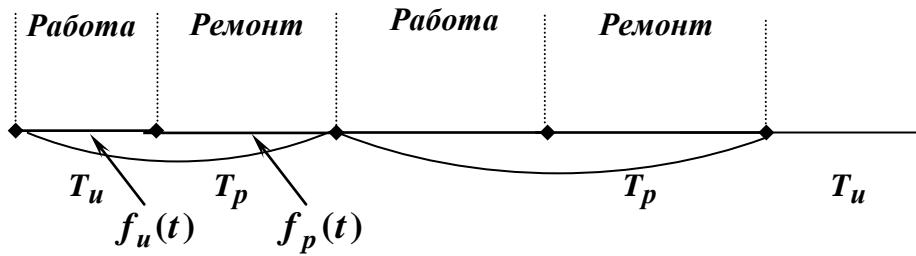


Рис. 1.7

$$T_u = T_u + T_p; \quad f_u(t) \rightarrow F_u(s); \quad f_p(t) \rightarrow F_p(s),$$

где T_u — длительность цикла.

Случайные величины T_u и T_p независимы, поэтому:

$$F_u^*(s) = F_u^*(s)F_p^*(s) \quad \text{и} \quad H_u^*(s) = \frac{F_u^*(s)F_p^*(s)}{1 - F_u^*(s)F_p^*(s)}, \quad (1.13)$$

где $H_u^*(s)$ — изображение по Лапласу скорости изменения среднего числа циклов “работа – ремонт”.

Как было показано ранее, простейший поток обладает свойством отсутствия последействия в любой момент, т. е. по всей оси времени. Рекур-

рентные потоки и вообще потоки Пальма, как показано в [2], обладают последействием всюду, кроме точек, в которых появляются события потока.

Время, произошедшее с момента появления предыдущего события, влияет на закон распределения интервала до ближайшего следующего события, но не влияет на закон распределения последующих полных интервалов. Область проявления последействия в потоках Пальма ограничена пределами одного интервала. Это объясняет происхождение термина потоки с ограниченным последействием, который иногда используется применительно к потокам Пальма.

1.4. Общие результаты теории СМО

В разделе приводятся результаты, справедливые для установившегося режима в СМО независимо от законов распределения длительности интервалов входящего потока, а также длительности обслуживания, т. е. справедливые для систем типа **G/G/1** и **G/G/K**.

Коэффициент использования для систем **G/G/1** задается как:

$$\rho = \lambda \bar{x} = \frac{\lambda}{\mu}, \quad (1.14)$$

т. е. произведение средней скорости λ поступления требований в систему на среднее время обслуживания $\bar{x} = 1/\mu$ каждого требования.

Для многоканальных систем типа **G/G/K** принято использовать следующий показатель загрузки:

$$\rho_c = \frac{\lambda \bar{x}}{K} = \frac{\lambda}{K\mu}. \quad (1.15)$$

Стабильной, или устойчивой системой (т. е. такой системой, которая имеет конечные средние задержки и длины очередей), является система, для которой:

$$\rho < 1 \text{ или } \rho_c < 1. \quad (1.16)$$

Чем ближе ρ или ρ_c к единице, тем больше очереди и время ожидания, тем более существенно отражается изменение средней нагрузки СМО на изменении ее характеристик.

Среднее время пребывания требования в обслуживающей системе определяется как:

$$\bar{t}_c = \bar{t}_{ож} + \bar{x}, \quad (1.17)$$

где \bar{x} — среднее время обслуживания; $\bar{t}_{ож}$ — среднее время ожидания в очереди.

Одной из универсальных формул теории массового обслуживания является закон Литтла [1]:

$$\bar{j} = \lambda \bar{t}_c, \quad (1.18)$$

т. е. среднее число требований в системе равно произведению средней скорости поступления требований на среднее время пребывания требования в системе для разомкнутых систем массового обслуживания.

Аналогично для среднего числа требований, ожидающих в очереди \bar{n}_0 , и среднего времени $\bar{t}_{ож}$ ожидания в очереди имеем:

$$\bar{n}_0 = \lambda \bar{t}_{ож}. \quad (1.19)$$

Интерпретация данных соотношений для конечного источника, т. е. для замкнутой системы выглядит следующим образом:

$$\begin{aligned}\bar{j} &= \lambda(n - \bar{j})\bar{t}_c; \\ \bar{n}_O &= \lambda(n - \bar{j})\bar{t}_{oж}.\end{aligned}\quad (1.20)$$

Для многоканальных систем с бесконечным источником имеем:

$$\bar{j} = \bar{n}_O + K\rho_c = \lambda\bar{t}_{oж} + \rho, \quad (1.21)$$

а для многоканальных систем с конечным источником:

$$\bar{j} = \bar{\lambda}(n - \bar{j})\bar{t}_{oж} + \bar{K}_3, \quad (1.22)$$

где \bar{K}_3 — среднее число занятых каналов в обслуживающей системе из K каналов.

1.5. Примеры формализации задач в терминах СМО

Пример 1.1

Рассматривается многопроцессорная система, составляющая вычислительный комплекс (ВК) обработки информации. Пусть число процессоров равно двум. На вход системы поступает поток задач. Если задача застает оба процессора занятыми, то она теряется. В процессе решения в результате сбоя может возникнуть ошибка. С целью повышения достоверности результатов счета рассматриваются два режима работы.

1. Использование временной избыточности: каждый процессор работает независимо от другого, т. е. решает свою задачу, и для повышения достоверности результата использует принцип голосования два из трех. На каждом из процессоров задача решается два раза. Если результаты совпали, то требование считается обслуженным; в противном случае, если возникла ошибка, задача решается третий раз. При совпадении двух результата

тов решения задачи из трех требование считается обслуженным, в противном случае — потерянным.

2. Использование структурной избыточности: каждую поступающую задачу оба процессора обрабатывают совместно, т. е. решают параллельно. Если результаты счета на обоих процессорах совпали, то требование считается обслуженным, в противном случае требование считается потерянным.

Провести сравнительный анализ двух режимов работы.

Рассмотрим формализацию данной задачи в терминах и понятиях СМО.

1. Источник требований — внешняя среда, генерирующая поток задач на обработку в ВК (источник — бесконечный).

2. Буфер для накопления поступающих задач (требований) в случае занятости процессоров отсутствует.

3. Система обработки информации, представленная ВК, включает: два канала, работающих автономно при временной избыточности; один канал, осуществляющий параллельную обработку одной задачи двумя процессорами — при структурной избыточности.

4. Дисциплина обслуживания в обоих случаях: первым пришел — первым обеспечен.

5. С учетом модели простейшего потока задач, поступающих в систему, а также показательного закона распределения длительности однократной обработки каждой задачи имеем:

при временной избыточности — модель СМО типа $M/E_2/2/0/\infty$;

при структурной избыточности — модель СМО типа $M/M/1/0/\infty$.

Для упрощения принято, что в случае временной избыточности длительность обработки каждой задачи в среднем в два раза превосходит длительность обработки при структурной избыточности.

6. Указанные модели СМО относятся к классу систем с потерями, и поэтому при сравнительном анализе качества обслуживания основным показателем целесообразно выбрать вероятность потерь P_{nom} .

Пример 1.2

Рассматривается автоматизированная система контроля (ACK), работающая в режиме параллельного выполнения n программ контроля. Каждая программа представляет собой последовательность чередующихся фаз работы процессора и специализированных внешних функциональных устройств (ФУ) (генераторов, преобразователей, коммутаторов, устройств ввода-вывода, регистраторов, внешних запоминающих устройств), управление которыми ведется от процессора. Средняя длительность фазы $\bar{\tau}_{\phi y}$ работы функциональных устройств значительно превосходит среднюю длительность фазы обработки команды в процессоре $\bar{\tau}_{np}$.

Обосновать целесообразный уровень совмещения при выполнении программ для условий.

1. Конфликты при обращении к функциональным устройствам не учитываются.
2. Дисциплина обслуживания программ процессором: первый ответил — первым обслужен.

Проведем формализацию данной задачи в терминах и понятиях СМО.

1. Источник требований — конечный и представляет собой совокупность n параллельно выполняемых программ.

2. Обслуживающий прибор — процессор.
3. Дисциплина обслуживания программ: первым пришел — первым обслужен.

4. С учетом модели показательных законов распределения длительностей фаз работы ФУ и процессора τ_{phy}, τ_{np} получим модель данной задачи в виде СМО типа $M/M/1/\infty/n$. Здесь ∞ означает, что физические ограничения для длины очереди в системе отсутствуют.

При оценке качества совмещения программ необходимо учесть как загрузку процессора (коэффициент загрузки обслуживающего прибора), так и задержки в выполнении программ (среднее время ожидания требований в очереди).

Задачи

Рассмотрите содержательную интерпретацию следующих задач в терминах и понятиях моделей СМО.

- 1.1. Обслуживание рабочим или бригадой рабочих парка станков на производстве.
- 1.2. Работа мастерской бытового обслуживания на вокзале.
- 1.3. Обслуживание разговоров на телефонной станции.
- 1.4. Работа пользователей за терминалами в дисплейном классе.
- 1.5. Управление взлетом — посадкой самолетов в аэропорту.
- 1.6. Обслуживание пациентов в поликлинике.
- 1.7. Погрузка и разгрузка судов в порту.
- 1.8. Работа производственного конвейера.
- 1.9. Обслуживание пассажиров в железнодорожных кассах.
- 1.10. Работа мультипрограммной и мультипроцессорной вычислительной системы.

- 1.11. Прохождение пассажиров через таможню в аэропорту.
- 1.12. Работа автотранспортной фирмы по перевозке грузов.
- 1.13. Обслуживание пассажиров на стоянках такси.
- 1.14. Комплектование штата секретарей-машинисток.
- 1.15. Работа управляющей вычислительной системы.
- 1.16. Обслуживание в столовой (буфете) с самообслуживанием.
- 1.17. Работа телеграфа по доставке телеграмм.
- 1.18. Обслуживание пассажиропотоков в метрополитене.

2. Опишите характер управляющих решений, которые могли бы иметь место в результате проведения анализа каждой из предложенных ранее задач.

3. Выберите для указанных задач (п. 1) варианты формирования потока, очереди, закона обслуживания из следующего набора.

- 3.1. Требования поступают в обслуживающую систему по одному.
- 3.2. Требования поступают в систему пачками.
- 3.3. Порядок поступления требований известен заранее.
- 3.4. Порядок поступления требований – случайный.
- 3.5. Источник требований – бесконечный.
- 3.6. Источник требований имеет ограниченную мощность.
- 3.7. В очередь становится каждое из поступивших требований.
- 3.8. Возможен отказ от присоединения к очереди со стороны некоторых из поступивших требований.
- 3.9. Дисциплина обслуживания: первым пришел – первым обслужен.
- 3.10. Дисциплина обслуживания: пришел последним – обслужен первым.
- 3.11. Имеет место случайный порядок обслуживания.

3.12. Учет определенной системы приоритетов.

3.13. Присоединившиеся к очереди требования не покидают систему до тех пор, пока их не обслужат.

3.14. Некоторые из требований могут покинуть систему, не дождавшись пока их обслужат (т. е. ожидают обслуживания лишь ограниченное время).

3.15. Имеется единственная очередь.

3.16. Имеется несколько параллельных очередей.

3.17. Обслуживающие приборы функционируют параллельно.

3.18. Обслуживающие приборы расположены последовательно.

3.19. Скорость обслуживания не зависит от длины очереди.

3.20. Скорость обслуживания зависит от длины очереди.

4. Определите четыре-пять важных операционных характеристик, которые необходимо измерить количественно, чтобы оценить эффективность и качество работы соответствующих функциональных систем.

2. Марковские системы массового обслуживания

2.1. Марковские случайные процессы и их классификация

Данный класс случайных процессов достаточно подробно исследован в работах выдающегося русского математика Андрея Андреевича Маркова, поэтому и получил такое название. В последнее время такие понятия как цепи Маркова, марковские процессы, свойства марковости стали достаточно привычными и не сходят со страниц отечественных и иностранных книг, диссертаций и статей.

Случайный процесс $x(t)$ называется марковским, если для любого момента t' при фиксированном значении $x(t')$ (каково бы ни было x) значение процесса $x(t)$ при $t > t'$ не зависят от значений процесса $x(t)$ при $t < t'$.

Иначе такие процессы называются процессами без последействия. Случайный процесс $x(t)$ считается заданным, если для любых t_1, t_2, \dots, t_n известен закон распределения (плотность) $f(x_1, x_2, \dots, x_n)$.

Для марковских процессов характерно задание только плотности

$$f(x_i / x_{i-1}), \text{ так как } f(x_i / x_1, x_2, \dots, x_{i-1}) = f(x_i / x_{i-1}).$$

Тогда:

$$f(x_1, x_2, \dots, x_n) = f(x_n / x_{n-1}) \cdot f(x_{n-1} / x_{n-2}) \cdot \dots \cdot f(x_2 / x_1) \cdot f(x_1). \quad (2.1)$$

Для выделения классификационных групп марковских процессов введем $x \in X, t \in T$, а также вектор $Z = (z_x, z_T)$, где z_x принимает значение $\mathbf{0}$, если множество X дискретно и 1 , если множество X непрерывно. Аналогичные значения принимает z_T в отношении множества T . В соот-

ветствии с этим имеем следующие классы процессов:

- 00** — процессы с дискретным числом состояний и дискретным временем (дискретные марковские цепи или последовательности);
- 01** — процессы с дискретным числом состояний и непрерывным временем (непрерывные марковские процессы);
- 10** — процессы с независимыми приращениями;
- 11** — диффузионные процессы.

В дальнейшем нас будут интересовать только первые две группы процессов, так как они активно используются для описания поведения СМО, хотя это и условно. В [3] показано, что при построении приближенных оценок показателей качества могут использоваться и модели процессов типа **10** и **11**.

2.2. Дискретные марковские последовательности

В этом случае процесс $x(t)$ переходит в последовательность $x(n) = x_n$, где $n = 0, 1, 2 \dots$; $x \in X$, а X — счетное или конечное множество.

Дискретная марковская последовательность считается заданной, если:

- определены вероятности

$$P_{i,j}^{n,n+1} = P\{x_{n+1} = j / (x_n = i)\}, \text{ где } i = 0, 1, 2, \dots; j = 0, 1, 2, \dots;$$

- определены начальные вероятности

$$P\{x_0 = i\} = P_i(0), \quad \sum_i P_i(0) = 1, \quad \text{где } i = 1, 2, \dots.$$

Если переходные вероятности не зависят от n , то $P_{i,j}^{n,n+1} = P_{i,j}$, и такая марковская цепь называется однородной и задается матрицей переходов $P = [P_{i,j}]$ и вектором распределения начального состояния $P^T(0) = [P_1(0) P_2(0) \dots P_i(0) \dots]$. Если множество X конечно, то цепь конечная, если нет, то цепь бесконечная.

При конечном или счетном множестве значений X удобно ввести другую интерпретацию марковской цепи. Сопоставим множеству X множество состояний $S : X \Leftrightarrow S$. Если $x = j$, то имеем состояние S_j . В этом случае марковская цепь представлена последовательностью переходов из состояния S_i в состояние S_j .

Проведем классификацию состояний.

1. Состояния S_i и S_j называются *взаимно-связанными*, если S_i связано с S_j и S_j связано с S_i (не обязательно непосредственно).

2. Состояние S_i называется *возвратным*, если $\sum_{n=1}^{\infty} P_{ii}(n) = 1$,

где $P_{ii}(n)$ — вероятность возврата в i -е состояние на n -м шаге. Если

$\sum_{n=1}^{\infty} P_{ii}(n) < 1$, то S_i — невозвратное состояние. Марковская цепь, со-

стоящая из взаимно-связанных (возвратных) состояний называется *эргоидической*.

3. *Поглощающим* (концевым или тупиковым) называется состояние, которое процесс не может покинуть, но в которое он может перейти хотя бы из одного другого состояния.

4. Марковская цепь, которая содержит хотя бы одно поглощающее состояние, называется *поглощающей*.

Эргодические однородные марковские цепи

Исследование таких цепей обычно проводится в двух режимах: переходном и установившемся.

Переходный режим

В переходном режиме требуется найти вектор вероятностей состояний $P(n)$ в n -й момент дискретного времени, где $n = 1, 2, \dots$, при условии, что начальный вектор $P(0)$ известен. В этом случае при известной матрице переходов P имеем:

При этом возможно также и аналитическое решение для $P(i)$ с использованием Z -преобразования (аналога преобразования Лапласа для решетчатых функций). Умножаем каждое уравнение из системы (2.2) на Z^i , где $i = 1, 2, \dots$ и суммируем, используя определение производящей функции,

$$\mathbf{P}^T(z) = \sum_{i=0}^{\infty} z^i \cdot \mathbf{P}^T(i), \quad (2.3)$$

получим следующее векторное уравнение:

$$P^T(z) - P^T(\mathbf{0}) = z \cdot P^T(z) \cdot P \quad . \quad (2.4)$$

Отсюда при существовании обратной матрицы $[E - z \cdot P]^{-1}$,

где E — единичная матрица, получим:

$$P^T(z) = P^T(0)[E - z \cdot P]^{-1}. \quad (2.5)$$

Используя обратное Z -преобразование, получим оценку вектора $P^T(n)$, где $n = 1, 2, \dots$.

Установившийся режим

Условия существования установившегося режима.

1. Если множество состояний S конечно, то необходимым условием является эргодичность структуры графа переходов.
2. Если S счетно, то необходимым условием является эргодичность графа, а также наличие некоторых дополнительных условий, накладываемых на матрицу переходов P .

Рассмотрим конечный случай, тогда из (2.2) имеем:

$$P^T(n) = P^T(n-1) \cdot P.$$

Реализуя предельный переход

$$\lim_{n \rightarrow \infty} P^T(n) = \pi^T,$$

$$\lim_{n \rightarrow \infty} P^T(n-1) = \pi^T,$$

получим:

$$\pi^T = \pi^T \cdot P; \quad \pi^T(E - P) = O^T. \quad (2.6)$$

Система (2.6) имеет нетривиальное решение при условии $\sum_i \pi_i = 1$ и

замены одного из уравнений системы (2.6) на нормировочное условие $\sum_i \pi_i = 1$.

2.3. Общие уравнения непрерывного марковского процесса

Для описания поведения системы в классе непрерывных марковских процессов необходимо:

- определить понятие состояния;
- составить полное множество состояний, в которых может находиться система;
- составить граф состояний, т. е. указать пути возможных переходов системы из состояния в состояние;
- указать, в каком состоянии находится система в начальный момент времени, или задать распределение начальных состояний;
- для каждого возможного перехода указать соответствующую интенсивность перехода $\lambda_{i,j}(t)$ потока событий, переводящих систему из состояния i в состояние j :

$$\lambda_{i,j}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{i,j}(t, t + \Delta t)}{\Delta t}. \quad (2.7)$$

В дальнейшем рассматриваются только случаи, когда $\lambda_{i,j}(t) = \lambda$, т. е. интенсивности переходов не зависят от времени, что справедливо для однородных марковских процессов. Общая система дифференциальных уравнений, описывающая поведение марковского процесса, выглядит следующим образом [15]:

$$P'_j(t) = - \left(\sum_{\substack{i=1 \\ i \neq j}}^n \lambda_{ji} \right) P_j(t) + \sum_{\substack{i=1 \\ i \neq j}}^n P_i(t) \lambda_{ij}, \quad (2.8)$$

где i, j — индексы состояния, $j = 0, 1, \dots, n$, $i = 0, 1, \dots, n$; $P_j(t)$ — вероятность того, что система в момент t находится в состоянии j .

Данная система дифференциальных уравнений составляется по графу состояний. Левая часть каждого уравнения содержит производную ве-

роятности состояния, а правая — столько членов, сколько дуг связано с данным состоянием. Если дуга направлена из состояния, соответствующий член имеет знак минус; если дуга направлена в состояние — знак плюс. Каждый член равен произведению интенсивности перехода, соответствующей данной стрелке, и вероятности того состояния, из которого выходит дуга. При решении данной системы дифференциальных уравнений необходимо учитывать, что $\sum_{j=1}^n P_j(t) = 1$, т. е. одно из уравнений отбрасывается.

Если число состояний конечно и из каждого состояния графа можно перейти за то или иное число шагов в любое другое, то существует уставившийся режим, т. е. предельные вероятности состояний:

$$P_j = \lim_{t \rightarrow \infty} P_j(t), \quad j = \overline{0, n},$$

причем их значения не зависят от начального состояния системы.

Система дифференциальных уравнений (2.8) переходит в систему линейных алгебраических уравнений, решение которой с учетом условия

$$\sum_{j=0}^n P_j = 1 \text{ позволяет найти все предельные вероятности } P_j.$$

Пример 2.1

Специализированная вычислительная система состоит из трех процессоров и общей оперативной памяти. Задания, поступающие на обработку через интервалы времени τ_1 мин., занимают объем оперативной памяти размером в страницу. После трансляции первым процессором в течение τ_2 мин. их объем увеличивается до двух страниц, и они поступают в оперативную память. Затем после редактирования во втором процессоре, которое занимает τ_3 мин на страницу, объем возрастает до трех страниц. Отре-

дактированные задания через оперативную память поступают в третий процессор на решение, требующее τ_4 мин, и покидают систему, минуя оперативную память. $\tau_i, i = \overline{1, 4}$ — независимые случайные величины, распределенные по показательному закону с интенсивностями $\lambda_i, i = \overline{1, 4}$. Объем оперативной памяти ограничен пятью страницами. Определить средний объем занятой оперативной памяти.

Состояние системы можно описать вектором $\{N_1, N_2, N_3\}$, где

N_1 — количество заданий на трансляцию,

N_2 — количество заданий на редактирование,

N_3 — количество заданий, подготовленных к решению.

Для того чтобы система не попадала в тупиковые состояния в связи с ограниченностью оперативной памяти, необходимо исключить из графа все состояния, приводящие к превышению объема оперативной памяти сверх заданной величины. В этом случае марковский график (рис. 2.1) будет эргодическим и позволит составить уравнения Чепмена-Колмогорова для определения вероятностей стационарных состояний системы.

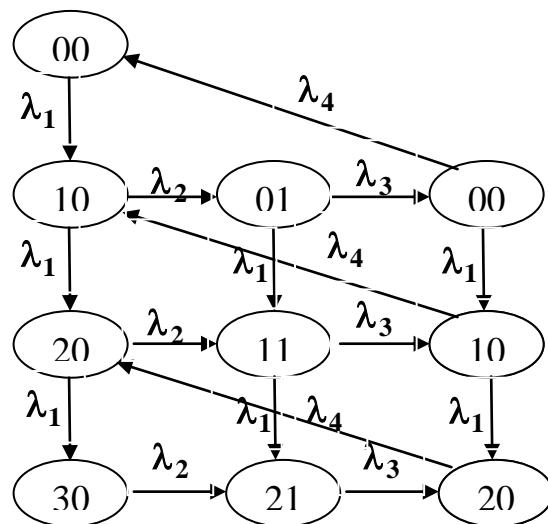


Рис. 2.1

Математическое ожидание объема занятой оперативной памяти рассчитывается по следующей вполне очевидной формуле, так как каждому состоянию системы можно поставить в соответствие количество N занятых страниц памяти, рассчитываемое по формуле $N = N_1 + 2N_2 + 3N_3$:

$$P_{100} + 2[P_{010} + P_{200}] + 3[P_{110} + P_{300} + P_{001}] + 4[P_{101} + P_{210}] + 5P_{201}.$$

Пример 2.2

Детали, необходимые для работы цеха, находятся на цеховом и центральном складах. На цеховом складе хранится 20 комплектов деталей, потребность в которых возникает через τ_1 минут и составляет один комплект. В случае снижения запасов до трёх комплектов формируется в течение τ_2 минут заявка на пополнение запасов цехового склада до полного объема в 20 комплектов, которая посыпается на центральный склад, где в течение τ_3 минут происходит комплектование и за τ_4 минут осуществляется доставка деталей в цех. Определить вероятность простоя цеха. Величины $\tau_i, i = 1, 2, 3, 4$ являются независимыми случайными и распределены по показательному закону с интенсивностью **1/60**.

Состояние системы можно описать вектором $N = \{N_1, N_2, N_3, N_4\}$,

где

N1— текущее число комплектов в цеху ($0 \leq N_1 \leq 20$),

N2— число сформированных заявок на пополнение цехового склада ($0 \leq N_2 \leq 1$),

N3— число заявок на комплектование деталей на центральном складе ($0 \leq N_3 \leq 1$),

N4— число заявок на доставку деталей из центрального на цеховой склад ($0 \leq N_4 \leq 1$).

Граф переходов, представленный на рис. 1, является марковским, вследствие показательного закона распределения всех временных задержек системы. Граф эргодический (все состояния взаимно-связанные и возвратные), число состояний графа конечно, поэтому в системе возможен установившийся режим. Дальнейший расчет стационарных вероятностей системы предполагает составление и решение уравнений Чепмена-Колмогорова. Вероятность простого цеха есть сумма вероятностей состояний, в которых $N1=0$, то есть $P_{0000} + P_{0001} + P_{0010} + P_{0100}$.

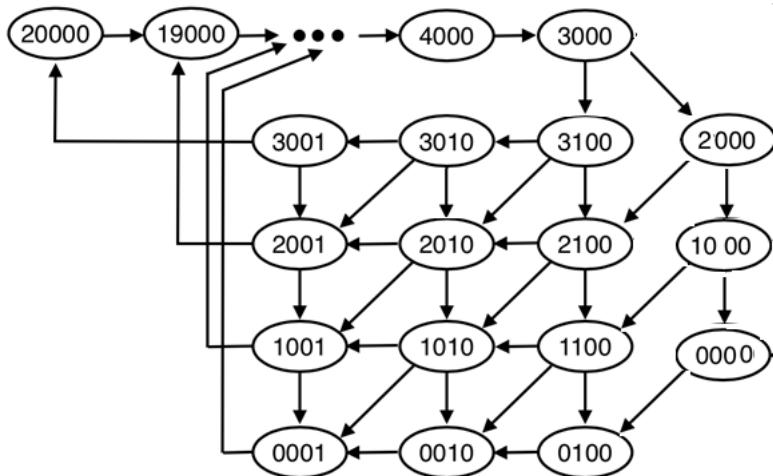


Рис 2.2

2.4. Агрегирование состояний марковского процесса

Для некоторых видов графов, описывающих весьма широкий класс поведения систем, удается снизить размерность системы уравнений (2.8) при анализе характеристик в установившемся режиме.

Пусть график динамики системы выглядит, как показано на рис. 2.3.

Использование обычного способа расчета стационарных вероятностей состояний приводит к системе (2.8) с числом уравнений, равным

(m+1) × (n+1). Агрегирование, т. е. объединение состояний, позволяет перейти к системе **(n+1)** алгебраических уравнений в матричной форме.

Для каждого i -го агрегированного состояния ведем вектор-строку $P_i = \{P_{i0}, P_{i1}, \dots, P_{in}\}, i = \overline{0, m}$.

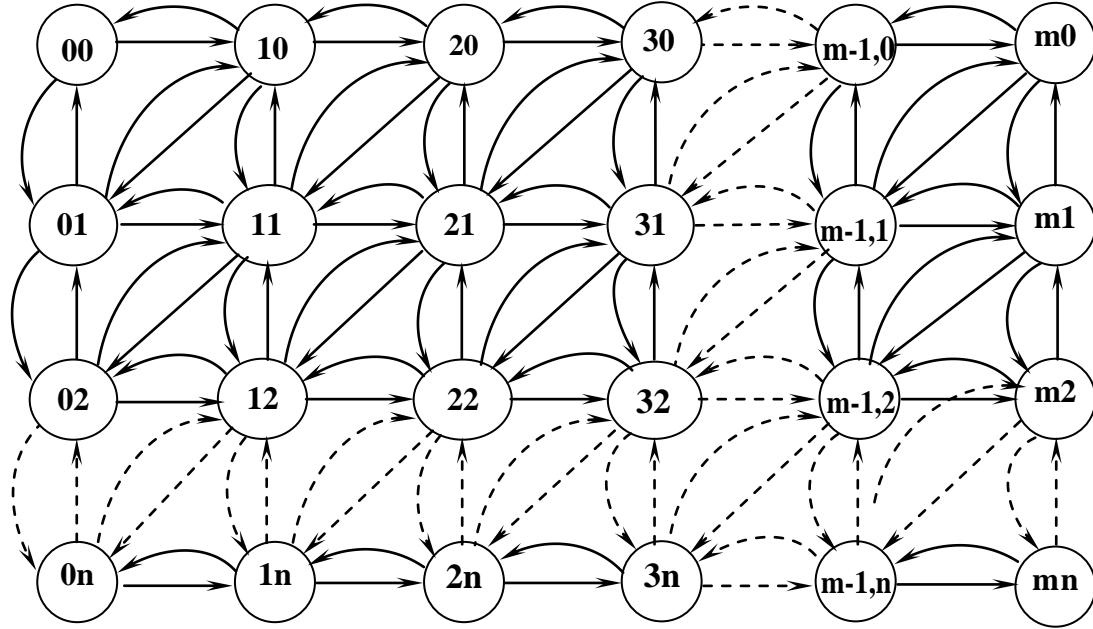


Рис. 2.3

Тогда обобщенная система алгебраических уравнений в случае установившегося режима выглядит следующим образом:

Для ряда структур марковского графа подобная система матричных уравнений баланса потоков приводит к конструктивным методам расчета стационарных вероятностей состояний системы [19, 20]. Коэффициентами

системы уравнений (2.9) служат матрицы интенсивностей переходов A_i , F_i , B_i , D_i . Принцип построения этих матриц поясняется в Приложении. Отметим, что часто для рассматриваемого графа по крайней мере одна из матриц A_i или B_i при всех i может оказаться диагональной. Предположим, что матрицы B_i диагональные. Тогда ищем решение системы уравнений (2.9) в виде:

$$P_i = P_0 W_i, \quad (2.10)$$

где W_i — матрицы размерности $(n+1) \times (n+1)$.

Подставляя (2.10) в (2.9), получим рекуррентные соотношения для вычисления матриц:

$$\begin{cases} W_0 = E, \quad W_1 = -(F_0 - D_0)B_1^{-1} \\ W_2 = -[W_1(F_1 - D_1) + W_0 A_0]B_2^{-1} \\ \dots \dots \dots \dots \dots \\ W_{i+1} = -[W_i(F_i - D_i) + W_{i-1} A_{i-1}]B_{i+1}^{-1} \quad i = \overline{1, m-1}, \\ P_m(F_m - D_m) + P_{m-1} A_{m-1} = P_0[W_m(F_m - D_m) + W_{m-1} A_{m-1}], \end{cases} \quad (2.11)$$

где E — единичная матрица,

а последнее уравнение системы (2.9) перепишем, введя матрицу W^* :

$$P_m(F_m - D_m) + P_{m-1} A_{m-1} = P_0[W_m(F_m - D_m) + W_{m-1} A_{m-1}] = P_0 W^*,$$

причем:

$$P_0 W^* = \mathbf{0}. \quad (2.12)$$

Для определения P_0 из (2.12) необходимо учесть условие нормировки:

$$\left[\sum_{i=0}^m P_0 W_i \right] \mathbf{1}_{(n+1)} = \mathbf{1}, \quad (2.13)$$

где

$$\mathbf{1}_{(n+1)} = \underbrace{\mathbf{1}, \mathbf{1}, \dots, \mathbf{1}}_{n+1}.$$

Составим теперь неоднородную систему уравнений с матрицей коэффициентов W , состоящей из элементов матрицы W^* , с первым столб-

цом, замененным на элементы вектора $\left[\sum_{i=0}^m W_i \right] \mathbf{1}_{(n+1)}$ и вектором-строкой L

свободных членов $L = \left(\underbrace{\mathbf{1}, \mathbf{0}, \mathbf{0}, \dots, \mathbf{0}}_{n+1} \right)$. Тогда:

$$P_0 W = L \quad (2.14)$$

Матрица W невырожденная, так как получена из исходной системы уравнений (2.9) методом последовательного исключения неизвестных и заменой одного уравнения на нормировочное условие. Таким образом, из (2.14) находим P_0 , а далее, используя (2.10), можно найти всё распределение вероятностей.

Предположим теперь, что A_i — диагональные. В этом случае решение для P_i ищем в виде:

$$P_i = P_m V_i, \quad (2.15)$$

где V_i – матрицы той же размерности $(n+1) \times (n+1)$.

Тогда для V_i справедливы следующие рекуррентные соотношения:

$$\begin{aligned} \mathbf{V}^* &= \mathbf{V}_0(\mathbf{F}_0 - \mathbf{D}_0) + \mathbf{V}_1 \mathbf{B}_1. \\ \mathbf{P}_m \mathbf{V}^* &= \mathbf{0}. \end{aligned} \quad (2.17)$$

Условие нормировки в этом случае выглядит следующим образом:

$$\left[\sum_{i=0}^m \mathbf{P}_m \mathbf{V}_i \right] \mathbf{1}_{(n+1)} = \mathbf{1} \quad (2.18)$$

Аналогично, как и в первом случае, заменив в (2.17) первую строку соотношением (2.18), получим систему уравнений для определения \mathbf{P}_m , а, следовательно, используя (2.15), и всех векторов \mathbf{P}_i .

Составим теперь неоднородную систему уравнений с матрицей коэффициентов \mathbf{V} , состоящей из элементов матрицы \mathbf{V}^* , с первым столбцом, замененным на элементы вектора $\left[\sum_{i=0}^m \mathbf{V}_i \right] \mathbf{1}_{(n+1)}$ и вектором-строкой \mathbf{L} свободных членов $\mathbf{L} = \left(\underbrace{\mathbf{1}, \mathbf{0}, \mathbf{0}, \dots, \mathbf{0}}_{n+1} \right)$. Тогда система линейных неоднородных уравнений $\mathbf{P}_m \mathbf{V} = \mathbf{L}$ позволяет найти значения компонентов вектора \mathbf{P}_m , а следовательно, с использованием (2.15) и всех остальных векторов \mathbf{P}_i , $i = \overline{1, m-1}$.

Если оказывается, что обе матрицы \mathbf{B}_i , \mathbf{C}_i диагональные, то можно выбрать любой из алгоритмов. Обратимся к примеру.

Пример 2.3

Рассматривается двухфазная система периферийных технических средств, входящая в комплекс автоматизированных систем управления обслуживанием.

Первая фаза включает накопитель емкостью n_1 и K_1 каналов обслуживания, вторая — K_2 каналов и накопитель емкостью $n_2 = K_1 - K_2 \geq 0$.

В систему поступает пуассоновский поток требований с интенсивностью λ и обслуживается в каналах 1-й фазы с интенсивностью μ_1 , 2-й фазы — с интенсивностью μ_2 , как показано на рис. 2.4.

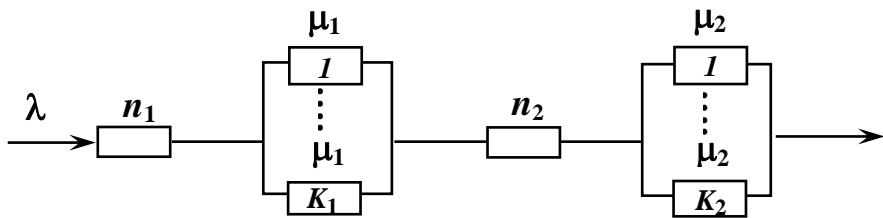


Рис. 2.4

Если при этом на второй фазе имеется свободный канал, требование, обслуженное в первой фазе, немедленно переходит на этот канал, а канал первой фазы принимает, но не обслуживает новые требования до тех пор, пока обслуженное им требование не покинет вторую фазу.

Так как перед прибором второй фазы имеется накопитель емкостью $n_2 = K_1 - K_2$, то потери внутри системы невозможны.

Построим математическую модель функционирования данной системы. Пространство состояний марковского процесса:

$$Z = [(i, j), i = \overline{0, k_1 + n_1}; j = \overline{0, k_2 + n_2}],$$

где i — число требований на первой фазе, j — число требований на второй фазе. Граф марковского процесса для $n_1 = 2; K_1 = 2; n_2 = 1; K_2 = 1$ показан на рис. 2.5. Рассмотрим более подробно схему вычислений предельных вероятностей для случая $n_1 = 1, K_1 = 1, K_2 = 1$. Граф процесса представлен на рис. 2.6.

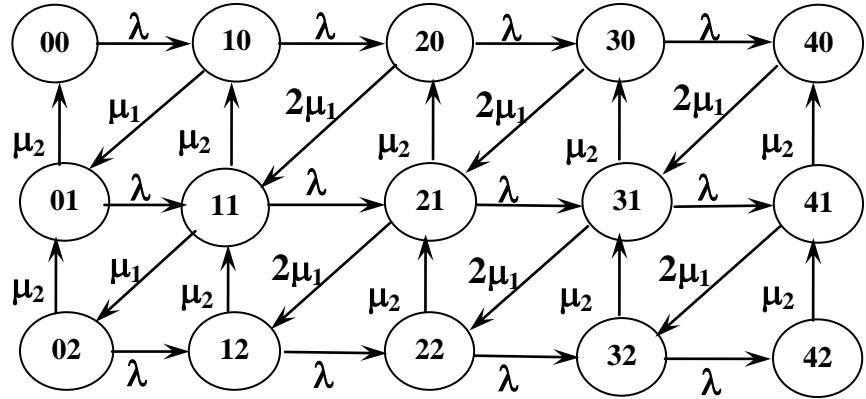


Рис. 2.5

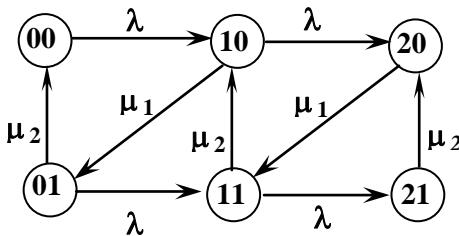


Рис. 2.6

Введем вектор вероятностей состояний:

$$\mathbf{P}_i = (p_{i0} \quad p_{i1}), \quad i = \overline{0, 2}.$$

Система уравнений (2.9) для данного примера выглядит следующим образом:

$$\mathbf{P}_0 \underbrace{\begin{bmatrix} F_0 - D_0 \\ -\lambda & \mathbf{0} \\ \mu_2 & -(\lambda + \mu_2) \end{bmatrix}}_{A_0} + \mathbf{P}_1 \underbrace{\begin{bmatrix} B_1 \\ \mathbf{0} & \mu_1 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}}_{B_1} = \mathbf{0};$$

$$\mathbf{P}_0 \underbrace{\begin{bmatrix} A_0 \\ \lambda & \mathbf{0} \\ \mathbf{0} & \lambda \end{bmatrix}}_{F_1 - D_1} + \mathbf{P}_1 \underbrace{\begin{bmatrix} F_1 - D_1 \\ -(\lambda + \mu_1) & \mathbf{0} \\ \mu_2 & -(\lambda + \mu_2) \end{bmatrix}}_{B_2} + \mathbf{P}_2 \underbrace{\begin{bmatrix} B_2 \\ \mathbf{0} & \mu_1 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}}_{B_2} = \mathbf{0};$$

$$\mathbf{P}_1 \underbrace{\begin{bmatrix} A_1 \\ \lambda & \mathbf{0} \\ \mathbf{0} & \lambda \end{bmatrix}}_{F_2 - D_2} + \mathbf{P}_2 \underbrace{\begin{bmatrix} F_2 - D_2 \\ -\mu_1 & \mathbf{0} \\ \mu_2 & -\mu_2 \end{bmatrix}}_{B_2} = \mathbf{0}.$$

Так как матрицы A_i диагональные, ищем решение в виде:

$$P_i = P_2 V_i, \quad i = 0, 1, 2; \quad V_2 = E,$$

где E — единичная матрица.

Определим матрицы V_0, V_1 :

$$\begin{aligned} V_1 &= -(F_2 - D_2) A_1^{-1} = -\begin{bmatrix} -\mu_1 & \mathbf{0} \\ \mu_2 & -\mu_2 \end{bmatrix} \begin{bmatrix} 1/\lambda & \mathbf{0} \\ \mathbf{0} & -1/\lambda \end{bmatrix} = \begin{bmatrix} \mu_1/\lambda & \mathbf{0} \\ -\mu_2/\lambda & \mu_2/\lambda \end{bmatrix}; \\ V_0 &= -[V_1(F_1 - D_1) + V_2 B_2] A_0^{-1} = \\ &= -\left\{ \begin{bmatrix} \frac{\mu_1}{\lambda} & \mathbf{0} \\ -\frac{\mu_2}{\lambda} & \frac{\mu_2}{\lambda} \end{bmatrix} \begin{bmatrix} -(\lambda + \mu_1) & \mathbf{0} \\ \mu_2 & -(\lambda + \mu_2) \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mu_1 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right\} \begin{bmatrix} 1/\lambda & \mathbf{0} \\ \mathbf{0} & 1/\lambda \end{bmatrix} = \\ &= \begin{bmatrix} \frac{(\lambda + \mu_1) \cdot \mu_1}{\lambda^2} & -\frac{\mu_1}{\lambda} \\ \frac{-(\lambda + \mu_1) \cdot \mu_2 + \mu_2^2}{\lambda^2} & \frac{(\lambda + \mu_2) \cdot \mu_2}{\lambda} \end{bmatrix}. \end{aligned}$$

Примем в дальнейшем, что $\lambda = 1$, $\mu_1 = 2$, $\mu_2 = 1$. Тогда:

$$V_1 = \begin{bmatrix} 2 & -1 \\ \mathbf{0} & 1 \end{bmatrix}, \quad F_0 - D_0 = \begin{bmatrix} -1 & \mathbf{0} \\ 1 & -2 \end{bmatrix}, \quad V_0 = \begin{bmatrix} 6 & -2 \\ -4 & 2 \end{bmatrix}, \quad B_1 = \begin{bmatrix} \mathbf{0} & 2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Первое матричное уравнение приобретает следующий вид:

$$P_2 V^* = P_2 [V_0(F_0 - D_0) + V_1 B_1] = \begin{bmatrix} -8 & 6 \\ 8 & -6 \end{bmatrix} P_2 = \mathbf{0}. \quad (2.19)$$

Условие нормировки:

$$\{P_2 E + P_2 V_1 + P_2 V_0\} \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \end{pmatrix} = \mathbf{1},$$

то есть

$$\left\{ P_2 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} + P_2 \begin{bmatrix} 2 & \mathbf{0} \\ -1 & 1 \end{bmatrix} + P_2 \begin{bmatrix} 6 & -2 \\ -4 & 2 \end{bmatrix} \right\} \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \end{pmatrix} = \mathbf{1}.$$

В результате преобразований получим:

$$7P_{20} - P_{21} = 1. \quad (2.20)$$

Вычерткнув первое уравнение в (2.19) и, заменив его на (2.20), получим систему двух уравнений, разрешение которой дает значение вектора P_2 :

$$P_2 = \begin{pmatrix} 3/17 \\ 4/17 \end{pmatrix}.$$

Тогда остальные векторы P_1 и P_0 определяются в соответствии с матрицами V_1 и V_0 :

$$P_1 = [3/17 \ 4/17] \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix} = [2/17 \ 4/17];$$

$$P_0 = [3/17 \ 4/17] \begin{bmatrix} 6 & -2 \\ -4 & 2 \end{bmatrix} = [2/17 \ 2/17].$$

Для рассматриваемой задачи вероятность потерь требований равна сумме вероятностей $P_{20} + P_{21}$, равной $7/17$.

Пример 2.4

Вычислительная система включает ряд подсистем, рис. 2.7.

- терминалы (среднее время обдумывания результатов $\overline{T_{об\partial}} = 1/\lambda$);
- процессор (среднее время решения задачи в процессоре $\bar{T}_{пес} = 1/\mu$);
- каналы обмена с памятью, включающие селекторные каналы и накопители на магнитных дисках (среднее время обмена $\bar{T}_{обм} = 1/v$).

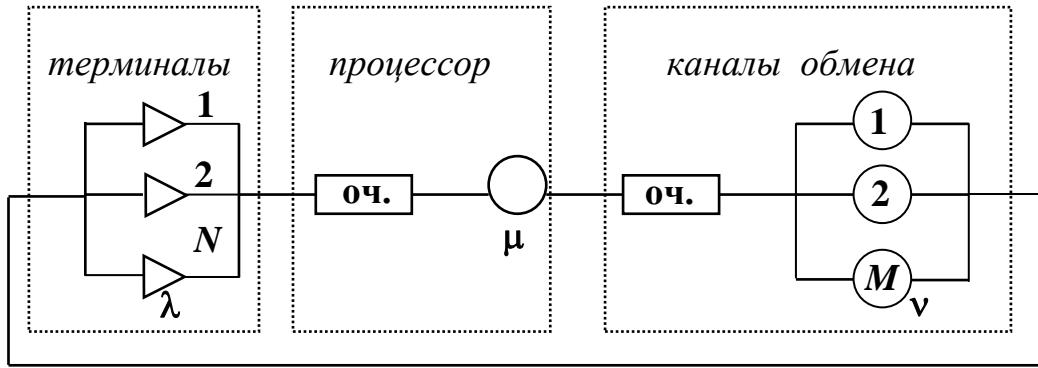


Рис. 2.7

Найти среднее время реакции (время пребывания заявки в системе).

Будем считать, что случайные величины $T_{обд}, T_{реши}, T_{обм}$ подчиняются показательному закону распределения с соответствующими параметрами.

Для описания поведения системы в классе марковских процессов примем следующее определение состояния: $\mathbf{J}(t) = [j_1(t) \ j_2(t)]$, где $j_1(t)$ — число заявок на терминалной фазе (число пользователей, получивших ответ системы и непосредственно работающих за монитором); $j_2(t)$ — число заявок на процессорной фазе. Тогда число заявок на фазе каналов обмена равно $N - j_1(t) - j_2(t)$.

Граф соответствующего марковского процесса представлен на рис. 2.8 при значениях параметров системы $M = 2, N = 3$.

В результате решения соответствующей системы уравнений типа (2.8) находим вероятности состояний P_{ij} . Среднее число занятых терминалов равно: $\bar{N}_{терм} = 1 \cdot (P_{10} + P_{11} + P_{12}) + 2 \cdot (P_{20} + P_{21}) + 3 \cdot P_{30}$.

Используя закон Литтла, найдем среднее время реакции:

$$\bar{t}_p = \frac{N - \bar{N}_{терм}}{\bar{N}_{терм} \cdot \lambda}.$$

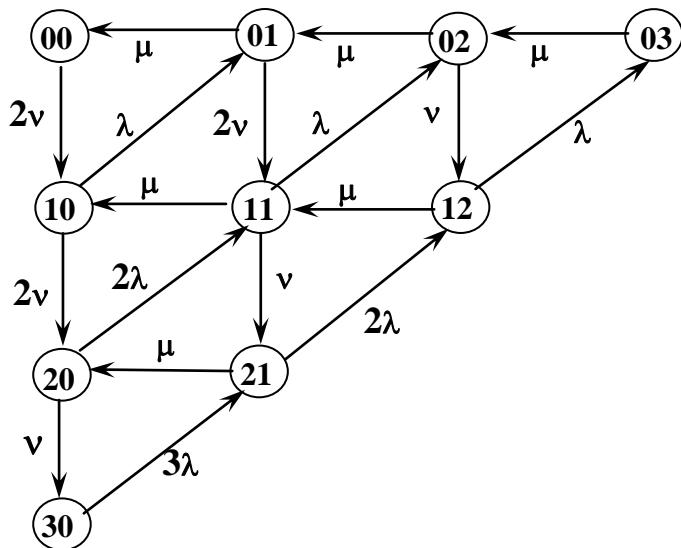


Рис. 2.8

2.5. Процесс гибели и размножения

Граф поведения марковского процесса такого вида представляет собой цепь состояний, в которой каждое из состояний ($j = \overline{1, n-1}$) связано прямой и обратной связями с каждым из соседних состояний, а состояния $j = n$ и $j = 0$ — только с одним из соседних состояний. Если n бесконечно, то цепь может быть продолжена на неограниченное число звеньев в направлении увеличения.

Общая система дифференциальных уравнений, описывающая динамику такого процесса, имеет следующий вид [15]:

Вопросы выяснения существования и единственности решения по-

лученной системы дифференциальных уравнений подробно рассмотрены в [2].

Для того чтобы найденные вероятности $P_j(t)$ удовлетворяли условию:

$\sum_{j=1}^{\infty} P_j(t) = 1$, достаточно иметь ряд $\sum_{j=1}^{\infty} \prod_{k=1}^j \mu_k / \lambda_{k-1}$ расходящимся.

Если при этом оказывается, что ряд

$$\sum_{j=1}^{\infty} \prod_{k=1}^j \lambda_{k-1} / \mu_k$$

сходится, то существуют пределы вероятностей

$$P_j = \lim_{t \rightarrow \infty} P_j(t), \quad j = 0, 1, 2, \dots$$

Последнее условие, в частности, выполняется во всех случаях, когда, начиная с некоторого k , выполнено неравенство $\lambda_{k-1} / \mu_k < 1$, что для большинства практических случаев имеет место.

Предельные вероятности определяются следующим образом:

$$P_0 = \left[1 + \sum_{j=1}^{\infty} \prod_{k=1}^j \lambda_{k-1} / \mu_k \right]^{-1}, \quad P_j = P_0 \prod_{k=1}^j \lambda_{k-1} / \mu_k. \quad (2.22)$$

Пример 2.5

Рассматривается двухпроцессорная управляющая система, на вход которой поступают три простейших входящих потока с интенсивностями:

$$\lambda_1 = 6c^{-1}, \quad \lambda_2 = 15c^{-1}, \quad \lambda_3 = 9c^{-1}.$$

Процессоры считаются однотипными с быстродействием $B = 50 \cdot 10^3$ опер/с. Обслуживание требования заключается в выполнении на любом из процессоров соответствующей прикладной программы, причем средняя трудоемкость всех трех прикладных программ принимается примерно одинаковой $\bar{\Theta} = 2,5 \cdot 10^3$ операций.

От заявки к заявке конкретная трудоемкость меняется случайным образом. Будем для простоты считать закон ее распределения экспоненциальным. Для хранения заявок, которые не могут быть немедленно поставлены на обслуживание, выделена буферная зона памяти емкостью в восемь ячеек; служебная информация об одной заявке занимает две ячейки. Время пребывания заявки в системе не должно превышать случайной величины τ_δ , распределенной экспоненциально с математическим ожиданием $\bar{\tau}_\delta = 0,1 \text{ с}$. Операционная система реализует бесприоритетные дисциплины ожидания и обслуживания. В ее же функции входит удаление “устаревших” заявок из системы. Критерий эффективности управляющей системы имеет вид: $E = \lambda \cdot (e_{OTK} P_{OTK} + e_y P_y) + e_H (\bar{K}_3 - K_3)$,

где e_{OTK} — штраф за отказ СМО принять заявку, $e_{OTK} = 3$ усл.ед./заявка; P_{OTK} — вероятность отказа в обслуживании; e_y — штраф за устаревшие заявки во время обработки ее системой; $e_y = 2$ усл. ед./заявка; P_y — вероятность ухода требования по причине устаревания, e_H — штраф за неиспользование одного канала обслуживания; $e_H = 10$ усл. ед./кан.; \bar{K}_3 — среднее число занятых каналов. Определить E .

Формализация в терминах и понятиях СМО приводит к тому, что мы имеем дело с разомкнутой системой массового обслуживания **M/M/K/m**, где интенсивность обслуживания $\mu = B / \Theta = 50 \cdot 10^3 / 2,5 \cdot 10^3 = 20 \text{ с}^{-1}$. $K = 2$, $m = 4$. Интенсивность ухода заявок из очереди и обслуживающего прибора $\alpha = 1 / \bar{\tau}_\delta = 10 \text{ с}^{-1}$. Общая интенсивность потока при бесприоритетном обслуживании равна: $\lambda = \lambda_1 + \lambda_2 + \lambda_3 = 30 \text{ с}^{-1}$, при этом $\rho = \lambda / \mu = 30 / 20 = 1,5$; $\rho_c = 0,75 < 1$. Граф процесса

гибели и размножения для данной модели представлен на рис. 2.9.

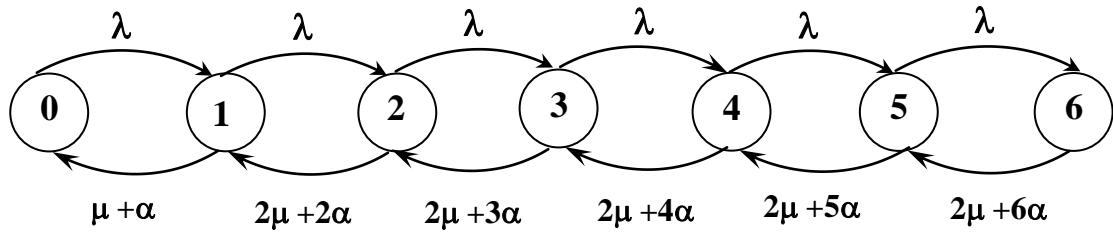


Рис. 2.9

Граф построен с учетом того, что переходы от $(j+1)$ -го состояния к j -му возможны как за счет завершения обслуживания, так и ухода “нетерпеливых” заявок из очереди и из обслуживающих приборов. Используя соотношения (2.22), можно найти вероятностные характеристики состояний:

$$\begin{aligned} P_0 &= 0,3534; & P_1 &= 0,3534; & P_2 &= 0,1767; & P_3 &= 0,0757; \\ P_4 &= 0,0284; & P_5 &= 0,0095; & P_6 &= 0,029. \end{aligned}$$

Среднее число занятых каналов:

$$\bar{K}_3 = P_1 + 2P_2 + 2(1 - P_0 - P_1 - P_2) = 0,9398.$$

Средняя длина очереди: $\bar{n} = 1P_3 + 2P_4 + 3P_5 + \dots + 4P_6 = 0,1726$.

Коэффициент загрузки: $\eta_3 = \bar{K}_3 / K = 0,4699$.

Вероятность отказов по причине ограниченности буфера:

$$P_{\text{отк}} = P_6 = 0,0029.$$

Вероятность ухода заявки во время обслуживания:

$$P_y' = \bar{K}_3 \cdot \alpha / \lambda = 0,9398 \cdot 10 / 30 = 0,3133.$$

Вероятность ухода заявки из очереди:

$$P_y'' = \bar{n} \cdot \alpha / \lambda = 0,1726 \cdot 10 / 30 = 0,0575.$$

Общая вероятность ухода: $P_y = P_y' + P_y'' = 0,3133 + 0,0575 = 0,3708$.

Получено следующее значение критерия эффективности E :

$$E = 30 \cdot (3 \cdot 0,0029 + 2 \cdot 0,3708) + 10 \cdot (2 \cdot 0,9398) = 33,111 \text{ усл. ед.}$$

2.6. Основные соотношения для простейших СМО

Ниже в табл. 2.1 – 2.7 приводятся без вывода базисные расчетные соотношения для основных показателей качества обслуживания в элементарных СМО, описываемых в классе процессов гибели и размножения.

Таблица 2.1

Показатели	M/M/1 — одноканальная система с неограниченной очередью
P_0	$1-\rho; \rho = \frac{\lambda}{\mu}$
P_j	$\rho^j(1-\rho); j \geq 1$
\bar{j}	$\frac{\rho}{(1-\rho)}$
\bar{n}_0	$\frac{\rho^2}{(1-\rho)}$
$t_{ож}$	$\frac{\rho}{(1-\rho)\mu}$
\bar{t}_C	$\frac{1}{\mu(1-\rho)}$
$P(j > k)$	ρ^{K+1}
$P(n_0 > k)$	ρ^{K+2}
$P(t_{ож} < t)$	$1-\rho e^{-\mu(1-\rho)t}; t \geq 0$
$P(t_C < t)$	$1 - e^{-\mu(1-\rho)t}; t \geq 0$
σ_j^2	$\frac{\rho}{(1-\rho)^2}$

Таблица 2.2

Показатели	M/M/1/m — одноканальная система с ограниченной очередью
P_0	$\frac{1-\rho}{1-\rho^{m+2}}$
P_j	$P_0 \rho^j$
\bar{n}_0	$\frac{\rho^2 P_0}{(1-\rho)^2} [1 - \rho^m (m + 1 - m\rho)]$
\bar{j}	$\bar{n}_0 + \frac{\rho - \rho^{m+2}}{1 - \rho^{m+2}}$
$\bar{t}_{ож}$	$\frac{\rho P_0}{\mu (1-\rho)^2} [1 - \rho^m (m + 1 - m\rho)]$
P_{OTK}	$\frac{\rho^{m+1} (1-\rho)}{1 - \rho^{m+2}}$
\bar{t}_c	$\bar{t}_{ож} + \frac{1}{\mu} (1 - P_{OTK})$
Относительная пропускная способность	
$q = 1 - P_{отк}$	
Абсолютная пропускная способность	
$A = \lambda \cdot q$	

Таблица 2.3

Показатели	M/M/K — многоканальная система с неограниченной очередью
P_0	$\left[1 + \sum_{j=1}^K \rho^j \frac{1}{j!} + \rho^{K+1} \frac{1}{K!(K-\rho)} \right]^{-1}$.
P_j	$\begin{cases} \rho^j P_0 / j! & ; \quad j \leq K; \\ \rho^j P_0 / K! K^{j-k} & ; \quad j > K. \end{cases} .$
\bar{n}_0	$\frac{\rho^{K+1} \cdot P_0}{K \cdot K! (1 - \rho_c)^2}; \quad \rho_c = \frac{\lambda}{K\mu}.$
\bar{K}_3	ρ
\bar{j}	$\bar{n}_0 + \bar{K}_3$
$\bar{t}_{ож}$	$\frac{\rho^K \cdot P_0}{K \cdot \mu \cdot K! (1 - \rho_c)^2}$
\bar{t}_c	$\bar{t}_{ож} + \frac{1}{\mu}$
$P(t_{ож} < t)$	$1 - \pi e^{-\mu(K-\rho)t}; \quad \pi = \sum_{j=K}^{\infty} P_j.$

Таблица 2.4

Показатели	M/M/∞ — многоканальная система с неограниченным числом каналов K .
$P_0(t)$	$e^{-\frac{\lambda}{\mu}(1-e^{-\mu t})}$
$P_0 = \lim_{t \rightarrow \infty} P_0(t)$	$e^{-\frac{\lambda}{\mu}}$
$P_j(t)$	$\frac{1}{j!} \cdot \left(\frac{\lambda}{\mu}\right)^j \cdot (1 - e^{-\mu t})^j \cdot P_0(t)$
$P_j = \lim_{t \rightarrow \infty} P_j(t)$	$\frac{1}{j!} \cdot \left(\frac{\lambda}{\mu}\right)^j e^{-\frac{\lambda}{\mu}}$
$K_3(t)$	$\frac{\lambda}{\mu} (1 - e^{-\mu t})$
$\overline{K}_3 = \lim_{t \rightarrow \infty} K_3(t)$	$\frac{\lambda}{\mu}$

Таблица 2.5

Показатели	M/M/K/m — многоканальная система с числом каналов K и ограниченной очередью m
P_0	$\left[1 + \sum_{j=1}^K \rho^j \frac{1}{j!} + \frac{\rho^{K+1}(1-\rho_c^m)}{K \cdot K!(1-\rho_c)} \right]^{-1}; \quad \rho_c = \frac{\lambda}{K\mu}.$
P_j	$\begin{cases} \rho^j \frac{1}{j!} P_0, & 1 < j \leq K; \\ \rho^j \frac{1}{K!} \frac{1}{K^{j-K}} P_0; & K < j \leq K+m. \end{cases}$
P_{omk}	$\frac{\rho^{K+m}}{K^m \cdot K!} \cdot P_0$
\bar{K}_3	$\rho (1 - P_{omk})$
\bar{n}_0	$\frac{\rho^{K+1} \cdot P_0}{K \cdot K!} \left\{ \frac{1 - \rho_c^m [m + 1 - m\rho_c]}{(1 - \rho_c)^2} \right\}$
$\bar{t}_{ож}$	\bar{n}_0 / λ
\bar{j}	$\bar{n}_0 + \bar{K}_3$
\bar{t}_c	$\bar{t}_{ож} + \frac{1}{\mu} (1 - P_{omk})$

Таблица 2.6

Показатели	M/M/1/n — одноканальная система с конечным числом требований
P_0	$\left[\sum_{j=0}^n \frac{n!}{(n-j)!} \rho^j \right]^{-1}$
P_j	$\frac{n!}{(n-j)!} \rho^j P_0 ; \quad j = \overline{1, n}$
\bar{j}	$n - \frac{1 - P_0}{\rho}$
\bar{n}_0	$n - \frac{1 + \rho}{\rho} (1 - P_0)$
$\bar{t}_{ож}$	$\frac{1}{\mu} \left[\frac{n}{1 - P_0} - \frac{1 + \rho}{\rho} \right]$
\bar{t}_c	$\frac{1}{\mu} \left[\frac{n}{1 - P_0} - \frac{1}{\rho} \right]$

Таблица 2.7

Показатели	M/M/K/n — многоканальная система с числом каналов K и с конечным числом требований n , $n > K$
P_0	$\left[1 + \sum_{j=1}^K \rho^j C_n^j + \sum_{j=K+1}^n \rho^j C_n^j \frac{j!}{K! K^{j-K}} \right]^{-1}$
P_j	$\begin{cases} \rho^j \frac{n!}{(n-j)! j!} P_0; & 1 \leq j \leq K, \\ \rho^j \frac{n!}{(n-j)! K! K^{j-K}} P_0; & K < j \leq n. \end{cases}$
\bar{n}_0	$\sum_{j=K+1}^n (j - K) P_j$
\bar{K}_3	$\sum_{j=1}^{K-1} j P_j + K \sum_{j=K}^n P_j$
\bar{j}	$\bar{n}_0 + \bar{K}_3$
\bar{t}_c	$\frac{\bar{j}}{\lambda(n - \bar{j})}$
$\bar{t}_{ож}$	$\bar{t}_c - \frac{1}{\mu}$

Пример 2.6

Провести сравнительный анализ двух структур организации системы обработки информации (рис. 2.10 а, б).

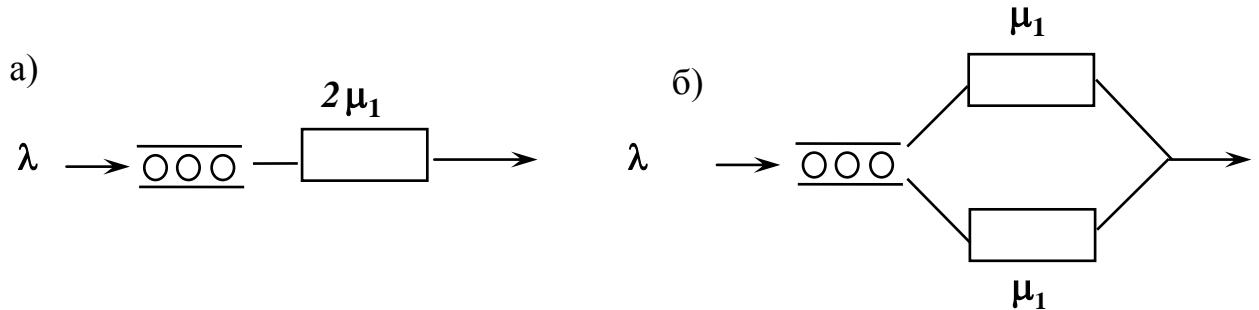


Рис. 2.10

В качестве показателя для проведения сравнительного анализа выберем среднюю длину очереди в системе \bar{n}_0 . Для структуры на рис 2.10 (а) в соответствии с моделью типа **M/M/1** имеем:

$$\bar{n}_0^{(a)} = \frac{\rho^2}{1-\rho} = \frac{(\lambda/2\mu_1)^2}{1-\lambda/2\mu_1} = \frac{1}{2} \frac{\rho_*^2}{(2-\rho_*)}, \quad \rho_* = \frac{\lambda}{\mu_1}.$$

Для структуры на рис. 2.10 (б) в соответствии с моделью типа **M/M/K**:

$$\bar{n}_0^{(b)} = \frac{\rho^{K+1} P_0}{K \cdot K! (1-\rho_c)^2},$$

где $P_0 = \left[1 + \sum_{j=1}^K \rho^j \frac{1}{j!} + \rho^{K+1} \frac{1}{K! (K-\rho)} \right]^{-1}$.

При $K = 2$ получим:

$$\bar{n}_0^{(b)} = \rho_*^3 P_0 \cdot \frac{1}{(2-\rho_*)^2};$$

$$P_0 = \left[1 + \rho_* + \rho_*^2 \frac{1}{2!} + \rho_*^3 \frac{1}{2} \frac{1}{(2 - \rho_*)} \right]^{-1} = \frac{2 - \rho_*}{2 + \rho_*};$$

$$\bar{n}_0^{(b)} = \rho_*^3 \frac{1}{(2 - \rho_*)(2 + \rho_*)} = \frac{\rho_*^3}{4 - \rho_*^2}.$$

Выражение для коэффициента предпочтительности одной структуры по сравнению с другой по данному показателю имеет следующий вид:

$$F_1(\rho_*) = \frac{\bar{n}_0^{(a)}}{\bar{n}_0^{(b)}} = \frac{1}{2} \frac{\rho_*^2}{(2 - \rho_*)} \cdot \frac{4 - \rho_*^2}{\rho_*^3} = \frac{1 + \rho_* / 2}{\rho_*} > 1.$$

Отсюда видно, что во всей возможной области изменения $0 < \rho_* < 2$ средняя длина очереди в первой структуре на рис. 2.10 (а) больше, чем во второй (рис. 2.10 (б)), при одинаковой средней производительности обслуживающей системы.

Рассмотрим другой показатель качества обслуживания, среднее время \bar{t}_c пребывания требования в системе. Величина \bar{t}_c в соответствии с моделью **M/M/1** равна:

$$\bar{t}_c^{(a)} = \frac{1}{\mu} \cdot \frac{1}{1 - \rho} = \frac{1}{2\mu_1} \cdot \frac{1}{1 - \lambda / 2\mu_1} = \frac{1}{\mu_1} \cdot \frac{1}{2 - \rho_*}.$$

Аналогично для \bar{t}_c в соответствии с моделью **M/M/K** имеем:

$$\bar{t}_c^{(b)} = \frac{1}{\mu} + \frac{\rho_*^K \cdot P_0}{K \mu K! (1 - \rho_c)^2}.$$

При $K = 2$ получим:

$$\bar{t}_c^{(b)} = \frac{1}{\mu_1} + \frac{\rho_*^2 \cdot \frac{2-\rho_*}{2+\rho_*} \cdot 2}{2\mu_1 \cdot 2 \left(1 - \frac{\lambda}{2\mu_1}\right)^2} = \frac{1}{\mu_1} \cdot \frac{4}{4 - \rho_*^2} .$$

Коэффициент предпочтительности:

$$F_2(\rho_*) = \frac{\bar{t}_c^{(a)}}{\bar{t}_c^{(b)}} = \frac{2 + \rho_*}{4} < 1 .$$

Отсюда видно, что в области $0 < \rho_* < 2$ среднее время пребывания в системе согласно рис. 2.10 (а) меньше, чем в случае рис. 2.10 (б).

Из данного примера видно, что от выбора показателя качества обслуживания зависит и выбор структуры СМО.

3. НЕМАРКОВСКИЕ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

3.1. Классификация методов исследования немарковских СМО

Как было показано выше, допущение о более общих законах распределения длительности интервалов между требованиями в потоке и длительности обслуживания нарушает марковское свойство и требует более сложных методов исследования.

Отметим основные подходы к исследованию немарковских СМО:

- метод фаз (этапов) Эрланга;
- использование фазовых аппроксимаций законов распределения интервалов поступления и обслуживания требований;
- введение избыточной переменной;
- метод вложенных цепей Маркова;
- интегральный подход.

3.1.1. Метод фаз (этапов)

Метод фаз разработан в самом начале прошлого столетия Эрлангом. А.К. Эрланг заметил, что ряд распределений могут быть разложены в набор показательных распределений. Это в первую очередь касается распределения:

$$f_k(x) = \frac{k\mu(k\mu x)^{k-1} e^{-k\mu x}}{(k-1)!},$$

которое при замене $k\mu=\mu^*$ переходит в гамма-распределение с параметром μ^* :

$$f_k(x) = \frac{\mu^*(\mu^*x)^{k-1} e^{-\mu^*x}}{(k-1)!}. \quad (3.1)$$

В дальнейшем будем считать распределение (3.1) базовым в моделях СМО типа **M/E_k/1**, **E_k/M/1**.

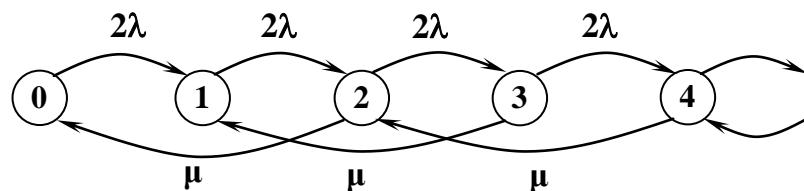
В соответствии с этим длительность обслуживания в системе типа **M/E_k/1** раскладывается в набор k фаз, длительность каждой из которых имеет плотность распределения:

$$f(x) = k \mu e^{-k \mu x}, \quad x \geq 0.$$

Аналогичную ситуацию имеем для системы типа **E_k/M/1**.

На рис. 3.1 представлены графы переходов для случая $k = 2$ систем **E₂/M/1** и **M/E₂/1**. Номер состояния $N(t)$ марковского процесса есть число этапов обслуживания находящихся в системе заявок.

a) Система типа **E₂/M/1**



б) Система типа **M/E₂/1**

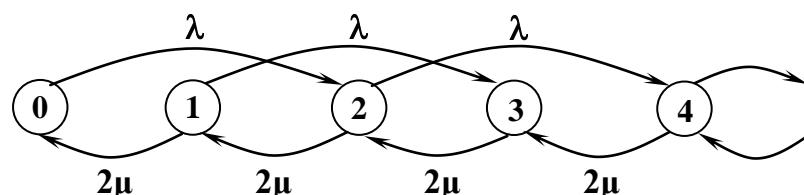


Рис. 3.1. Графы переходов для случая $k = 2$

Так как распределение длительности пребывания в каждом из состояний графов рис. 3.1 является показательным, обладающим отсутствием последействия, то соответствующие процессы являются марковскими и поэтому к ним может быть применен рассмотренный выше в разделе 9.3 метод исследования с использованием общих уравнений непрерывного марковского процесса.

Этот метод связан с увеличением размерности графа состояний и невыгоден тем, что он дает только процедуру решения, но не позволяет получить решение в явном виде.

3.1.2. Использование фазовых аппроксимаций законов распределения интервалов поступления и обслуживания заявок для расчета СМО типа G/G/m

Этот метод опирается на подбираемые по методу моментов фазовые аппроксимации исходных распределений интервалов поступления и обслуживания заявок в СМО G/G/m [10].

Рассмотрим пример трансформации системы массового обслуживания G/G/1 в систему C2/C2/1, порождающую, в отличие от общего случая системы G/G/1, марковский процесс перехода из состояния. Для «марковизации» исходных законов распределения поступления и обслуживания заявок используется подмена указанных распределений дважды коксовыми распределениями с выровненными первыми тремя начальными моментами [20]. Этот прием предполагает существенно более сложную кодировку состояний по сравнению с элементарными системами типа M/M/1, M/M/m, в которых состояние есть скалярная величина — количество заявок в системе. В то же время процесс перехода системы C2/C2/1 из состояния в состояние становится марковским со всеми возможностями его использования. Предположим, что задача выравнивания первых трех начальных мо-

ментов завершена и получены параметры дважды коксовой аппроксимации для входного потока и закона обслуживания.

Распределение Кокса – фазовое распределение, в котором переход к очередной фазе осуществляется с заданной вероятностью, а длительность фаза распределена по показательному закону. На рис. 3.2 представлена схема, поясняющая смысл распределения С2 Кокса [1] для входного потока заявок и закона их обслуживания. Здесь λ_1 и λ_2 – интенсивности 1-й и 2-й фаз поступления заявок в систему, сумма вероятностей u и \bar{u} равна 1, μ_1 и μ_2 – интенсивности 1-й и 2-й фаз обслуживания заявок а, сумма вероятностей y и \bar{y} также равна 1.

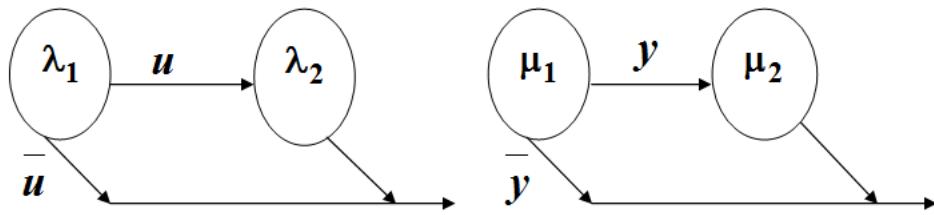


Рис. 3.2

Модель входного потока с использованием распределения С2 предполагает, что с вероятностью \bar{u} на интервале, распределенном по показательному закону с интенсивностью λ_1 появится заявка, которая поступит в систему, а с вероятностью $u = 1 - \bar{u}$ эта заявка не поступит в систему, а останется в приемнике в течение дополнительного интервала времени, распределенного по показательному закону с интенсивностью λ_2 , после чего поступит в систему. Аналогичным образом формируется длительность обслуживания заявки: с вероятностью \bar{y} достаточно одной фазы обслуживания с интенсивностью μ_1 , а с вероятностью $y = 1 - \bar{y}$ к 1-й фазе добавится 2-я фаза с интенсивностью μ_2 .

Состояние системы C2/C2/1 можно описать вектором
 $N = \{N_1, N_2, N_3, N_4\}$, где

N_1 — число заявок в системе, $0 \leq N_1 \leq \infty$,

N_2 — число заявок на обслуживании в 1-й фазе, $0 \leq N_2 \leq 1$,

N_3 — число заявок на обслуживании во 2-й фазе, $0 \leq N_3 \leq 1$,

N_4 — число заявок в приемнике, которым предстоит пройти 2-ю фазу, $0 \leq N_4 \leq 1$.

На рис. 3.3 представлен марковский граф системы C2/C2/1/2 с ограниченной очередью. Максимальное число заявок в этой системе достигает трех: двух заявок в очереди и одной заявки на обслуживании. Каждый из 4-х ярусов графа объединяет подмножество микросостояний, соответствующее определенному числу (0, 1, 2, 3) заявок в системе, равному элементу N_1 вектора состояний. Граф состояний системы C2/C2/1 с неограниченной очередью имеет бесконечное число подобных ярусов, что требует определенных дополнительных усилий для расчета вероятностей состояний этой системы [10, 20].

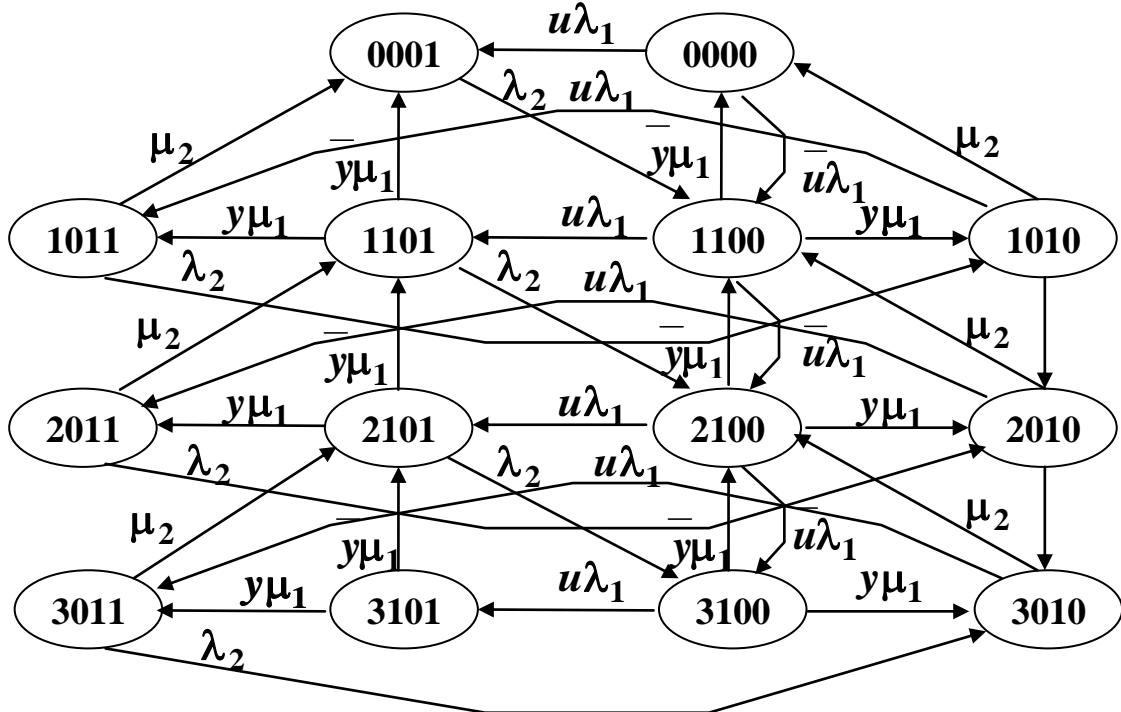


Рис. 3.3

Матрица интенсивностей переходов из состояния в состояние, построенная по графу на рис. 3.3, имеет размерность **14×14** . При увеличении размера накопителя количество ярусов графа возрастает. При увеличении числа каналов обслуживания в системе C2/C2/m или C2/C2/m/k увеличивается также и количество микросостояний яруса. С учетом обеих тенденций размерность матрицы интенсивностей стремительно растет, что делает актуальным переход к агрегированию состояний системы, рассмотренный в разделе 2.4.

Выделим векторные состояния $P_j = \{P_{j0}, P_{j1}, \dots, P_{j\sigma_j}\}$, $i = \overline{0, 3}$,

где $\sigma_0 = 2$, $\sigma_1 = \sigma_2 = \sigma_3 = 4$. Построим матрицы интенсивностей A , B , F переходов из микросостояний соседних векторных состояний (ярусов). Весь набор $A0$, $A1$, $A2$, $F0$, $F1$, $F2$, $F3$, $B1$, $B2$, $B3$ этих матриц,

встроенных в матрицу интенсивностей $\Lambda = [\lambda_{ij}]_{i=1,14 \atop j=1,14}$ для СМО C2/C2/1/2

представлен на рис. 3.4.

	0001	0000	1011	1101	1100	1010	2011	2101	2100	2010	3011	3101	3100	3010
0001		F0			A0	$\bar{\lambda}_2$								
0000	$u\lambda_1$					$u\lambda_1$								
1011	μ_2										λ_2			
1101	$\bar{y}\mu_1$		$y\mu_1$		F1				A1	λ_2				
1100		B1	$\bar{y}\mu_1$		$u\lambda_1$		$y\mu_1$			$u\lambda_1$				
1010		μ_2	$u\lambda_1$							$u\lambda_1$				
2011				μ_2										λ_2
2101				$\bar{y}\mu_1$	B2		$y\mu_1$		F2			A2	λ_2	
2100					$\bar{y}\mu_1$			$u\lambda_1$		$y\mu_1$			$u\lambda_1$	
2010					μ_2		$u\lambda_1$							$u\lambda_1$
3011								μ_2						
3101								$\bar{y}\mu_1$	B3		$y\mu_1$		F3	
3100								$\bar{y}\mu_1$			$u\lambda_1$			$y\mu_1$
3010								μ_2		$u\lambda_1$				

Рис. 3.4

Нетрудно сделать очевидный вывод: матрица интенсивностей Λ , объединяющая переходы в пределах всех состояний системы C2/C2/1/2 может быть представлена как блочная матрица, определенным образом объединяющая подматрицы $A0, A1, A2, F0, F1, F2, F3, B1, B2, B3$. Для того, чтобы сделать этот вывод, достаточно проанализировать рис. 3.4. Заметим, что, начиная с определенного яруса подматрицы начинают повторяться: $A1=A2, F1=F2=F3, B2=B3$, что позволяет алгоритмизировать процедуру построения матрицы Λ по известным подматрицам A_i, F_i, B_i .

3.1.3. Метод введения избыточной переменной

В этом случае одним из возможных способов приведения немарковского процесса к марковскому является восполнение недостающей информации с помощью дополнительной переменной.

Метод был предложен Кендаллом и дальнейшее развитие получил в работах Кокса и Смита [3]. Для иллюстрации основного принципа рассмотрим систему типа **M/G/1**. Как ранее было показано для системы типа **M/M/1**, описываемой в классе процессов гибели и размножения, в качестве состояния процесса достаточно было выбрать число требований j , находящихся в системе. В случае системы **M/G/1** с произвольным законом распределения длительности обслуживания этого недостаточно, и приходится расширить размерность состояния процесса за счет введения в него дополнительной переменной x , времени, прошедшего в системе от начала обслуживания текущей заявки до рассматриваемого момента t . Таким образом, состоянием процесса становится вектор (j, x) и описание дисциплины такого векторного процесса может быть получено на основе уравнений в частных производных.

Естественным является стремление к упрощению такого описания за счет рассмотрения не любых моментов t наблюдения за процессом, а только определенных в соответствии со специально подобранный последовательностью точек на оси времени. Если для этих моментов сохраняются марковские свойства для последовательности переходов в графе состояний, то такая последовательность называется *вложенной марковской цепью*.

3.1.4. Метод вложенных марковских цепей

Для системы **M/G/1** такой последовательностью являются последовательность моментов ухода требований из обслуживающего прибора, при этом состояние процесса определяется числом требований, остающихся в эти моменты в системе. Так как состояние процесса в эти моменты может изменяться только за счет поступающих требований из пуассоновского по-

тока, не обладающего последействием, то соответствующая цепочка переходов будет обладать марковскими свойствами. Аналогичное рассмотрение можно провести и для системы типа **G/M/1**.

В соответствии с данным подходом в [3] выводится известная формула Поллячека-Хинчина для среднего числа требований в системе **M/G/1**:

$$\bar{j} = \rho + \rho^2 \frac{1 + \sigma_x^2 / (\bar{x})^2}{2(1-\rho)}, \quad (3.2)$$

где σ_x^2 и \bar{x} определяются для конкретного закона обслуживания.

3.1.5. Интегральный метод

Все рассмотренные выше методы были основаны на использовании марковских свойств за счет применения тех или иных преобразований пространства состояний или выбора необходимой последовательности наблюдений. При изучении более сложных систем типа **G/G/1** и **G/G/k** приходится отказаться от многих упрощений, вытекающих из марковских свойств и искать новые методы анализа. Рассмотрим основную идею подхода на примере системы типа **G/G/1**.

Все заявки имеют естественный порядок поступления их в систему. Обозначим через C_n заявку с порядковым номером n ($n = 1, 2, \dots$), а через W_n — время ожидания ее в очереди. Тогда, если на обслуживание этой заявки требуется ровно x_n времени, то длительность ее пребывания в системе от момента поступления до завершения обслуживания на приборе определяется:

$$S_n = x_n + W_n. \quad (3.3)$$

Случайное время между приходом заявок C_{n-1} и C_n обозначим t_n .

Установим рекуррентную связь между этими величинами. Каждая вновь появляющаяся заявка может застать обслуживающий прибор свободным или занятым. Если прибор свободен, то заявка сразу начнет обслуживаться и $W_n = 0$; если же прибор оказывается занятым, то заявка C_n прежде чем начать обслуживаться, должна ждать завершения обслуживания заявки C_{n-1} . В результате время ожидания заявки C_n оказывается меньше времени пребывания заявки C_{n-1} на интервал t_n между поступлением в систему, т. е. $W_n = S_{n-1} - t_n$.

Объединяя эти условия, находим, что

$$W_n = \max[0, S_{n-1} - t_n] \quad (3.4)$$

и время S_n пребывания в системе заявки C_n равно

$$S_n = \max[0, S_{n-1} - t_n] + x_n. \quad (3.5)$$

Введем функции распределения случайной величины времени ожидания в очереди и пребывания в системе:

$$W_n(x) = P[W_n < x]; \quad S_n(x) = P[S_n \leq x]. \quad (3.6)$$

Тогда из (3.5) и формулы полной вероятности на основании независимости величин t_n и S_n получим:

$$\begin{aligned} W_n(x) &= P\{W_n = \max[0, S_{n-1} - t_n] \leq x\} = \\ &= \sum_{\Delta\xi} P\{t_n \in (\xi + \Delta\xi)\} P\{S_{n-1} < x + \xi\} \Big|_{\xi=\overline{0,\infty}} = \int_0^\infty S_n(x + \xi) dA(\xi) \end{aligned} \quad (3.7)$$

где $A(x)$ — функция распределения длительности интервала в потоке.

$$S_n(x) = P\{S_n(x) = W_n(x) + x_n < x\} = \int_0^x W_n(x - \xi) dB(\xi), \quad (3.8)$$

где $B(x)$ — функция распределения длительности обслуживания.

Переходя к пределу при $n \rightarrow \infty$ в выражениях (3.7) и (3.8), получим, что в установившемся режиме функции $W(x)$ и $S(x)$ связаны системой двух интегральных уравнений Винера-Хопфа:

$$W(x) = \int_0^{\infty} S(x + \xi) dA(\xi), \quad S(x) = \int_0^x W(x - \xi) dB(\xi), \quad (3.9)$$

разрешение которых позволяет получить $W(x)$ и $S(x)$ для заданных законов $A(x)$ и $B(x)$. Так, например, для системы типа **M/M/1** и законов

$$A(t) = 1 - e^{-\lambda t} \quad \text{и} \quad B(t) = 1 - e^{-\mu t}$$

получим: $W(t) = 1 - \rho e^{-\mu(1-\rho)t}, t \geq 0;$

$$S(t) = 1 - e^{-\mu(1-\rho)t}, \quad t \geq 0.$$

3.2. Характеристики простейших систем массового обслуживания типа M/D/1, M/E_K/1, M/G/1

Важнейшим результатом использования вложенных цепей Маркова для анализа характеристик СМО является формула Поллячека-Хинчина для определения среднего числа требований в одноканальной системе в момент ухода обслуженного требования при произвольном законе распределения длительности обслуживания, т. е. для систем **M/G/1**:

$$\bar{j} = \rho + \rho^2 \frac{\left(1 + \frac{\sigma_x^2}{(\bar{x})^2}\right)}{2(1-\rho)}, \quad (3.10)$$

где σ_x^2 и \bar{x} определяются для конкретного закона обслуживания.

В соответствии с (3.10) имеем:

— для системы **M/M/1**

$$\sigma_x^2 = \frac{1}{\mu^2}, \quad \bar{x} = \frac{1}{\mu}, \quad \bar{j} = \frac{\rho}{1-\rho}; \quad (3.11)$$

– для системы **M/D/1**

$$\sigma_x^2 = 0, \quad \bar{x} = \frac{1}{\mu}, \quad \bar{j} = \rho + \frac{\rho^2}{2(1-\rho)} = \frac{\rho}{1-\rho} - \frac{\rho^2}{2(1-\rho)}; \quad (3.12)$$

что свидетельствует о том, что система **M/D/1** в среднем на $\rho^2/2(1-\rho)$ содержит меньше требований, чем система **M/M/1**;

– для системы **M/E_k/1**

$$\bar{j} = \rho + \frac{\rho^2 \left(1 + \frac{1}{k} \right)}{2(1-\rho)}, \quad (3.13)$$

что свидетельствует о том, что система **M/E_k/1** при $k \rightarrow \infty$ приближается к системе **M/D/1**.

Из формулы (3.14) и закона Литтла (1.18) можно получить соотношения, определяющие время ожидания требования в очереди:

$$\bar{t}_{ож} = \frac{\rho \bar{x} \left(1 + \frac{\sigma_x^2}{\bar{x}^2} \right)}{2(1-\rho)}; \quad (3.14)$$

– для системы **M/D/1**

$$\bar{t}_{ож} = \frac{\rho \frac{1}{\mu}}{2(1-\rho)} = \frac{\rho}{2\mu(1-\rho)}; \quad (3.15)$$

– для системы **M/E_k/1**

$$\bar{t}_{ож} = \frac{\rho \left(1 + \frac{1}{k} \right)}{2\mu(1-\rho)}. \quad (3.16)$$

Использование принципа вложенных марковских цепей позволяет получить не только характеристики моментов, но и законы распределения времени пребывания и времени ожидания в системе.

Если обозначить через $W^*(s)$ и $S^*(s)$ преобразование Лапласа для плотностей распределения $t_{ож}$ и t_c , а через $B^*(s)$ — преобразование Лапласа для закона распределения времени обслуживания, то из [1] известно:

$$S^*(x) = B^*(s) \frac{s(1-\rho)}{s - \lambda + \lambda B^*(s)}, \quad (3.17)$$

$$W^*(s) = \frac{s(1-\rho)}{s - \lambda + \lambda B^*(s)}. \quad (3.18)$$

Реализуя обратное преобразование Лапласа величин $W^*(s)$ и $S^*(s)$ для системы **M/M/1**, получим:

$$P(t_c < t) = 1 - e^{-\mu(1-\rho)t}; \quad (3.19)$$

$$P(t_{ож} < t) = 1 - \rho e^{-\mu(1-\rho)t}. \quad (3.20)$$

Заметим, что для экспоненциального закона распределения времени обслуживания в системе **M/M/1**:

$$B^*(s) = \frac{\mu}{s + \mu}.$$

Одной из важных характеристик систем **M/G/1** является среднее остаточное время до завершения обслуживания требования, находящегося в системе, \bar{T}_o определяемое как:

$$\bar{T}_o = \frac{1}{2} \lambda v^{(2)}, \quad (3.21)$$

где $v^{(2)}$ — второй начальный момент распределения времени обслуживания требования в системе [3]. Данная характеристика является весьма важной при анализе приоритетных систем массового обслуживания и будет использована в разделе 4.

4. СТОХАСТИЧЕСКИЕ СЕТЕВЫЕ МОДЕЛИ СМО

4.1. Предпосылки и цели применения аппарата сетей СМО

Методы теории массового обслуживания, основанные на представлении моделируемой системы в пространстве дискретных состояний, являются универсальными и активно используются для исследования различных объектов. К ним могут относиться не только технические системы, такие как вычислительные системы, системы автоматизации или компьютерные сети, но и организационно-производственные системы широкого назначения.

Однако усложнение структур и режимов реальных систем затрудняет применение классических методов теории массового обслуживания из-за возрастающей размерности решаемой задачи (в частности, порядка уравнений Чепмена–Колмогорова для случая протекания марковского непрерывного стохастического процесса). Один из возможных путей преодоления размерности — использование моделей в форме сетей массового обслуживания.

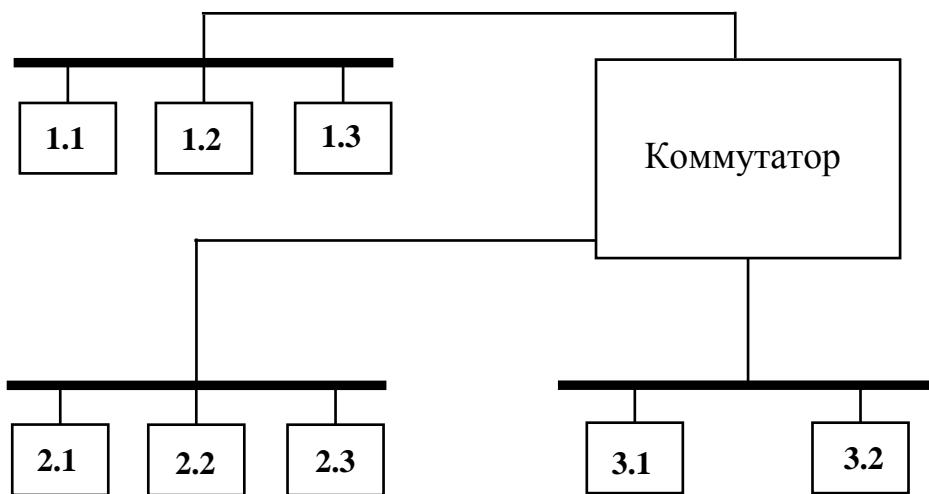


Рис. 4.1

Сеть массового обслуживания представляет собой совокупность конечного числа обслуживающих узлов, в которой циркулируют заявки, передающиеся в соответствии с маршрутной матрицей из одного узла в другой. Узел всегда является разомкнутой СМО.

В качестве примера может быть рассмотрена упрощенная структура локальной вычислительной сети (ЛВС), в состав которой входят коммутатор и три сегмента, объединяющих 8 рабочих станций (рис. 4.1). Рабочие станции нагружают сегменты, генерируя независимо друг от друга пакеты, которые перемещаются по элементам ЛВС в соответствии с вероятностной матрицей передач. Если магистраль сегмента организована по стандарту ETHERNET, то в каждый момент времени в ней может находиться не более одного пакета.

Для построения сети СМО необходимо определить, что является заявкой, в чем состоит “обслуживание заявки”, какими ресурсами моделируемой системы оно осуществляется. Ответы на эти вопросы не всегда очевидны.

Топология сети СМО может не повторять топологию моделируемой системы. Адекватность сетевой модели, в первую очередь, определяется точностью воспроизводимых ею временных задержек.

В данном случае пакеты интерпретируются как заявки, а относительная независимость генерации пакетов от загруженности ЛВС означает присутствие в модели внешнего источника заявок. На рис. 4.2 приведена сеть СМО для рассматриваемого примера. Заявки попадают в сеть из внешнего источника, причем его интенсивность равна суммарной интенсивности пакетов-заявок, поступающих в ЛВС от всех рабочих станций. Как только заявка получает полное обслуживание в сети СМО, т. е. пакет принимается адресатом, она покидает сеть и возвращается в источник Λ_0 .

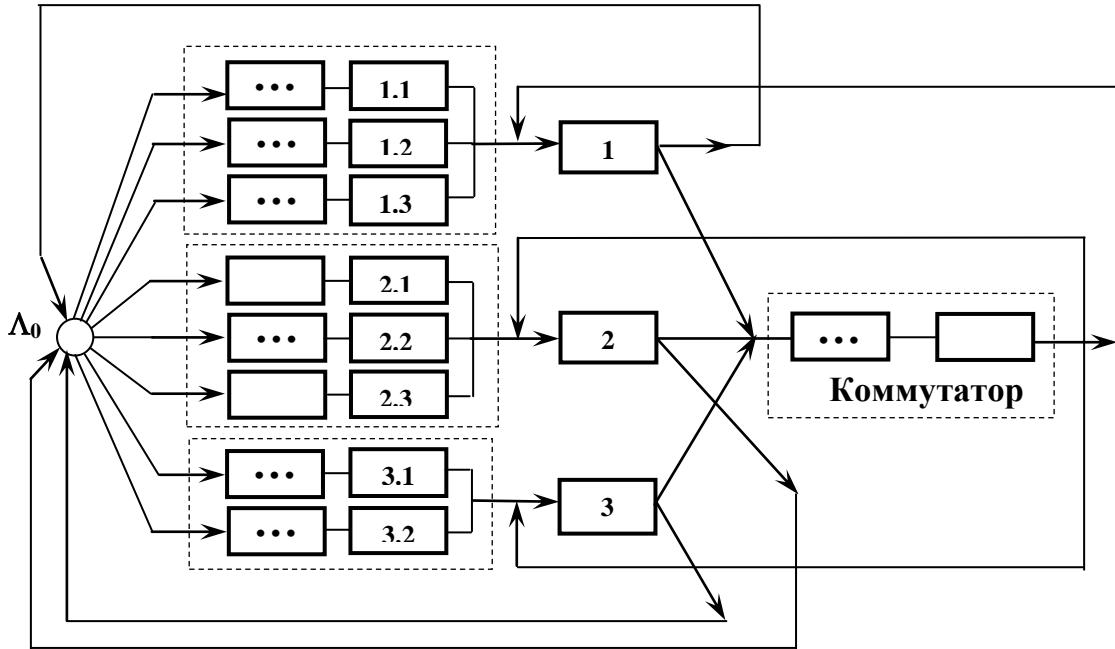


Рис. 4.2

Каждому ресурсу ЛВС в сети массового обслуживания соответствует узел — СМО определенного типа, а именно:

рабочей станции — одноканальный узел с ограниченным накопителем, позволяющим хранить некоторое количество заявок-пакетов, ожидающих отправки;

ETHERNET-магистрали сегмента — одноканальный узел без накопителя;

коммутатору — одноканальный узел с ограниченным накопителем.

Если число выполняемых заданий в моделируемой системе характеризуется постоянством, то для нее, скорее всего, будет построена замкнутая сеть СМО, в которой отсутствует внешний источник заявок. Естественно, это не единственный (хотя и важный) признак, по которому различаются сети массового обслуживания.

Приведенная далее классификация сетей массового обслуживания позволяет оценить спектр используемых сетевых моделей и их возможности адекватного представления реальных систем (рис. 4.3).



Рис. 4.3

Цель классификации состоит в определении типа сети, задающего способы расчета параметров ее функционирования, как использующих аналитические методы, так и методы имитационного моделирования.

В основе представленной классификации — две группы признаков: одна из которых определяет маршрутизацию заявок, другая — характеристики их обслуживания в узлах сети.

4.2. Однородные экспоненциальные сети СМО

4.2.1. Разомкнутые сети СМО. Анализ их характеристик

Аналитическое исследование возможно для стохастических сетей со следующими допущениями:

- сеть формируют узлы, объединяющие каналы с показательным законом обслуживания;
- сеть нагружена пуассоновским потоком однородных заявок;
- в каждом узле сети, очередь не ограничена, дисциплина обслуживания: первым пришел — первым обслужен;
- сеть состоит из узлов типа **M/M/K**;
- сеть является линейной, т. е. вероятность поступления заявки в узел C_j за интервал времени $[t, t + \Delta t]$ — линейная комбинация с постоянным коэффициентом p_{ij} вероятностей выхода заявок из различных узлов сети, включая и узел с номером j .

Формальное задание

линейных экспоненциальных разомкнутых сетей СМО

Линейная экспоненциальная разомкнутая сеть СМО задается кортежем

$$\left\{ \lambda_0; M; \pi = \{ p_{ij} \}, i, j = \overline{0, M}; \vec{\mu} = \{ \mu_1, \dots, \mu_M \}; \vec{m} = \{ m_1, \dots, m_M \} \right\},$$

где

λ_0 — интенсивность внешнего пуассоновского источника заявок;

M — число узлов сети;

$\pi = \{ p_{ij} \}, \quad i, j = \overline{0, M}$ — стохастическая матрица передач;

$\vec{\mu} = \{ \mu_1, \dots, \mu_M \}$ — вектор, i -м компонентом которого является интенсивность обслуживания заявки в канале i -го узла сети;

$\vec{m} = \{ m_1, \dots, m_M \}$ — вектор, i -м компонентом которого является число каналов в i -м узле сети.

Для линейных экспоненциальных сетей рассмотрим только установившийся режим, как наиболее часто встречающийся при решении практических задач. При наличии установившегося режима в разомкнутой сети справедлива следующая закономерность: интенсивность среднего суммарного потока на входе любой подсистемы равна средней суммарной интенсивности выходного потока из данной подсистемы, т. е. справедлива система уравнений:

$$\lambda_j = \sum_{i=0}^M \lambda_i p_{ij}, \quad j = \overline{0, M}, \quad (4.1)$$

где λ_i — средняя суммарная интенсивность на выходе i -й подсистемы; λ_j — суммарная интенсивность на выходе j -й подсистемы; p_{ij} — вероятность поступления требований из i -й подсистемы по окончании обслуживания в j -ю подсистему.

В матричной форме можно записать данную систему уравнений следующим образом:

$$\pi' \lambda = \lambda, \quad (4.2)$$

где π' — транспонированная матрица передач; λ — вектор-столбец интенсивностей λ_i потоков требований, проходящих через подсистемы сети в

установившемся режиме. Уравнение (4.2) можно представить в несколько ином виде, используя понятие единичной матрицы E :

$$(\pi' - E)\lambda = \mathbf{0}.$$

Решение данного матричного уравнения существует, если определитель $|\pi' - E| = 0$. Так как $\sum_{j=0}^M p_{ij} = 1$, $i = \overline{0, M}$, и учитывая, что прибавление к элементам какой-либо строки элементов любой другой строки, умноженных на одно и то же число, не изменяет значения определителя, можно показать, что это условие выполняется. Замыкание в разомкнутой сети происходит через источник требований, отображающий внешнюю среду, следовательно, можно считать, что каждый узел сети является возвратным, т. е. из любой подсистемы через некоторое количество шагов можно попасть в любую другую подсистему, в том числе и в источник требований. Поэтому в такой сети существует установившийся режим, и ранг матрицы $\pi' - E$ равен M , т. е. для каждого значения интенсивности λ_0 источника требований существует единственное решение системы уравнений (4.1):

$$\lambda_j = \alpha_j \lambda_0, \quad j = \overline{1, M},$$

где α_j — коэффициент передачи от источника требований к j -му узлу.

Условие предотвращения образования бесконечной очереди в каждой из подсистем очевидно: $\alpha_j \lambda_0 < m_j \mu_j$, $j = \overline{1, M}$,

где m_j — число каналов в j -м узле.

Это же условие определяет ограничение на интенсивность требований, поступающих из источника

$$\lambda_0 < \min_{(j)} \left(\frac{m_j \mu_j}{\alpha_j} \right). \quad (4.3)$$

Для иллюстрации применения системы уравнений (4.1) рассмотрим

пример.

Пример 4.1. Определить стационарные вероятности стохастической сети, представленной вместе с матрицей передач на рис. 4.4. Пусть интенсивность источника $\lambda_0 = 5 \text{ c}^{-1}$. Тогда в соответствии с (4.1) справедлива следующая система уравнений:

$$\begin{cases} \lambda_0 = \lambda_1 p_{10}; \\ \lambda_1 = \lambda_0 + \lambda_2 + \lambda_3; \\ \lambda_2 = \lambda_1 p_{12}; \\ \lambda_3 = \lambda_1 p_{13}, \end{cases}$$

откуда следует, что $\alpha_1 = [p_{10}]^{-1}$, $\alpha_2 = p_{12}[p_{10}]^{-1}$, $\alpha_3 = p_{13}[p_{10}]^{-1}$.

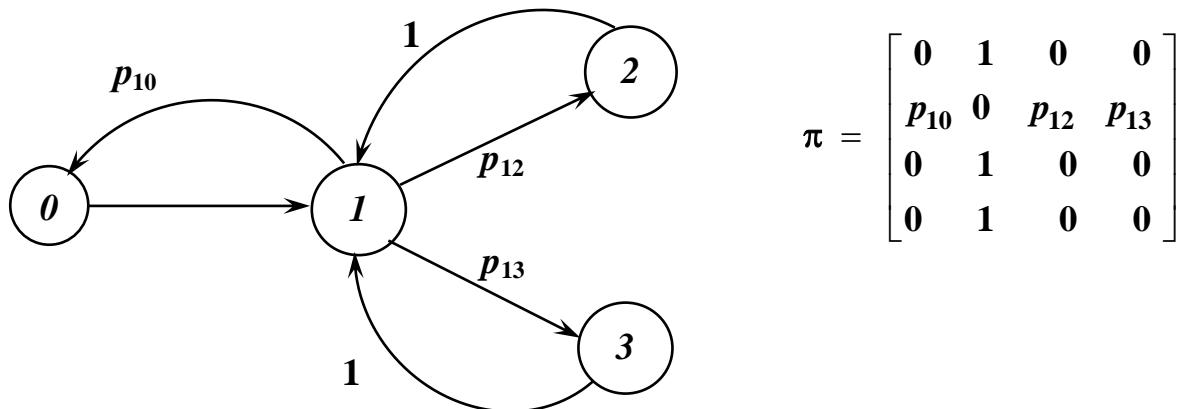


Рис. 4.4

Рассмотрим основные характеристики разомкнутых сетей в целом и их отдельных узлов в установившемся режиме. При определении характеристик сети используется понятие состояния сети.

Под состоянием сети понимается вектор $\mathbf{n} = (n_1, n_2, \dots, n_M)$, в соответствии с которым в узле 1 содержится n_1 требований, в узле 2 — n_2 требований, в узле M — n_M требований. Поэтому наиболее полной характеристикой сети является вероятность $P(n_1, n_2, \dots, n_M)$. Если ра-

зомкнутая сеть представляет собой совокупность узлов с неограниченными очередями, то множество состояний $\{n_i\}$ каждого узла и множество состояний $S = \{n\}$ сети бесконечны.

Важным аналитическим инструментом анализа характеристик сетей является теорема разбиения Джексона [1], содержащая достаточные условия, при выполнении которых разомкнутая сеть общего вида есть совокупность независимых полнодоступных систем с очередями. Далее эта теорема приводится без доказательства, которое можно найти в [1].

Теорема Джексона

Если для рассматриваемой сети выполнены условия:

- все входящие потоки являются пуассоновскими;
- все маршруты переходов требований между подсистемами управляются вероятностной матрицей передач;
- все длительности обслуживания имеют экспоненциальное распределение;
- дисциплина выбора заявок из очереди в каждой подсистеме не зависит от длительностей обслуживания и маршрута заявок, то

$$P(n_1, n_2, \dots, n_j, \dots, n_M) = \prod_{j=1}^M P_j(n_j), \quad (4.4)$$

где

$$P_j(n_j) = P_j(0) \prod_{i=1}^{n_j} \frac{\lambda_j}{\mu_j(i)} ; \quad (4.5)$$

$$\lambda_j = \alpha_j \lambda_0; \quad \mu_j(i) = i \mu_j, \quad i \leq m_j; \quad \mu_j(i) = m_j \mu_j, \quad i > m_j.$$

Интерпретация соотношений (4.5) для разомкнутой сети, когда все узлы сети — одноканальные системы массового обслуживания, может

быть представлена следующим образом:

$$P_j(n_j) = P_j(0)(\rho_j)^{n_j}, \quad (4.6)$$

где

$$\rho_j = \frac{\alpha_j \lambda_0}{\mu_j}, \quad P_j(0) = 1 - \rho_j,$$

а для многоканальных систем:

$$P_j(n_j) = \begin{cases} P_j(0) \left(\frac{\lambda_j}{\mu_j}\right)^{n_j} \frac{1}{n_j!}, & n_j \leq m_j; \\ P_j(0) \left(\frac{\lambda_j}{\mu_j}\right)^{n_j} \frac{1}{m_j!(m_j)^{n_j-m_j}}, & n_j > m_j; \end{cases}, \quad (4.7)$$

$$P_j(0) = \left[\sum_{n_j=0}^{m_j-1} \frac{1}{n_j!} \left(\frac{\lambda_j}{\mu_j}\right)^{n_j} + \frac{\rho_j^{m_j}}{m_j!(1-\rho_j/m_j)} \right]^{-1},$$

где $\rho_j = \frac{\lambda_j}{\mu_j} = \frac{\alpha_j \lambda_0}{\mu_j}$.

В соответствии с рассмотренным принципом декомпозиции задачи анализа сети на задачи анализа отдельных подсистем получим следующие формулы для среднего числа требований в сети \bar{n} ; среднего числа требований, ожидающих обслуживания в сети \bar{l} ; среднего времени пребывания требования в сети \bar{t}_c ; среднего времени ожидания требования в сети $\bar{t}_{ож}$:

$$\bar{n} = \sum_{j=1}^M \bar{n}_j; \quad \bar{l} = \sum_{j=1}^M \bar{l}_j; \quad \bar{t}_c = \sum_{j=1}^M \alpha_j \bar{t}_j; \quad \bar{t}_{ож} = \sum_{j=1}^M \alpha_j \bar{t}_{ожj}, \quad (4.8)$$

где $\alpha_j = \lambda_j / \lambda_0$ определяет среднее число прохождений требования через j -й узел, соответствующее одному требованию, поступившему из 0-го узла в сеть.

Формулы для \bar{n} , \bar{l} , \bar{t} , $\bar{t}_{ож}$ и других параметров, характеризующих отдельные подсистемы (узлы), сведены в табл. 4.1.

Т а б л и ц а 4.1

Основные соотношения для расчета параметров узлов		
Параметр	Узел $M/M/1$	Узел $M/M/m$
P_0	$1 - \rho$	$\left[1 + \sum_{j=1}^m \frac{\rho^j}{j!} + \frac{\rho^{m+1}}{m!(m-\rho)} \right]^{-1}$
\bar{l}	$\frac{\rho^2}{1 - \rho}$	$\frac{P_0 \rho^{m+1}}{m \left(1 - \frac{\rho}{m} \right)^2 m!}$
\bar{n}	$\frac{\rho}{(1 - \rho)}$	$\bar{l} + \rho$
$\bar{t}_{ож}$	$\frac{\rho}{(1 - \rho) \mu}$	$\frac{P_0 \rho^m}{m \mu m! \left(1 - \frac{\rho}{m} \right)^2}$
\bar{t}	$\frac{1}{\mu - \lambda}$	$\bar{t}_{ож} + \frac{1}{\mu}$

Рассмотренные этапы анализа однородных экспоненциальных разомкнутых сетей СМО можно кратко изложить в виде методики.

Методика анализа

однородных экспоненциальных разомкнутых сетей СМО

1. Расчет интенсивностей потоков заявок в узлах сети по формуле (4.1).
2. Проверка условий (4.3) наличия установившегося режима в сети.

3. “Разрезание” сети — расчет локальных показателей узлов сети по формулам (4.6), (4.7) и соотношениям табл. 4.1.

4. “Сборка” сети — расчет интегральных показателей сети по теореме Джексона (формулы (4.4), (4.5) и (4.6), (4.7)) и соотношениям (4.8).

Пример 4.2. Определить среднее время пребывания заявки в разомкнутой сети, структура которой представлена на рис. 4.5.

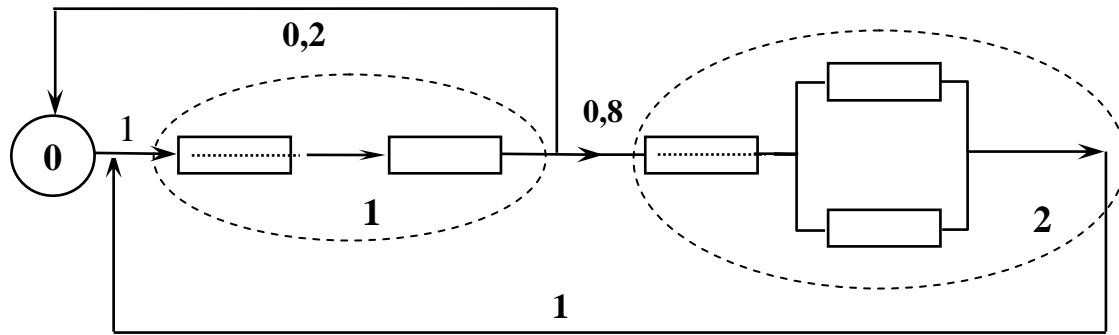


Рис. 4.5

Структура данной сети отображает работу вычислительной системы (ВС) в составе процессора и двух внешних однородных запоминающих устройств, управляемых мультиплексорным каналом. Пусть задана следующая матрица передач:

$$\pi = \begin{bmatrix} 0 & 1 & 0 \\ 0,2 & 0 & 0,8 \\ 0 & 1 & 0 \end{bmatrix};$$

$$\lambda_0 = 0,1 \text{ c}^{-1}; \quad \mu_1 = 2 \text{ c}^{-1}; \quad \mu_2 = 1 \text{ c}^{-1}.$$

Тогда систему уравнений (4.1) можно записать так:

$$\begin{cases} \lambda_0 = 0,2 \lambda_1; \\ \lambda_1 = \lambda_0 + \lambda_2; \\ \lambda_2 = 0,8 \lambda_1, \end{cases}$$

откуда

$$\lambda_1 = 5\lambda_0 = 0,5 \text{ c}^{-1}; \quad \lambda_2 = 0,8 \cdot 5\lambda_0 = 0,4 \text{ c}^{-1}.$$

В рассматриваемой сети стационарный режим существует, так как условие (4.3) выполняется: $0,1 < \min(2/5, 1/2)$.

Среднее время пребывания требования в первом узле есть сумма средних времен ожидания в очереди и обслуживания:

$$\bar{t}_1 = \frac{1}{\mu_1} + \frac{\lambda_1}{\mu_1(\mu_1 - \lambda_1)} = \frac{1}{\mu_1 - \lambda_1} = \frac{1}{2 - 0,5} = 0,67 \text{ с.}$$

Среднее время пребывания требования во втором узле определяется аналогично, но с учетом наличия двух каналов обслуживания:

$$\bar{t}_2 = \frac{1}{\mu_2} + \frac{\rho_2^2 P_0}{2\mu_2 \cdot 2! \left(1 - \frac{\rho_2}{2}\right)^2} \quad \text{или} \quad \bar{t}_2 = 1 + \frac{(0,4)^2}{2 \cdot 2! \left(1 - \frac{0,4}{2}\right)^2} \cdot 0,67 = 1,04 \text{ с.}$$

Среднее время пребывания требования в сети по формуле (4.8)

$$\bar{t}_c = \sum_{j=1}^2 \alpha_j \bar{t}_j = 5 \cdot 0,67 + 4 \cdot 1,04 = 7,51 \text{ с.}$$

Анализ характеристик систем, структура которых отображается в виде многофазных систем массового обслуживания, иногда можно рассматривать как частный случай разомкнутых сетей массового обслуживания.

4.2.2. Замкнутые сети СМО. Анализ их характеристик

В замкнутых сетях всегда имеется конечное число требований, которое сохраняется постоянным при изменении их распределения по отдельным узлам. Мощность $|S|$ множества $S(N, M)$ состояний замкнутой сети определяется по правилам комбинаторики как число вариантов размещения N заявок по M узлам:

$$|S| = C_{N+M-1}^N.$$

*Формальное задание линейных
экспоненциальных замкнутых сетей СМО*

Линейная экспоненциальная замкнутая сеть СМО задается кортежем

$$\left\{ N; M; \pi = \{ p_{ij} \}, i, j = \overline{1, M}; \vec{\mu} = \{ \mu_1, \dots, \mu_M \}; \vec{m} = \{ m_1, \dots, m_M \} \right\},$$

где

N — постоянное число заявок в сети;

M — число узлов сети;

$\pi = \{ p_{ij} \}, i, j = \overline{1, M}$ — стохастическая матрица передач;

$\vec{\mu} = \{ \mu_1, \dots, \mu_M \}$ — вектор, i -м компонентом которого является интенсивность обслуживания заявки в канале i -го узла сети;

$\vec{m} = \{ m_1, \dots, m_M \}$ — вектор, i -м компонентом которого является число каналов в i -м узле сети.

В замкнутой сети обязательно должно выполняться условие:

$$\pi_{i0} = 0; \quad \pi_{0j} = 0, \quad i = \overline{0, M}, \quad j = \overline{0, M},$$

так как требования не покидают сеть, а просто перераспределяются по ее узлам. В этом случае матрица передач имеет размерность $M \times M$, и система уравнений, эквивалентная системе уравнений (4.1), выглядит следующим образом:

$$\lambda_j = \sum_{i=1}^M \lambda_i p_{ij}, \quad j = \overline{1, M}. \quad (4.9)$$

Следовательно, для существования решения определитель матрицы $\pi' - E$ должен быть равен нулю. Можно показать так же, как и для разомкнутых систем, что в этом случае ранг матрицы равен $M - 1$, т. е. в установленном режиме справедлива следующая система решений:

$$\lambda_j = \alpha_j \lambda_1; \quad \alpha_j = \frac{\lambda_j}{\lambda_1},$$

при условии, что интересующие нас интенсивности λ_j выражены через интенсивность λ_1 .

Система (4.9) имеет бесконечное множество решений $\{\lambda_1, \lambda_2, \dots, \lambda_M\}$, получаемых с точностью до мультипликативной константы. Это означает, что точное решение возможно только для относительных интенсивностей α_j , $j = \overline{1, M}$:

$$\begin{cases} \alpha_j = \sum_{i=1}^M \alpha_i p_{ij}, & j = \overline{1, M}; \\ \alpha_1 = 1. \end{cases} \quad (4.10)$$

Так как в замкнутой сети содержится конечное число требований, то в любом узле может образоваться только конечная очередь. Поэтому анализ решения системы (4.10) позволяет выявить только наиболее нагруженную подсистему, характеризуемую условием

$$\max_{(j)} \left\{ \alpha_j / m_j \mu_j \right\}.$$

Для определения вероятностей состояний замкнутой сети используется тот же подход, что и для разомкнутой сети, строгое доказательство которого приведено в [3] — вероятности состояний сети выражаются в мультипликативной форме через вероятности состояний составляющих сеть СМО. Однако полная декомпозиция задачи анализа характеристик замкнутых сетей невозможна, так как в сети содержится конечное число требований. Для учета этого факта вводится дополнительный множитель A , смысл которого может быть пояснен следующим образом.

Представим себе разомкнутую сеть, состоящую из тех же самых узлов, что и рассматриваемая замкнутая сеть. Пусть интенсивность входного

источника и матрица передач таковы, что на входы узлов разомкнутой сети заявки поступают с той же средней интенсивностью, что и на аналогичные входы замкнутой сети. Отличие двух сетей состоит в том, что число состояний разомкнутой сети бесконечно, а замкнутой конечно и равно C_{N+M-1}^N . Выделим подмножество состояний разомкнутой сети, характеризуемое условием

$$\sum_{j=1}^M n_j = N,$$

где n_j — число заявок в j -м узле сети, и определим в соответствии с формулой (4.4) сумму вероятностей всех C_{N+M-1}^N интересующих нас состояний “эквивалентной” разомкнутой сети. Величина, обратная этой сумме, есть множитель A .

В соответствии с этим:

$$P(n_1, n_2, \dots, n_j, \dots, n_M) = A \prod_{j=1}^M P_j(n_j). \quad (4.11)$$

Так как $\sum_{n \in S(N, M)} P(n_1, n_2, \dots, n_j, \dots, n_M) = 1$,

то:

$$A = \left[\sum_{n \in S(N, M)} \prod_{j=1}^M P_j(n_j) \right]^{-1};$$

$$P(n_1, n_2, \dots, n_j, \dots, n_M) = \frac{\prod_{j=1}^M P_j(n_j)}{\sum_{n \in S(N, M)} \prod_{j=1}^M P_j(n_j)}. \quad (4.12)$$

Из (4.6) и (4.7) следует, что вероятности $P_j(n_j)$ для одноканальных и многоканальных систем могут быть представлены произведением:

$$P_j(n_j) = d_j(n_j) P_j(0), \quad (4.13)$$

причем для одноканальных систем

$$d_j(n_j) = (\rho)^{n_j}, \quad n_j = \overline{1, N}, \quad (4.14)$$

а для многоканальных систем

$$d_j(n_j) = \begin{cases} (\rho_j)^{n_j} \frac{1}{n_j!}, & n_j \leq m_j; \\ (\rho_j)^{n_j} \frac{1}{m_j! m_j^{(n_j-m_j)}}, & n_j > m_j. \end{cases} \quad (4.15)$$

На основании (4.14) и (4.15) можно записать:

$$d_j(n_j) = \frac{\lambda_j^{n_j}}{\prod_{k=1}^{n_j} \mu_j(k)}, \quad (4.16)$$

где $\mu_j(k) = \begin{cases} k \mu_j, & k < m_j; \\ m_j \mu_j, & k \geq m_j. \end{cases}$

С учетом этих преобразований из формулы (4.12) получаем:

$$P(n_1, n_2, \dots, n_j, \dots, n_M) = \frac{\prod_{j=1}^M P_j(0) \prod_{j=1}^M d_j(n_j)}{\prod_{j=1}^M P_j(0) \sum_{n \in S(N, M)} \prod_{j=1}^M d_j(n_j)}. \quad (4.17)$$

Можно показать, что вероятность $P(n_1, \dots, n_M)$ не зависит от абсолютного значения интенсивности λ_1 при условии, что все другие интенсивности λ_i выражены через интенсивность λ_1 :

$$P(n_1, \dots, n_M) = \frac{\lambda_1^N \prod_{j=1}^M C_j(n_j)}{\lambda_1^N \sum_{\sum n_j=N} \prod_{j=1}^M C_j(n_j)}, \quad (4.18)$$

причем для одноканальных систем:

$$C_j(n_j) = (\alpha_j / \mu_j)^{n_j}, \quad (4.19)$$

а для многоканальных систем:

$$C_j(n_j) = \begin{cases} \left(\frac{\alpha_j}{\mu_j}\right)^{n_j} \frac{1}{n_j!}, & n_j \leq m_j; \\ \left(\frac{\alpha_j}{\mu_j}\right)^{n_j} \frac{1}{m_j! m_j^{(n_j-m_j)}}, & n_j > m_j. \end{cases} \quad (4.20)$$

По аналогии с (4.16)

$$C_j(n_j) = \frac{\alpha_j^{n_j}}{\prod_{k=1}^{n_j} \mu_j(k)}, \quad (4.21)$$

где $\mu_j(k) = \begin{cases} k \mu_j, & k < m_j; \\ m_j \mu_j, & k \geq m_j. \end{cases}$

Анализ проведенных преобразований позволяет сделать вывод: структура формулы (4.18) такова, что параметр α_j достаточно определить с точностью до мультипликативной константы. Поэтому в соотношениях (4.19)–(4.21) можно вместо параметра α_j ввести параметр ω_j , отличающийся от него постоянным множителем. По смыслу ω_j представляет собой вероятность поступления в j -й узел некоторого помеченного требования при его очередном переходе из узла в узел замкнутой сети:

$$\omega_j = \frac{\alpha_j}{\sum_{i=1}^M \alpha_i} .$$

Вероятности ω_j удовлетворяют нормирующему условию:

$$\sum_{j=1}^V \omega_j = 1 .$$

Можно показать, что значения ω_j могут быть определены непосредственно из системы уравнений, подобной (4.10):

$$\omega_j = \sum_{i=1}^M \omega_i p_{ij} , \quad j = \overline{1, M} , \quad (4.22)$$

с учетом нормирующего условия $\sum_{j=1}^M \omega_j = 1$.

Замена λ_j на ω_j обуславливает следующую запись формулы для определения вероятности состояния замкнутой сети:

$$P(n_1, \dots, n_M) = \frac{1}{G_M(N)} \prod_{i=1}^M Z_i(n_i) , \quad (4.23)$$

где

$$Z_i(n_i) = \frac{\omega_i^{n_i}}{\prod_{j=1}^{n_i} \mu_i(j)} . \quad (4.24)$$

Здесь $G_M(N)$ — нормирующая константа замкнутой сети с M узлами и N заявками:

$$G_M(N) = \sum_{n \in S(N, M)} \prod_{i=1}^M Z_i(n_i) . \quad (4.25)$$

Пример 4.3. Определить коэффициент загрузки процессора в вычислительной системе (ВС), работающей в мультипрограммном режиме,

структурой и матрица передач которой представлены на рис. 4.6.

$$\pi = \begin{bmatrix} p_1 & p_2 & p_3 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

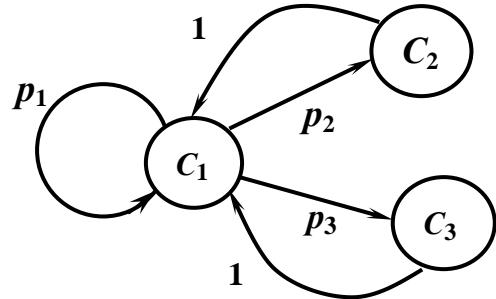


Рис. 4.6

Структура сети СМО отображает взаимодействие процессора C_1 с устройствами внешней памяти C_2 и C_3 . Постоянное число N заявок равно числу выполняемых программ: $N = 3$. Число узлов M равно числу элементов ВС: $M = 3$. Все узлы являются одноканальными. Матрица π описывает передачу управления между элементами ВС в рамках выполнения программы. Решив систему уравнений (4.22), найдем вероятности ω_1 , ω_2 и ω_3 :

$$\begin{cases} \omega_1 = \omega_1 p_1 + \omega_2 + \omega_3; \\ \omega_2 = \omega_1 p_2; \quad \omega_3 = \omega_1 p_3. \end{cases}$$

С учетом нормирующего условия $\sum_{i=1}^3 \omega_i = 1$:

$$\omega_1 = \frac{1}{p_2 + p_3 + 1}, \quad \omega_2 = \frac{p_2}{p_2 + p_3 + 1}, \quad \omega_3 = \frac{p_3}{p_2 + p_3 + 1}.$$

Интенсивности обслуживания μ_j в одноканальных узлах сети считаются заданными.

В соответствии с (4.24) $Z_i(n_i) = \left(\frac{\omega_i}{\mu_i} \right)^{n_i}$.

Коэффициент η загрузки процессора определяется как разность между единицей и суммой вероятностей состояний сети, в которых имеет ме-

сто простой узла C_1 :

$$\eta = 1 - (P(0, 3, 0) + P(0, 0, 3) + P(0, 1, 2) + P(0, 2, 1)).$$

В соответствии с (4.23)

$$\eta = 1 - \frac{\left(\frac{\omega_2}{\mu_2}\right)^3 + \left(\frac{\omega_3}{\mu_3}\right)^3 + \left(\frac{\omega_2}{\mu_2}\right)\left(\frac{\omega_3}{\mu_3}\right)^2 + \left(\frac{\omega_2}{\mu_2}\right)^2\left(\frac{\omega_3}{\mu_3}\right)}{G_3(3)},$$

где $G_3(3) = \sum_{n_1+n_2+n_3=3} \left(\frac{\omega_1}{\mu_1}\right)^{n_1} \left(\frac{\omega_2}{\mu_2}\right)^{n_2} \left(\frac{\omega_3}{\mu_3}\right)^{n_3}$.

Число произведений, суммируемых при расчете нормирующей константы $G_3(3)$, равно числу состояний сети: $C_{N+M-1}^N = C_5^3 = 10$.

4.2.3. Определение основных показателей качества обслуживания в узлах замкнутой сети СМО

Введем следующие обозначения для средних значений показателей качества обслуживания в j -м узле сети при условии пребывания в ней N заявок:

$\overline{n_j}(N)$ — среднее число заявок в узле;

$\overline{n_j^{ож}}(N)$ — среднее число заявок в очереди на обслуживание;

$\overline{n_j^{обсл}}(N)$ — среднее число занятых приборов;

$\overline{t_j}(N)$ — среднее время пребывания заявки в узле;

$\overline{t_j^{ож}}(N)$ — среднее время ожидания заявкой обслуживания;

$\overline{t_j^{обсл}}(N)$ — среднее время обслуживания заявки (равное $1/\mu_j$).

Исходными данными, необходимыми для определения любых показателей качества обслуживания в узлах замкнутой сети, являются вероят-

ности ее состояний, определяемые в соответствии с формулой (4.23). Это в полной мере относится к среднему числу заявок в j -м узле, рассчитываемому по формуле для математического ожидания дискретной случайной величины:

$$\overline{n_j}(N) = \sum_{n \in S(N, M)} n_j P(n_1, \dots, n_j, \dots, n_M). \quad (4.26)$$

Тот же подход используется для определения показателей:

$$\overline{n_j^{ож}}(N) = \sum_{\substack{n \in S(N, M) \\ n_j > m_j}} (n_j - m_j) P(n_1, \dots, n_j, \dots, n_M); \quad (4.27)$$

$$\begin{aligned} \overline{n_j^{обсл}}(N) &= \sum_{\substack{n \in S(N, M) \\ n_j < m_j}} n_j P(n_1, \dots, n_j, \dots, n_M) + \\ &+ m_j \left[\sum_{\substack{n \in S(N, M) \\ n_j > m_j}} P(n_1, \dots, n_j, \dots, n_M) \right]. \end{aligned} \quad (4.28)$$

Определение временных показателей обслуживания в узле замкнутой сети

Заявки, находящиеся в узле сети, заняты ожиданием либо обслуживанием, поэтому:

$$\overline{t_j}(N) = \overline{t_j^{ож}}(N) + \overline{t_j^{обсл}}(N); \quad (4.29)$$

$$\overline{n_j}(N) = \overline{n_j^{ож}}(N) + \overline{n_j^{обсл}}(N). \quad (4.30)$$

При наличии стационарного режима в сети справедливо следующее тождество, представляющее собой вариант записи правила Литтла:

$$\frac{\overline{t_j}(N)}{\overline{n_j}(N)} = \frac{\overline{t_j^{ож}}(N)}{\overline{n_j^{ож}}(N)} = \frac{\overline{t_j^{обсл}}(N)}{\overline{n_j^{обсл}}(N)} = \frac{1/\mu_j}{\overline{n_j^{обсл}}(N)}. \quad (4.31)$$

Каждый член равенства (4.31) есть средняя интенсивность потока заявок в j -м узле, одинаковая также и для его элементов: очереди и обслуживающего прибора. Из (4.31) следует:

$$\bar{t}_j(N) = \frac{\bar{n}_j(N)}{\bar{n}_j^{обсл}(N)} \frac{1}{\mu_j}; \quad (4.32)$$

$$\bar{t}_j^{ож}(N) = \frac{\bar{n}_j^{ож}(N)}{\bar{n}_j^{обсл}(N)} \frac{1}{\mu_j}. \quad (4.33)$$

Подводя промежуточный итог, можно отметить основные этапы расчета однородных экспоненциальных замкнутых сетей СМО в форме упрощенной методики.

Методика анализа однородных экспоненциальных замкнутых сетей СМО

1. Расчет вероятностей ω_j , $j=1, M$, посещения заявками узлов замкнутой сети по формуле (4.22).
2. Расчет вероятностей состояний сети по формуле (4.23).
3. Расчет среднего числа заявок в узлах, очередях и каналах узлов сети по известным вероятностям ее состояний.
4. Расчет средних времен пребывания заявок в узлах, очередях и каналах узлов сети по данным п. 3 с использованием правила Литтла (формулы (4.32), (4.33)).

Пример 4.4. Задана одноканальная сеть с двумя узлами (рис. 4.7), число требований в сети $N = 1$. Определить для каждого из узлов коэффициент загрузки и среднее время пребывания в нем заявки.

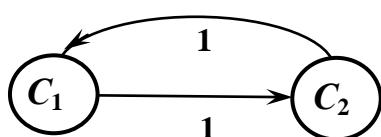


Рис. 4.7

Р е ш е н и е.

1. Расчет вероятностей ω_j : $\omega_1 = \omega_2 = 0,5$.
2. Расчет коэффициентов η_1 и η_2 загрузки узлов:

$$\eta_1 = 1 - \eta_2 = P(1,0) = \frac{\omega_1/\mu_1}{\omega_1/\mu_1 + \omega_2/\mu_2} = \frac{1/\mu_1}{1/\mu_1 + 1/\mu_2} .$$

3. Расчет среднего числа заявок в узле сети

$$\overline{n_1(1)} = 1 - \overline{n_2(1)} = P(1,0) .$$

4. Расчет среднего времени пребывания заявки в узле сети:

$$\overline{t_1(1)} = \frac{\overline{n_1(1)}}{n_1^{\text{обсл}}(1)} \frac{1}{\mu_1} = \frac{\overline{n_1(1)}}{\eta_1} \frac{1}{\mu_1} = \frac{P(1,0)}{\eta_1} \frac{1}{\mu_1} ;$$

$$\overline{t_2(1)} = \frac{\overline{n_2(1)}}{n_2^{\text{обсл}}(1)} \frac{1}{\mu_2} = \frac{\overline{n_2(1)}}{\eta_2} \frac{1}{\mu_2} = \frac{P(0,1)}{1 - \eta_1} \frac{1}{\mu_2} .$$

Строго говоря, последовательность расчета замкнутых сетей СМО, начиная с п. 2 методики, может быть изменена. Преобразования исходной формулы (4.23) позволяют вывести соотношения, обеспечивающие расчет локальных характеристик узлов сети без предварительного определения вероятностей ее состояний. Эти результаты, имеющие важное практическое значение, приводятся в подразд. 4.2.5.

4.2.4. Расчет нормирующей константы однородной замкнутой сети СМО

Расчет нормирующей константы одноканальной замкнутой сети

В основе процедуры расчета нормирующей константы — ранее приведенная формула (4.25):

$$G_M(N) = \sum_{n \in S(N, M)} \prod_{i=1}^M Z_i(n_i).$$

Для одноканальных сетей $Z_i(n_i) = \left(\frac{\omega_i}{\mu_i}\right)^{n_i} = x_i^{n_i}$, тогда в следующем

виде можно представить выражение для

$$\begin{aligned} G_r(k) &= \sum_{\substack{n \in S(k, r) \\ n_r = 0}} \prod_{i=1}^r x_i^{n_i} = \sum_{\substack{n \in S(k, r) \\ n_r > 0}} \prod_{i=1}^r x_i^{n_i} + \sum_{\substack{n \in S(k, r) \\ n_r > 0}} \prod_{i=1}^r x_i^{n_i} = \\ &= G_{r-1}(k) + x_r G_r(k-1), \end{aligned} \quad (4.34)$$

где

$$\begin{cases} G_1(k) = x_1^k, \quad k = \overline{0, N}; \\ G_r(0) = 1, \quad r = \overline{1, M}. \end{cases}$$

С учетом полученных соотношений можно дать следующее формальное изложение процедуры расчета нормирующей константы:

Начальные значения параметров: $r = 2$, $k = 1$.

Шаг 1. $G_r(k) = G_{r-1}(k) + x_r G_r(k-1)$.

Если $k = N$, идти к шагу 2.

Иначе $k := k + 1$, идти к шагу 1.

Шаг 2. Если $r = M$, конец,

Иначе $r := r + 1$, $k := 1$, идти к шагу 1.

В этой процедуре, помимо $G_M(N)$, вычисляются также константы $G_M(r)$, $r = \overline{1, N-1}$, используемые при расчете ряда характеристик замкнутой сети.

Пример 4.5. Рассмотрим использование процедуры расчета нормирующей константы для задачи, изложенной в примере 4.3.

Шаг 1. $G_2(1) = G_1(1) + x_2 G_2(0) = x_1 + x_2$.

Шаг 2. $G_2(2) = G_1(2) + x_2 G_2(1) = x_1^2 + x_2(x_1 + x_2)$.

Шаг 3. $G_2(3) = G_1(3) + x_2 G_2(2) = x_1^3 + x_2 x_1^2 + x_2^2 x_1 + x_2^2$.

Шаг 4. $G_3(1) = G_2(1) + x_3 G_3(0) = x_1 + x_2 + x_3$.

Шаг 5. $G_3(2) = G_2(2) + x_3 G_3(1) = x_1^2 + x_2(x_1 + x_2) + x_3(x_1 + x_2 + x_3)$.

Шаг 6. $G_3(3) = G_2(3) + x_3 G_3(2) = x_1^3 + x_2 x_1^2 + x_2^2 x_1 + x_2^3 + x_3 x_1^2 +$
 $+ x_3 x_2(x_1 + x_2) + x_3^2(x_1 + x_2 + x_3)$.

Расчет нормирующей константы многоканальной замкнутой сети

Для общего случая многоканальной замкнутой сети формула (4.25) может быть представлена в следующем виде:

$$\begin{aligned}
 G_M(N) &= \sum_{\substack{n \in S(N, M) \\ n_M = 0}} \prod_{i=1}^M Z_i(n_i) = \sum_{\substack{n \in S(N, M) \\ n_M = 0}} \prod_{i=1}^M Z_i(n_i) + \\
 &+ \sum_{\substack{n \in S(N, M) \\ n_M = 1}} \prod_{i=1}^M Z_i(n_i) + \sum_{\substack{n \in S(N, M) \\ n_M = 2}} \prod_{i=1}^M Z_i(n_i) + \dots + \sum_{\substack{n \in S(N, M) \\ n_M = N}} \prod_{i=1}^M Z_i(n_i) = \\
 &= \sum_{l=0}^N Z_M(l) \sum_{n \in S(N-l, M-1)} \prod_{i=1}^{M-1} Z_i(n_i) = \sum_{l=0}^N Z_M(l) G_{M-1}(N-l), \quad (4.35)
 \end{aligned}$$

где $Z_i(n_i) = \frac{\omega_i^{n_i}}{\prod_{j=1}^{n_i} \mu_i(j)}$, а $\mu_i(j) = \begin{cases} j \mu_i, & j < m_i; \\ m_i \mu_i, & j \geq m_i; \end{cases}$

$$G_1(k) = Z_1(k) = \frac{\omega_i^k}{\prod_{j=1}^k \mu_1(j)}, \quad k = \overline{0, N};$$

$$G_r(0) = 1, \quad r = \overline{1, M};$$

$$Z_i(0) = 1, \quad i = \overline{1, M}.$$

Тогда с учетом (4.35) получим следующее выражение:

$$G_r(k) = \sum_{l=0}^k Z_r(l) G_{r-1}(k-l). \quad (4.36)$$

Эта формула составляет основу приведенной ниже процедуры расчета нормирующей константы многоканальной замкнутой сети.

Начальные значения параметров: $r = 2, k = 1$.

Шаг 1. $G_r(k) = \sum_{l=0}^k Z_r(l) G_{r-1}(k-l)$.

Если $k = N$, идти к шагу 2.

Иначе $k := k + 1$, идти к шагу 1.

Шаг 2. Если $r = M$, конец.

Иначе $r := r + 1, k := 1$, идти к шагу 1.

Здесь так же, как и для одноканальной сети, справедливо замечание о полезности промежуточных результатов процедуры — значений констант $G_M(r), r = \overline{1, N-1}$ — для расчета ряда характеристик многоканальных замкнутых сетей.

Отметим обстоятельство, которое может оказаться существенным для определения характеристик узлов замкнутой сети.

Процедура расчета нормирующей константы предполагает последовательное включение узлов в состав сети. Нумерация узлов соответствует порядку их внедрения в сеть ($r = 1, 2, \dots, M$). При этом указанный порядок, естественно, не имеет значения для результата расчета нормирующей константы $G_M(N)$, но влияет на смысл и значение получаемых промежуточных констант $G_r(k)$. С учетом последнего принято выделять M -й узел сети, получивший название граничного узла.

Использование операции свертки векторов

для расчета нормирующей константы

Пусть сверткой двух $(R+1)$ -мерных векторов

$A = \{a(0), a(1), \dots, a(R)\}$ и $B = \{b(0), b(1), \dots, b(R)\}$ считается $(R+1)$ -мерный вектор $C = \{c(0), c(1), \dots, c(R)\}$, где

$$c(i) = \sum_{j=0}^i a(j) b(i-j). \quad (4.37)$$

Обозначим операцию свертки

$$C = A * B.$$

Введем в рассмотрение вектор $Z_i = \{Z_i(0), Z_i(1), \dots, Z_i(N)\}$, где

$$Z_i(k) = \frac{\omega_i^k}{\prod_{j=1}^k \mu_i(j)} \quad \text{или} \quad Z_i(k) = \frac{Z_i(k-1)\omega_i}{\mu_i(k)}, \quad k = 1, 2, \dots, N; \quad Z_i(0) = 1.$$

Введем в рассмотрение систему векторов G_0, G_1, \dots, G_M одинаковой размерности $(N+1) \times 1$. Пусть $G_0 = (1, 0, \dots, 0)^T$. В соответствии с (4.36) нормирующая константа $G_r(k)$ есть не что иное, как $(k+1)$ -й компонент вектора G_r , являющегося сверткой $Z_r * G_{r-1}$ векторов Z_r и G_{r-1} :

$$G_r = Z_r * G_{r-1}; \quad (4.38)$$

$$G_r(k) = \sum_{l=0}^k Z_r(l) G_{r-1}(k-l), \quad k = \overline{0, N}.$$

Если последовательно раскрывать G_r в соответствии с (4.38), то в конечном итоге будет получено следующее соотношение:

$$G_r = G_0 * Z_1 * Z_2 * Z_3 * \dots * Z_i * \dots * Z_{r-1} * Z_r;$$

аналогично

$$G_M = G_0 * Z_1 * Z_2 * Z_3 * , \dots , * Z_i * , \dots , * Z_{M-1} * Z_M .$$

Отметим, что искомая нормирующая константа $G_M(N)$ является $(N+1)$ -м компонентом вектора \mathbf{G}_M .

4.2.5. Расчетные соотношения, используемые для определения характеристик узлов однородной замкнутой сети СМО

Маргинальным называется распределение отдельного компонента системы случайных величин. Применительно к замкнутой сети СМО число n_j заявок в j -м узле является компонентом вектора $(n_1, \dots, n_j, \dots, n_M)$, задающего состояние всей сети в целом.

Используя формулу (4.23), можно вывести общее соотношение для вероятности $P_M(n, N)$ состояния граничного M -го узла, характеризуемого наличием n заявок при общем числе заявок в сети, равном N :

$$\begin{aligned} P_M(n, N) &= \sum_{\substack{n \in S(N, M) \\ n_M = n}} \frac{1}{G_M(N)} \prod_{i=1}^M Z_i(n_i) = \\ &= Z_M(n) \frac{\left\{ \sum_{n \in S(N-n, M-1)} \prod_{i=1}^{M-1} Z_i(n_i) \right\}}{G_M(N)} = Z_M(n) \frac{G_{M-1}(N-n)}{G_M(N)}. \end{aligned} \quad (4.39)$$

*Маргинальное распределение
числа заявок в узле однородной замкнутой сети СМО*

Выражение (4.39) — ключевое для получения ряда эффективных формул расчета характеристик замкнутых сетей. Эти формулы, вывод которых можно найти в [5], приводятся далее.

Маргинальное распределение числа заявок в одноканальном узле однородной замкнутой сети СМО

$$P_i(n, N) = \frac{x_i^n [G_M(N-n) - x_i G_M(N-n-1)]}{G_M(N)}. \quad (4.40)$$

Интенсивность на выходе i -го узла однородной замкнутой сети СМО

$$\lambda_i(N) = \frac{\omega_i G_M(N-1)}{G_M(N)}. \quad (4.41)$$

Среднее число заявок в граничном узле однородной замкнутой сети СМО

$$\bar{n}_M(N) = \frac{\sum_{n=1}^N n Z_M(n) G_{M-1}(N-n)}{G_M(N)}. \quad (4.42)$$

Среднее число заявок в одноканальном узле однородной замкнутой сети СМО

$$\bar{n}_i(N) = \frac{\sum_{n=1}^N x_i^n G_M(N-n)}{G_M(N)}. \quad (4.43)$$

Среднее время пребывания заявки в одноканальном узле однородной замкнутой сети СМО

$$\bar{t}_i(N) = \frac{\sum_{n=1}^N x_i^n G_M(N-n)}{\omega_i G_M(N-1)}. \quad (4.44)$$

4.2.6. Метод анализа средних значений характеристик узлов однородной замкнутой сети СМО

В предыдущих пунктах рассмотрен традиционный подход к определению характеристик узлов замкнутых сетей, основанный на использовании теоремы Джексона и предполагающий обязательный расчет нормирующей константы.

Последнее обстоятельство привносит комбинаторно возрастающую вычислительную сложность в анализ замкнутых сетей с увеличением чис-

ла их состояний.

Описываемый в данном пункте итерационный метод не требует предварительного вычисления нормирующей константы, обеспечивая, тем не менее, точный расчет важных характеристик узлов замкнутой сети СМО, таких как средние времена ожидания, длины очередей, коэффициенты загрузки узлов и т. д.

Одноканальные однородные замкнутые сети СМО

Рекуррентная процедура расчета среднего времени пребывания заявки в узле одноканальной однородной замкнутой сети.

Приводится итерационная процедура расчета среднего времени $T_i(N)$ пребывания заявки в i -м узле, основанная на использовании формулы (4.43) в сочетании с правилом Литтла.

Начальные значения параметров: $r = 1$, $\bar{n}_i(0) = 0$, $i = \overline{1, M}$.

$$\text{Шаг 1. } \bar{t}_i(r) = \frac{1}{\mu_i} [1 + \bar{n}_i(r-1)], \quad i = \overline{1, M}.$$

$$\text{Шаг 2. } \lambda_1(r) = \frac{r}{\left[\sum_{i=1}^M \frac{\omega_i \bar{t}_i(r)}{\omega_1} \right]}.$$

$$\text{Шаг 3. } \bar{n}_i(r) = \frac{\omega_i}{\omega_1} \lambda_1(r) \bar{t}_i(r), \quad i = \overline{1, M}.$$

Шаг 4. Если $r = N$, конец.

Иначе $r := r + 1$, идти к шагу 1.

На выходе процедуры можно получить следующие характеристики сети.

1. Коэффициент загрузки i -го узла

$$\rho_i(N) = \frac{\lambda_i(N)}{\mu_i} = \frac{\bar{n}_i(N)}{\bar{t}_i(N)\mu_i}. \quad (4.45)$$

2. Вероятность пребывания заявки в i -м узле сети

$$q_i(N) = \frac{\omega_i \bar{t}_i(N)}{\sum_{j=1}^M \omega_j \bar{t}_j(N)}. \quad (4.46)$$

3. Среднее время ожидания в i -м узле сети

$$\bar{t}_i^{o\text{ж}}(N) = \bar{t}_i(N) - \frac{1}{\mu_i}. \quad (4.47)$$

4. Среднее число заявок в очереди i -го узла сети

$$\bar{n}_i^{o\text{ж}}(N) = \bar{n}_i(N) \frac{\bar{t}_i^{o\text{ж}}(N)}{\bar{t}_i(N)}. \quad (4.48)$$

5. Среднее время цикла для i -го узла сети

$$\bar{V}_i = \frac{[N - \bar{n}_i(N)]}{\lambda_i(N)}. \quad (4.49)$$

Многоканальные однородные замкнутые сети СМО

Рекуррентная процедура вычисления $\bar{t}_i(N)$ и маргинального распределения числа заявок в узлах однородной замкнутой сети.

Начальные значения параметров:

$$P_i(\mathbf{0}, \mathbf{0}) = 1, \quad i = \overline{1, M}; \quad r = 1. \quad (4.50)$$

Шаг 1. $\bar{t}_i(r) = \sum_{n=1}^r \frac{n}{\mu_i(n)} P_i(n-1, r-1), \quad i = \overline{1, M}. \quad (4.51)$

Шаг 2. $\lambda_1(r) = \frac{r}{\sum_{j=1}^M \frac{\omega_j}{\omega_1} \bar{t}_j(r)} \quad (4.52)$

Шаг 3. Для каждого i -го узла сети, $i = \overline{1, M}$, определить набор вероятностей $P_i(n, r)$, $n = \overline{0, r}$:

$$\begin{cases} P_i(n, r) = \frac{\omega_i \lambda_1(r)}{\omega_1 \mu_i(n)} P_i(n-1, r-1), & n = \overline{1, r}; \\ P_i(0, r) = 1 - \sum_{n=1}^r P_i(n, r), \end{cases} \quad (4.53)$$

$$\mu_i(n) = \begin{cases} n \mu_i, & n < m_i; \\ m_i \mu_i, & n \geq m_i, \end{cases}, \quad (4.54)$$

где m_i — число каналов в i -м узле.

Шаг 4. Если $r = N$, конец.

Иначе $r := r + 1$, идти к шагу 1.

На выходе процедуры формируются значения следующих характеристик:

$$\bar{t}_i(N), \quad i = \overline{1, M}; \quad \{P_i(n, N)\}, \quad i = \overline{1, M}; \quad n = \overline{0, N}.$$

П р и м е ч а н и е. Вместо формулы (4.51) может быть использована следующая эквивалентная формула:

$$\bar{t}_i(r) = \frac{1}{\mu_i} \left[1 + \frac{1}{m_i} \sum_{n=m_i}^{r-1} (n - m_i + 1) P_i(n, r-1) \right], \quad i = \overline{1, M}. \quad (4.55)$$

Ряд существенных характеристик узлов многоканальной однородной замкнутой сети так же, как и ранее, может быть получен с помощью формул (4.45)–(4.49).

4.2.7. Приближенный расчет замкнутых сетей с использованием метода баланса заявок

Точные аналитические методы расчета замкнутых сетей Джексона характеризуются высокой вычислительной трудоемкостью. Этот тезис в

равной степени относится к методам, предполагающим обязательный расчет нормирующих констант, и к реализациям метода средних значений для одноканальных и многоканальных сетей Джексона. Точный расчет замкнутых сетей Джексона, как нетрудно убедиться в результате анализа его методики, основан на использовании формул расчета показателей разомкнутых СМО типа $M/M/m$. Действительно, вероятность состояния сети есть с точностью до мультипликативной константы произведение вероятностей состояний разомкнутых СМО типа $M/M/m$, являющихся узлами сети, см. формулу (4.11). Весь расчет показателей замкнутых сетей Джексона основывается на этом их свойстве.

Предположим, что есть две сети Джексона, сформированные из одних и тех же узлов типа $M/M/m$. Пусть эти сети имеют также матрицы передач, обеспечивающие одинаковую маршрутизацию заявок в пределах узлов сети. Разница между сетями состоит в том, что поток заявок узлах разомкнутой сети обусловлен наличием внешнего потока заявок, а аналогичный поток заявок в узлах замкнутой сети есть следствие обслуживания и перемещения узлами сети постоянного числа заявок.

Приближенный расчет замкнутой сети Джексона состоит в том, чтобы поставить замкнутой сети в соответствие разомкнутую сеть, обеспечивающую среднее суммарное количество заявок в сети, равное фиксированному числу N заявок в замкнутой сети. Это свойство постоянства суммарного количества заявок в эквивалентной разомкнутой сети достигается в ходе итерационной процедуры. Идея такого расчета предложена Ю. И. Рыжиковым в [21] и более детально представлена в виде рекомендаций по построению соответствующего алгоритма расчета показателей сети в [10].

Ниже излагается оригинальная версия алгоритма расчета показателей замкнутой многоканальной сети Джексона, основанная на балансе заявок в сети. Формат задания замкнутой сети Джексона приводится в разделе

4.2.2. Расчет вероятностей $\{\omega_i\}, i = \overline{1, M}$ замкнутой сети осуществляется в результате решения системы уравнений (4.22). Итерации алгоритма завершаются по достижении условия $|\bar{L}(B_l) - N| \leq \epsilon$, где $\bar{L}(B_l)$ — среднее число заявок в эквивалентной разомкнутой сети, а ϵ — параметр алгоритма, задающий допустимую погрешность соблюдения условий баланса заявок.

Алгоритм расчета показателей замкнутой сети Джексона по методу баланса заявок

Шаг 1. Найти $\{\omega_i\}, i = \overline{1, M}$,

Шаг 2. Найти наиболее загруженный узел $i^* = \arg \max_{(i)} \left\{ \frac{\omega_i}{m_i \mu_i} \right\}$,

Шаг 3. $l = 0$,

Шаг 4. Принять $B_l = \frac{m_{i^*} \mu_{i^*}}{\omega_{i^*}} \left(1 - \frac{1}{N} \right)$,

Шаг 5. Найти $\{\lambda_i\}, i = \overline{1, M}$, где $\lambda_i = B_l \omega_i$,

Шаг 6. Найти $\{\bar{j}_i(B_l)\}, i = \overline{1, M}$ по формуле для СМО типа М/М/m,

Шаг 7. Найти $\bar{L}(B_l) = \sum_{i=1}^M \bar{j}_i(B_l)$ — суммарное число заявок в сети,

Шаг 8. Если $|\bar{L}(B_l) - N| > \epsilon$ идти к Шагу 9,

Иначе идти к Шагу 10,

Шаг 9. $B_{l+1} = \frac{N}{\bar{L}(B_l)} B_l$,

$l = l + 1$, идти к Шагу 5,

Шаг 10. Конец.

На выходе алгоритма формируются значения следующих показателей:

$$\{\lambda_i\}, i = \overline{1, M},$$

$$\{j_i\}, i = \overline{1, M},$$

Временные показатели сети могут быть получены из «штучных» с использованием правила Литтла:

$$\{\bar{t}_i\}, i = \overline{1, M} \text{ по формуле Литтла } \bar{t}_i = \frac{\bar{j}_i}{\lambda_i},$$

$$\{\bar{t}_i^{\text{ож}}\}, i = \overline{1, M} \text{ по формуле } \bar{t}_i^{\text{ож}} = \bar{t}_i - \frac{1}{\mu_i}.$$

Опыт практического использования предложенного алгоритма показал неидеальную сходимость реализуемого с его помощью метода простых итераций. Альтернативой этому методу является рассмотренный ниже метод Вегстейна [16], обеспечивающий более уверенную сходимость алгоритма к окончательному решению.

Метод Вегстейна

Применительно к данному алгоритму метод Вегстейна предполагает следующую коррекцию шагов алгоритма.

Шаги 1 – 8 остаются без изменений,

$$\text{Шаг 9. } \mathbf{B}_{l+1} = \frac{N}{L(\mathbf{B}_l)} \mathbf{B}_l, \quad y_{l+1} = \mathbf{B}_{l+1},$$

$l = l + 1$, идти к Шагу 5¹,

На всех последующих итерациях для $l > 0$ содержание шага 9 подвергается изменению:

¹ этот вариант шага 9 актуален только для $l = 0$

$$\text{Шаг 9. } \mathbf{y}_{l+1} = \frac{N}{L(\mathbf{B}_l)} \mathbf{B}_l, \quad \mathbf{B}_{l+1} = \mathbf{y}_{l+1} - \frac{(\mathbf{y}_{l+1} - \mathbf{y}_l)(\mathbf{y}_{l+1} - \mathbf{B}_l)}{(\mathbf{y}_{l+1} - \mathbf{y}_l) - (\mathbf{B}_l - \mathbf{B}_{l-1})},$$

$l = l + 1$, идти к Шагу 5,

Шаг 10. Конец.

Данный метод является приближенным. Исследования показали, что точность расчетов в меньшей степени зависит от числа узлов и существенно возрастает с увеличением числа N заявок в сети. При этом наименьшие погрешности расчета достигаются для такого показателя, как среднее число заявок в сети. Если $N \geq 10$, то, как правило, можно гарантировать относительную погрешность расчета среднего числа заявок в узле в пределах 20%, а среднего времени пребывания в узле в пределах 30%. С учетом реальной точности получаемых исходных данных предлагаемый метод может оказаться приемлемым.

4.3. Неоднородные сети СМО

4.3.1. Общие сведения о неоднородных сетях СМО

Рассматриваются разомкнутые линейные экспоненциальные неоднородные сети СМО. Неоднородность сети обусловлена принадлежностью заявок различным классам, определяющим разную маршрутизацию и обслуживание в узлах.

Состояние неоднородной сети задается матрицей

$$S = \begin{Bmatrix} \vec{n}_1 \\ \vec{n}_2 \\ \cdot \\ \cdot \\ \vec{n}_M \end{Bmatrix} = \begin{Bmatrix} n_{11}, n_{12}, \dots, n_{1V} \\ n_{21}, n_{22}, \dots, n_{2V} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ n_{M1}, n_{M2}, \dots, n_{MV} \end{Bmatrix}, \quad (4.56)$$

где $\vec{n}_i = (n_{i1}, n_{i2}, \dots, n_{iV})$ — вектор состояния i -го узла сети; M — число узлов сети; V — число классов заявок.

Основная часть представленных далее аналитических решений для разомкнутых и замкнутых сетей справедлива в том случае, если узлы сети принадлежат одному из четырех типов [3].

Узел типа 1. Дисциплина обслуживания FIFO (в порядке поступления). Длительность обслуживания заявок всех классов в i -м узле имеет одно и то же экспоненциальное распределение с интенсивностью $\mu_i(n_i)$, зависящей от числа n_i заявок в узле.

Узел типа 2. Обслуживание заявок в одноканальном узле осуществляется в соответствии с дисциплиной PS¹. Длительность обслуживания распределена по закону Кокса [3]².

Узел типа 3. Многоканальный узел с бесконечным числом каналов и дисциплиной IS (без ожидания). Длительность обслуживания задается также, как и для узла типа 2.

¹ Дисциплина Processor Sharing [3] предполагает предоставление бесконечно малого кванта времени обслуживания каждой заявке. Не успевшие обслужиться заявки становятся на дообслуживание в хвост очереди.

² Показательный закон является частным случаем закона Кокса.

Узел типа 4. Одноканальный узел с дисциплиной обслуживания LIFO (в порядке, обратном поступлению) и с прерыванием обслуживания. Длительность обслуживания задается так же, как и для узла типа 2.

4.3.2. Разомкнутые неоднородные сети СМО

Разомкнутые сети массового обслуживания характеризуются наличием внешнего источника требований. Требования (заявки) поступают в сеть, обслуживаются в узлах сети и покидают ее. В неоднородной сети имеет место несколько классов заявок, различающихся маршрутизацией и обслуживанием в узлах.

Рассмотрим общий класс разомкнутых неоднородных сетей СМО: заявки меняют класс при переходе из узла в узел. Это означает, что матрица передач в такой сети приобретает форму

$$\pi = \{p_{ik,jv}\}, \quad i, j = \overline{0, M}; \quad k, v = \overline{1, V}, \quad (4.57)$$

где M — число узлов сети; V — число классов заявок; $p_{ik,jv}$ — вероятность того, что заявка k -го класса, завершившая обслуживание в i -м узле, поступит в j -й узел и станет заявкой v -го класса.

Модель входного источника для неоднородной разомкнутой сети может быть задана следующим образом:

$\Lambda_0 = \{\lambda_{0k}\}$, $k = \overline{1, V}$, где λ_{0k} — интенсивность внешнего источника заявок k -го класса.

Если сеть линейная, то для каждого узла справедливо уравнение

$$\lambda_{jv} = \sum_{i=1}^M \sum_{k=1}^V \lambda_{ik} p_{ik,jv}, \quad j = \overline{1, M}; \quad k, v = \overline{1, V}. \quad (4.58)$$

Условие существования установившегося режима состоит в том, что суммарный коэффициент загрузки R_j каждого узла должен быть меньше

единицы:

$$R_j = \sum_{v=1}^V \frac{\lambda_{jv}}{\mu_{jv}} < 1, \quad j = \overline{1, M}, \quad (4.59)$$

где μ_{jv} — средняя интенсивность обслуживания заявок v -го класса в j -м узле.

Если заявки не меняют класса, то каждому из них соответствует своя матрица передач

$$\pi^{(v)} = \left\{ p_{iv, jv} \right\} = \left\{ p_{ij}^{(v)} \right\}, \quad i, j = \overline{0, M}; \quad v = \overline{1, V},$$

а система уравнений (4.58) приобретает следующий вид:

$$\lambda_{jv} = \sum_{i=0}^M \lambda_{iv} p_{ij}^{(v)}, \quad j = \overline{1, M}, \quad v = \overline{1, V}. \quad (4.60)$$

Решением системы будет

$$\lambda_{jv} = \alpha_{jv} \lambda_{0v}, \quad j = \overline{1, M}; \quad v = \overline{1, V}, \quad (4.61)$$

где λ_{0v} — средняя интенсивность внешнего источника заявок v -го класса; α_{jv} — коэффициент передачи от источника требования v -го класса к j -му узлу.

При этом условие (4.59) стационарности сети конкретизируется:

$$R_j = \sum_{v=1}^V \frac{\alpha_{jv} \lambda_{0v}}{\mu_{jv}} < 1, \quad j = \overline{1, M}.$$

Сделаем важный вывод. В разомкнутых линейных экспоненциальных сетях массового обслуживания, в установившемся режиме, интенсивности потоков в узлах зависят только от матрицы передач и входных потоков, а их значения определяются системой (4.58). Данное решение справедливо в условиях (4.59).

Распределение числа заявок в сети СМО без разделения на классы

Рассмотрим задачу определения вероятностей агрегированных состояний неоднородных разомкнутых сетей СМО. В этом случае состояние S сети задается вектором (n_1, n_2, \dots, n_M) , а выражение для стационарного распределения состояний сети принимает вид:

$$P(S) = P_1(n_1) P_2(n_2) \dots P_M(n_M), \quad (4.62)$$

где

$$P_i(n_i) = \begin{cases} (1 - R_i) R_i^{n_i} & \text{для узлов типов 1, 2, 4;} \\ \frac{R_i^{n_i} e^{-R_i}}{n_i!} & \text{для узлов типа 3.} \end{cases} \quad (4.63)$$

Отметим, что для узла сети, принадлежащего типу 1, (4.63) предполагает наличие только одноканальных узлов с однородным обслуживанием заявок разных классов ($\mu_{ir} = \mu_i$).

4.3.3. Замкнутые неоднородные сети СМО

Замкнутые сети массового обслуживания характеризуются постоянным числом заявок. В неоднородной замкнутой сети циркулируют различающиеся маршрутизацией и/или обслуживанием в узлах заявки нескольких классов с постоянным числом N_v ($v = \overline{1, V}$) в каждом из них.

Неоднородная замкнутая экспоненциальная сеть может быть задана следующим набором параметров:

$$\left(M, \vec{N}, \left\{ \pi^{(v)} \right\}, \left\{ \mu_{iv} \right\}, \vec{m} \right), \quad (4.64)$$

где

M — число узлов сети;

$\vec{N} = (N_1, N_2, \dots, N_v, \dots, N_V)$ — вектор распределения общего числа заявок сети по классам, $v = \overline{1, V}$;

$\{\pi^{(v)}\}$, $v = \overline{1, V}$ — набор стохастических матриц передач, в котором
 $\pi^{(v)} = \{p_{ij}^{(v)}\}$, $i, j = \overline{1, M}$ — матрица передач заявок v -го класса;
 $\{\mu_{iv}\}$, $i = \overline{1, M}$; $v = \overline{1, V}$ — множество, задающее интенсивности
 обслуживания неоднородных заявок в узлах сети, в котором μ_{iv} есть ин-
 тенсивность обслуживания заявки v -го класса в канале i -го узла сети;
 $\vec{m} = \{m_1, \dots, m_M\}$ — вектор, i -м компонентом которого является
 число каналов в i -м узле сети.

Мощность $|S|$ множества $S(\vec{N}, \vec{M})$ состояний неоднородной замк-

нутой сети определяется по правилам комбинаторики как число вариантов
 размещения заявок по узлам сети с учетом их принадлежности к разным
 классам:

$$|S| = \prod_{v=1}^V C_{N_v + M - 1}^{N_v}. \quad (4.65)$$

*Замкнутые сети
 с неоднородными по маршрутам циркуляции заявками*

Оставим для рассмотрения те дисциплины обслуживания, которые
 соответствуют четырем типам узлов, введенным в п. 4.3.1.

Тогда рассмотренный ранее алгоритм расчета однородных замкну-
 тых сетей может быть обобщен и для сетевых моделей с несколькими
 классами заявок, отличающимися маршрутами циркуляции. При этом,
 правда, значительно возрастают сложность и объем вычислений. Для ил-
 люстрации приведем основные расчетные соотношения.

Аналогично тому, как это сделано для однородных сетей, введем от-

носительные частоты ω_{iv} , удовлетворяющие системам уравнений

$$\omega_{jv} = \sum_{i=1}^M \omega_{iv} p_{ij}^{(v)}, \quad i, j = \overline{1, M}; \quad v = \overline{1, V}. \quad (4.66)$$

Состояние i -го узла сети зададим вектором $\vec{\mathbf{n}}_i = \left\{ \mathbf{n}_{iv} \right\}_{v=\overline{1, V}}$, тогда состояние сети определяется матрицей $\left\{ \mathbf{n}_{iv} \right\}_{i=\overline{1, M}; v=\overline{1, V}}$.

Возможными состояниями сети будут такие, для которых

$$\sum_{i=1}^M \mathbf{n}_{iv} = N_v, \quad v = \overline{1, V},$$

где $\mathbf{n}_{iv} \geq 0, \quad i = \overline{1, M}; \quad v = \overline{1, V}$.

Вероятность $P(\left\{ \mathbf{n}_{iv} \right\})$ состояния сети $\left\{ \mathbf{n}_{iv} \right\}$ имеет мультипликативную форму:

$$P(\left\{ \mathbf{n}_{iv} \right\}) = \frac{1}{G} \prod_{i=1}^M Z_i(n_{i1}, n_{i2}, \dots, n_{iV}) = \frac{1}{G} \prod_{i=1}^M Z_i(\vec{\mathbf{n}}_i). \quad (4.67)$$

В зависимости от типа узла i выделим три варианта вычисления сомножителей $Z_i(\vec{\mathbf{n}}_i)$:

$$1) \quad Z_i(\vec{\mathbf{n}}_i) = \frac{n_i!}{\prod_{k=1}^{n_i} \mu_i(k)} \prod_{v=1}^V \frac{\omega_{iv}^{n_{iv}}}{n_{iv}!} \text{ для многоканальных узлов типа 1}; \quad (4.68)$$

$$2) \quad Z_i(\vec{\mathbf{n}}_i) = n_i! \prod_{v=1}^V \frac{\left(\frac{\omega_{iv}}{\mu_{iv}} \right)^{n_{iv}}}{n_{iv}!} \text{ для одноканальных узлов типов 1, 2, 4}; \quad (4.69)$$

$$3) \quad Z_i \left(\vec{n}_i \right) = \prod_{v=1}^V \frac{1}{n_{iv}!} \left(\frac{\omega_{iv}}{\mu_{iv}} \right)^{n_{iv}} \quad \text{для узлов типа 3.} \quad (4.70)$$

В приведенных соотношениях

$$\vec{n}_i = \sum_{v=1}^V n_{iv}, \quad i = \overline{1, M}$$

обозначает общее число заявок всех классов в узле i , G — нормирующая константа. Отметим, что $Z_i(\mathbf{0}, \mathbf{0}, \dots, \mathbf{0}) = 1$.

Таким образом, так же как и в однородных замкнутых сетях возможность и эффективность точного анализа зависит от наличия и эффективности алгоритма вычисления нормирующей константы G . Покажем возможность использования так называемого алгоритма “свертки” для вычисления нормирующей константы.

*Использование операции свертки множеств
для расчета нормирующей константы
в замкнутых неоднородных сетях СМО*

Пусть Z_i есть множество значений функции $Z_i(n_{i1}, n_{i2}, \dots, n_{iV})$, определенных в пространстве размерности V , причем v -я координата n_{iv} может принимать значения $\mathbf{0}, \mathbf{1}, \mathbf{2}, \dots, N_v$. Мощность множества Z_i определяется вектором

$$\vec{N} = (N_1, N_2, \dots, N_v, \dots, N_V) : |Z_i| = \prod_{v=1}^V (N_v + 1).$$

Для двух множеств A и B со структурой, аналогичной структуре множества Z_i , определим свертку $C = A * B$ как множество C такой же структуры, элементы которого вычисляются по формуле

$$c(n_1, \dots, n_V) = \sum_{m_V=0}^{m_V=n_V} \cdots \sum_{m_1=0}^{m_1=n_1} a(m_1, \dots, m_V) b(n_1 - m_1, \dots, n_V - m_V). \quad (4.71)$$

Введем M множества G_i , $i = \overline{1, M}$, связанных соотношением $G_i = G_0 * Z_1 * Z_2 * \dots * Z_i$ или $G_i = G_{i-1} * Z_i$, где G_0 — множество, содержащее все нулевые элементы кроме $G_0(0, 0, \dots, 0) = 1$. Нормирующая константа G определяется аналогично рассмотренной ранее в п. 4.2.4 константе $G_M(N)$ и равна $G_M(N_1, N_2, \dots, N_V)$.

Отметим, что порядок включения множеств Z_i в свертку, приводящую к расчету $G_M(N_1, N_2, \dots, N_V)$, может быть произвольным.

Расчёт некоторых характеристик замкнутых неоднородных сетей СМО

Так же, как и для однородных сетей, теперь можно вывести ряд полезных расчетных соотношений. Например, вероятность нахождения в узле номер M ровно n_v заявок каждого класса v , $v = 1, 2, \dots, V$:

$$P_M(n_1, \dots, n_V) = Z_M(n_1, \dots, n_V) \frac{G_{M-1}(N_1 - n_1, \dots, N_V - n_v)}{G}; \quad (4.72)$$

интенсивность прохождения заявок класса v через узел с номером M :

$$\lambda_{Mv} = \omega_{Mv} \frac{G_M(N_1, N_2, \dots, N_{v-1}, N_v - 1, N_{v+1}, \dots, N_V)}{G}; \quad (4.73)$$

среднее число заявок v -го класса в M -м узле типа 1, 2 или 4:

$$\overline{n_{Mv}} = \frac{1}{G} \sum_{n_v=0}^{N_v} \cdots \sum_{n_1=0}^{N_1} n_v Z_M(n_1, \dots, n_V) G_{M-1}(N_1 - n_1, \dots, N_V - n_v), \quad (4.74)$$

а для узла типа 4 можно воспользоваться формулой Литтла:

$$\overline{n_{Mv}} = \lambda_{Mv} / \mu_i.$$

Среднее время пребывания заявок класса v в граничном M -м узле также может быть определено с помощью формулы Литтла:

$$\overline{t_{Mv}} = \overline{n_{Mv}} / \lambda_{Mv}.$$

Аналогичные характеристики для любого другого узла могут быть получены после перенумерации узлов. Вывод приведенных соотношений можно найти, например, в [5].

Таким образом, усложнение процедур численного анализа сетевых моделей с несколькими классами заявок, отличающимися маршрутами циркуляции, происходит главным образом за счет необходимости оперировать множествами, определенными на многомерном пространстве.

Пример 4.6. Приведем сначала пример, иллюстрирующий характер влияния неоднородности маршрутов циркуляции заявок на временные параметры сетевой модели. Пусть в простейшей сетевой модели со структурой, показанной на рис. 4.8, циркулируют заявки двух классов.

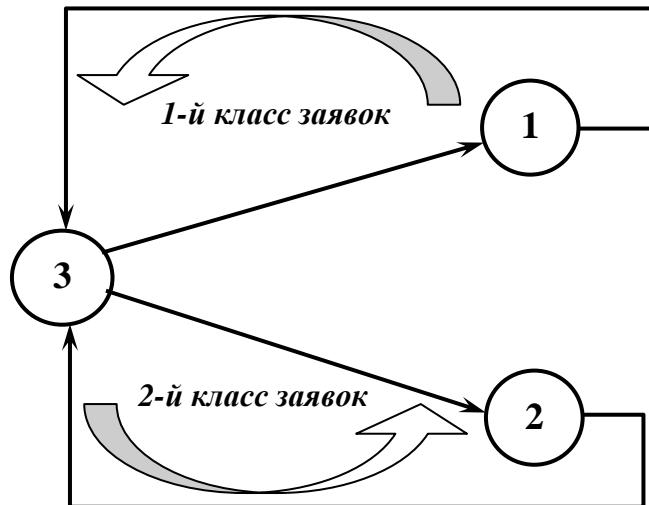


Рис. 4.8

Маршруты циркуляции заданы матрицами передач:

$$\pi^{(1)} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \pi^{(2)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Пусть во всех узлах модели заявки обслуживаются в порядке поступления, а интенсивности обслуживания μ_i , $i = 1, 2, 3$, не зависят от класса заявок, т. е. маршруты заявок разных классов пересекаются только в узле 3.

Рассмотрим случай, когда в каждом классе всего по одной заявке. Среднее время пребывания заявки v -го класса в i -м узле можно представить как сумму времени ожидания и времени обслуживания:

$$\overline{t_{iv}} = \overline{t_{iv}^{ож}} + 1/\mu_i, \quad i = 1, 2, 3, \quad v = 1, 2.$$

Для получения некоторых предельных характеристик предположим, что $\mu_1 \ll \mu_3 \ll \mu_2$. Тогда вероятность пребывания заявки 1-го класса в узле 1 стремится к единице: $q_{11} \rightarrow 1$ и соответственно $q_{31} \rightarrow 0$. Для 2-го класса можно утверждать, что $q_{32} \rightarrow 1$, а $q_{22} \rightarrow 0$. Таким образом, заявка 1-го класса в узле 3 практически отсутствует, и $\overline{t_{32}^{ож}} \rightarrow 0$. Заявка 2-го класса, напротив, почти все время находится в узле 3, и, следовательно, $\overline{t_{31}^{ож}} \rightarrow 1/\mu_3$. Теперь можно заключить, что времена пребывания заявок разных классов в узле 3 отличаются примерно в два раза: $\overline{t_{31}} \approx 2/\mu_3$, $\overline{t_{32}} \approx 1/\mu_3$, хотя заявки в этом узле обслуживаются в порядке поступления.

Приведенный пример наглядно иллюстрирует, что наличие хотя бы одной неоднородности (в данном случае неоднородности маршрутов циркуляции заявок) может привести к существенным различиям временных характеристик обслуживания заявок разных классов в узлах сети.

Пример 4.7. Этот пример служит иллюстрацией использования процедуры “свертки” при расчете характеристик сетевой модели.

Пусть структура сети та же, что и в предыдущем примере, и пусть по-прежнему в сети циркулируют две заявки разных классов в соответствии с матрицами передач $\pi^{(1)}$ и $\pi^{(2)}$. Таким образом, хотя рассматриваемая модель относительно проста, она все же является сетью с неоднородными заявками, и для ее анализа необходимо использовать наиболее общие расчетные соотношения (4.66)–(4.70).

Пусть для конкретности $\mu_1 = 1$, $\mu_2 = 0,5$, $\mu_3 = 2$. Определим в этих условиях интенсивности прохождения заявок 1-го и 2-го классов через узел 3.

Начнем расчет с нахождения относительных частот прохождения заявок через узлы сети ω_{iv} , где $i = 1, 2, 3$, $v = 1, 2$, в соответствии с (4.66). Поскольку любое решение системы (4.66) пригодно для использования, выберем конкретное значение ω_{iv} исходя из простоты дальнейших вычислений:

$$\begin{aligned}\omega_{11} &= 1, \quad \omega_{21} = 0, \quad \omega_{31} = 1; \\ \omega_{12} &= 0, \quad \omega_{22} = 1, \quad \omega_{32} = 1.\end{aligned}$$

Значения компонентов множества G_0 заданы его определением, а значения компонентов множеств Z_1 , Z_2 , Z_3 вычислим в соответствии с (4.68)–(4.70) и поместим в табл. 4.2.

Легко убедиться, что результат первой свертки $G_1 = G_0 * X_1 = X_1$, а потому перейдем к вычислению второй, а затем и третьей свертки, руководствуясь соотношением (4.71) (табл. 4.3).

Таблица 4.2

G_0	Z_1	Z_2	Z_3
$G_0(0,0)=1$	$Z_1(0,0)=1$	$Z_2(0,0)=1$	$Z_3(0,0)=1$
$G_0(0,1)=0$	$Z_1(0,1)=0$	$Z_2(0,1)=2$	$Z_3(0,1)=0,5$
$G_0(1,0)=0$	$Z_1(1,0)=1$	$Z_2(1,0)=0$	$Z_3(1,0)=0,5$
$G_0(1,1)=0$	$Z_1(1,1)=0$	$Z_2(1,1)=0$	$Z_3(1,1)=0,5$

Таблица 4.3

$G_2 = G_1 * Z_2$	$G_3 = G_2 * Z_3$
$G_2(0,0)=1$	$G_3(0,0)=1$
$G_2(0,1)=1 \cdot 2 + 0 \cdot 1 = 2$	$G_3(0,1)=1 \cdot 0,5 + 2 \cdot 1 = 2,5$
$G_2(1,0)=1 \cdot 0 + 1 \cdot 1 = 1$	$G_3(1,0)=1 \cdot 0,5 + 1 \cdot 1 = 1,5$
$G_2(1,1)=1 \cdot 0 + 1 \cdot 1 + 1 \cdot 2 + 0 \cdot 1 = 2$	$G_3(1,1)=1 \cdot 0,5 + 2 \cdot 0,5 + 1 \cdot 0,5 + 2 \cdot 1 = 4$

Таким образом, нормирующая константа G , определяемая как последний компонент множества G_3 , равна 4. Теперь можно определить исковые интенсивности, пользуясь формулой (4.73):

$$\lambda_3(1) = \frac{\omega_3(1)G_3(0,1)}{G} = \frac{1 \cdot 2,5}{4} = 0,625 \frac{1}{c};$$

$$\lambda_3(2) = \frac{\omega_3(2)G_3(1,0)}{G} = \frac{1 \cdot 1,5}{4} = 0,375 \frac{1}{c}.$$

В заключение примера заметим, что увеличение числа узлов сетевой модели, числа классов заявок, количества самих заявок не внесет никаких принципиальных затруднений в вычисления, так как основное расчетное соотношение свертки двух множеств легко поддается программированию на ЭВМ.

4.4. Приближенные методы расчета немарковских сетей СМО

Рассмотренные ранее сети СМО, так называемые сети Джексона, предполагающие показательные законы обслуживания в узлах и простейшие потоки на входе (для разомкнутых сетей), относительно несложны в анализе, поскольку порождают непрерывные марковские цепи, являющиеся теоретической основой элементарной теории сетей СМО.

При невыполнении требований следования показательным законам обслуживания и поступления заявок модели в виде систем СМО по-прежнему актуальны и широко используются, но уже не могут быть объектами приложения к ним теории сетей Джексона. Будем называть такие сети СМО немарковскими, подчеркивая тем самым не только отличие упомянутых законов от показательных, но и тот факт, что граф переходов такой сети из состояния в состояние уже не является марковским.

Принудительная подмена немарковской сети СМО сетью Джексона с теми же интенсивностями обслуживания и поступления заявок (но с другими, естественно, равными единице, коэффициентами вариации) может обусловить существенные погрешности при определении показателей эффективности исследуемых сетей СМО. Именно поэтому немарковские сети СМО предполагают использование собственных методов расчета, в той

или иной степени приближенных в отличие от упомянутых методов точной теории сетей Джексона.

Дальнейшее рассмотрение методов расчета немарковских сетей СМО ограничим классом разомкнутых сетей, подход к анализу которых рассмотрен в значительном числе источников [6–9].

Расчет разомкнутых немарковских сетей СМО предполагает декомпозицию — выделение отдельных узлов сети с назначением каждому из них соответствующего входного потока. Немарковские сети, как и сети Джексона, являются линейными, поэтому интенсивности на входах в узлы, не зависящие от законов поступления и обслуживания, рассчитываются точно так же. В сетях Джексона с узлами $M/M/m$ все входные потоки в узлах — пуассоновские (с единичными коэффициентами вариации). В немарковской сети приходится проводить расчет коэффициентов вариации входных потоков в узлах.

На этом этапе осуществляется формирование исходных данных для расчета локальных показателей эффективности в узлах немарковской сети, таких, например, как среднее время пребывания заявки, среднее время ожидания или среднее число заявок в узле. С помощью универсального правила Литтла, справедливого для систем, находящихся в установившемся режиме, рассчитываются интегральные показатели эффективности немарковской сети: среднее время пребывания и среднее число заявок в сети.

Далее дается описание приближенного двухмоментного метода расчета разомкнутых немарковских сетей, основанного на использовании результатов, представленных в [6]³.

³ Термин "двохмоментный" означает и подчеркивает, что входные потоки и законы обслуживания задаются только двумя моментами: математическим ожиданием и дисперсией.

4.4.1. Расчет показателей разомкнутых однородных одноканальных немарковских сетей СМО

1. Дадим формальное задание однородной одноканальной немарковской сети СМО (объединяющей узлы типа G/G/1). Сеть является однородной в связи с одинаковым обслуживанием заявок в узлах и единой маршрутизацией. Сеть задается следующим кортежем параметров:

$$\left\{ M; \lambda_0; ca_0; Q = \left\{ q_{ij} \right\}_{\substack{i=0, M \\ j=0, M}}; \left\{ \mu_i \right\}_{i=1, M}; \left\{ cs_i \right\}_{i=1, M} \right\},$$

где

M — число узлов;

λ_0 — интенсивность входного потока;

ca_0 — квадрат коэффициента вариации длительности интервала между заявками, поступающими на вход сети;

$Q = \left\{ q_{ij} \right\}_{\substack{i=0, M \\ j=0, M}}$ — матрица передач размерности $(M+1) \times (M+1)$,

определяющая маршрутизацию заявок;

$\{\mu_i\}_{i=1, M}$ — интенсивности обслуживания заявок в узлах сети;

$\{cs_i\}_{i=1, M}$ — квадраты коэффициентов вариации длительностей об-

служивания заявок в узлах сети;

$\lambda_{0j} = \lambda_0 q_{0j}$, а $ca_{0j} = ca_0$, $j : q_{0j} > 0$ (данное соотношение очевидным образом следует из того факта, что в j -й узел поступает равная q_{0j} доля внешнего потока с тем же коэффициентом вариации ca_0).

2. Расчет интенсивностей потоков на входах в узлах сети:

$$\lambda_j = \sum_{i=0}^M \lambda_i q_{ij}, \quad j = \overline{0, M}, \quad (4.75)$$

$$\lambda_{ij} = \lambda_i q_{ij}, \quad i = \overline{0, M}, \quad j = \overline{0, M}. \quad (4.76)$$

3. Расчет частот $\alpha_j = \frac{\lambda_j}{\lambda_0}, \quad j = \overline{1, M}$ (4.77)

4. Расчет $ca_j, \quad j = \overline{1, M}$, квадратов коэффициентов вариации потоков на входах в узлы сети осуществляется в результате решения следующей системы уравнений:

$$ca_j = w_j \sum_{i=0}^M \frac{\lambda_{ij}}{\lambda_j} ca_{ij} + 1 - w_j, \quad (4.78)$$

где

$$ca_{ij} = q_{ij} (\rho_i^2 cs_i + (1 - \rho_i^2) ca_i) + 1 - q_{ij};$$

$$w_j = \frac{1}{1 + 4(1 - \rho_j)^2(v_j - 1)}; \quad v_j = \frac{1}{\sum_{i=0}^M \left(\frac{\lambda_{ij}}{\lambda_j} \right)^2}; \quad \rho_j = \frac{\lambda_j}{\mu_j}.$$

5. Расчет математических ожиданий $E(W_j), \quad j = \overline{1, n}$ времен ожидания обслуживания в узлах G/G/1 сети:

$$E(W_j) = \frac{\rho_j (ca_j + cs_j) g(\rho_j, ca_j, cs_j)}{2\mu_j(1 - \rho_j)}, \quad (4.79)$$

где

$$g(\rho_j, ca_j, cs_j) = \begin{cases} \exp \left\{ \frac{-2(1 - \rho_j)(1 - ca_j)^2}{3\rho_j(ca_j + cs_j)} \right\}, & ca_j < 1; \\ 1, & ca_j \geq 1. \end{cases}$$

Формула (4.79) является эвристической, но для узла типа M/M/1 становится точной, в чем нетрудно убедиться, обратив внимание на равенство $ca_j = cs_j = g(\cdot) = 1$ в этом случае.

6. Расчет математического ожидания времени $E(T)$ пребывания

заявки в сети

$$E(T) = \sum_{j=1}^M \alpha_j \left(E(W_j) + \frac{1}{\mu_j} \right). \quad (4.80)$$

Другие реализации подобного метода изложены в гл. 6 "Приближенные методы расчета открытых сетей по двум моментам" в книге Г. П. Башарина и др. [7].

4.4.2. Расчет показателей разомкнутых однородных многоканальных немарковских сетей СМО

1. Сеть объединяет узлы G/G/m, и к кортежу исходных данных добавляется вектор $\{\mathbf{m}_i\}_{i=\overline{1, M}}$, задающий число каналов в узлах сети.

Метод расчета среднего времени пребывания заявки в сети строится аналогично, но с некоторым изменением системы уравнений и формул предыдущего раздела. При этом по-прежнему

$$\lambda_{0j} = \lambda_0 q_{0j}, \text{ а } ca_{0j} = ca_0, \quad j : q_{0j} > 0.$$

2. Расчет ca_j , $j = \overline{1, M}$, квадратов коэффициентов вариации потоков на входах в узлы сети осуществляется в результате решения системы уравнений:

$$ca_j = \alpha_j + \sum_{i=1}^M ca_i \beta_{ij}, \quad j = \overline{1, M}, \quad (4.81)$$

где

$$\alpha_j = 1 + w_j \left\{ (p_{0j} ca_{0j} - 1) + \sum_{i=1}^M p_{ij} \left[(1 - q_{ij}) + q_{ij} \rho_i^2 x_i \right] \right\};$$

$$\beta_{ij} = w_j p_{ij} q_{ij} (1 - \rho_i^2);$$

$$w_j = \frac{1}{1 + 4(1 - \rho_j)^2 (v_j - 1)}; \quad v_j = \frac{1}{\sum_{i=0}^M \left(\frac{\lambda_{ij}}{\lambda_j} \right)^2};$$

$$p_{ij} = \frac{\lambda_i}{\lambda_j} q_{ij} = \frac{\lambda_{ij}}{\lambda_j}; \quad x_i = 1 + \frac{\max\{cs_i, 0.2\} - 1}{\sqrt{m_i}}.$$

3. Расчет математических ожиданий $E(W_j)$, $j = \overline{1, M}$, времен ожидания обслуживания в узлах G/G/m сети:

$$E(W_j) = \frac{(ca_j + cs_j)}{2} E(W_j(M / M / m_j)), \quad (4.82)$$

где $E(W_j(M / M / m_j))$ — среднее время ожидания обслуживания в узле типа $M / M / m_j$, рассчитываемое по формуле (см. табл. 2.5).

4. Расчет математического ожидания времени $E(T)$ пребывания заявки в сети производится так же, как и ранее по формуле (4.80).

4.4.3. Расчет показателей разомкнутых неоднородных многоканальных ненарковских сетей СМО

1. Сеть объединяет узлы G/G/m и задается следующим кортежем параметров⁴:

$$\left\{ M; \left\{ \lambda_0^{(r)} \right\}_{r=\overline{1, R}}; \left\{ ca_0^{(r)} \right\}_{r=\overline{1, R}}; \left\{ Q^{(r)} \right\}_{r=\overline{1, R}}; \left\{ m_i \right\}_{i=\overline{1, M}}; \left\{ \mu_{ir} \right\}_{r=\overline{1, R}, i=\overline{1, M}}; \left\{ cs_{ir} \right\}_{r=\overline{1, R}, i=\overline{1, M}} \right\},$$

где

M — число узлов; R — число классов;

$\left\{ \lambda_0^{(r)} \right\}_{r=\overline{1, R}}$ — интенсивности входных потоков разных классов;

$\left\{ ca_0^{(r)} \right\}_{r=\overline{1, R}}$ — квадраты коэффициентов вариации длительностей

интервалов входных потоков разных классов;

⁴ Излагается интерпретация метода, приведенного в [6]. Стогое следование [6] требует детерминированной маршрутизации заявок разных классов, означающей, что элементами матриц $\left\{ Q^{(r)} \right\}_{r=\overline{1, R}}$ могут быть только нули и единицы.

$$\left\{ \mathbf{Q}^{(r)} \right\}_{r=\overline{1,R}} = \left\{ q_{ij}^{(r)} \right\}_{\substack{i=\overline{0,M} \\ j=\overline{0,M} \\ r=\overline{1,R}}} \quad \text{— матрицы передач размерности}$$

$(M+1) \times (M+1)$, определяющие маршрутизацию заявок разных классов;

$\{m_i\}_{i=\overline{1,M}}$ — числа каналов в узлах сети;

$\{\mu_{ir}\}_{\substack{i=\overline{1,M} \\ r=\overline{1,R}}}$ — интенсивности обслуживания заявок разных классов в

узлах сети;

$\{cs_{ir}\}_{\substack{i=\overline{1,M} \\ r=\overline{1,R}}}$ — квадраты коэффициентов вариации длительностей об-

служивания заявок разных классов в узлах сети.

Аналогично тому, как это имеет место в однородных сетях СМО,

$$\lambda_{0j}^{(r)} = \lambda_0^{(r)} q_{0j}^{(r)}, \quad \text{а} \quad ca_{0j}^{(r)} = ca_0^{(r)}, \quad j : q_{0j}^{(r)} > 0.$$

2. Составим и решим системы уравнений для нахождения интенсивностей потоков заявок в узлах.

Для каждого $r = \overline{1, R}$ найдем $\lambda_j^{(r)} = \sum_{i=0}^M \lambda_i^{(r)} q_{ij}^{(r)}, \quad j = \overline{1, M}$.

Затем найдем $\left\{ \lambda_{ij}^{(r)} \right\}_{\substack{i=\overline{1,M} \\ j=\overline{1,M}}}$ — интенсивности потоков разных классов, поступающих из узла i в узел j :

$$\lambda_{ij}^{(r)} = \lambda_i^{(r)} q_{ij}^{(r)}.$$

Теперь найдем $\lambda_j = \sum_{r=1}^R \lambda_j^{(r)}$.

3. Дальнейший расчет неоднородной немарковской сети в соответствии с методикой, изложенной в [6], предполагает превращение неоднородной сети в однородную с единой матрицей передач, осредненными интенсивностями обслуживания в узлах и остальными параметрами, также подвергнутыми осреднению. Переход к однородной сети СМО производится

по формулам (4.83)–(4.87).

$$q_{ij} = \frac{\sum_{r=1}^R \lambda_{ij}^{(r)}}{\lambda_i}, \quad i = \overline{0, M}, \quad j = \overline{1, M}; \quad (4.83)$$

$$\frac{1}{\mu_j} = \frac{\sum_{r=1}^R \left(\frac{1}{\mu_{jr}} \right) \lambda_j^{(r)}}{\lambda_j}, \quad j = \overline{1, M}; \quad (4.84)$$

$$cs_j = \frac{\sum_{r=1}^R \lambda_j^{(r)} \left(\frac{1}{\mu_{jr}} \right)^2 (cs_{jr} + 1)}{\lambda_j \left(\frac{1}{\mu_j} \right)^2} - 1, \quad j = \overline{1, M}; \quad (4.85)$$

$$ca_{0j} = w_j \sum_{r=1}^R \frac{ca_0^{(r)} \lambda_0^{(r)} q_{0j}^{(r)}}{\sum_{r=1}^R \lambda_0^{(r)} q_{0j}^{(r)}} + 1 - w_j; \quad w_j = \frac{1}{1 + 4(1 - \rho_j)^2 (v_j - 1)}; \quad (4.86)$$

$$v_j = \frac{1}{\sum_{r=1}^R \left(\frac{ca_0^{(r)} \lambda_0^{(r)} q_{0j}^{(r)}}{\sum_{r=1}^R \lambda_0^{(r)} q_{0j}^{(r)}} \right)^2}. \quad (4.87)$$

Расчеты по формулам (4.86)–(4.90) позволяют получить массив исходных данных $\left\{ \left\{ q_{ij} \right\}_{i=\overline{0, M}}; \left\{ \frac{1}{\mu_j}; cs_j; ca_{0j}; w_j; v_j \right\}, \quad j = \overline{1, M} \right\}$, необходимый для расчета однородной немарковской сети.

4. Расчет математических ожиданий $E(W_j)$, $j = \overline{1, M}$, и математического ожидания времени $E(T)$ пребывания заявки в сети производится так же, как и ранее, по формулам (4.82) и (4.80).

Задачи

Задача 1. Задана сеть массового обслуживания (рис. 4.9).

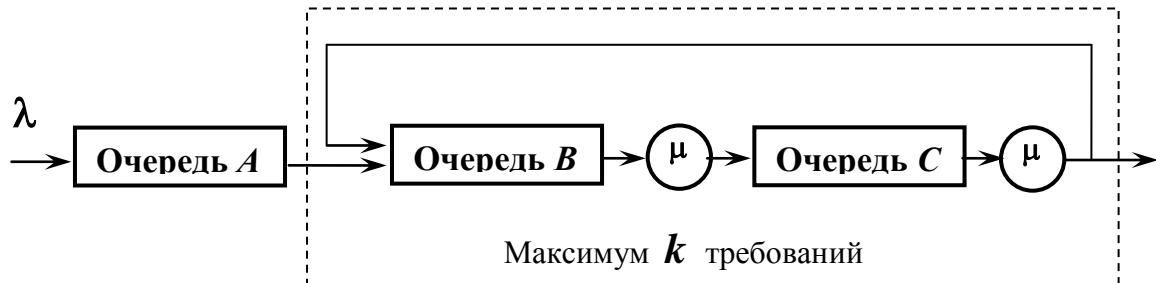


Рис. 4.9

На входе сети – пуассоновский поток с интенсивностью λ . При поступлении требование проходит в блок, выделенный штриховой линией, только если число таких требований меньше k . В противном случае оно становится в очередь A , из начала которой требования поступают в штриховой блок.

Поступив в штриховой блок, требование становится в очередь B и со временем в соответствии с дисциплиной "первым пришел – первым обслужен" получает обслуживание с показательным распределением времени и интенсивностью μ .

Затем оно переходит в очередь C и получает аналогичное обслуживание на независимом приборе с такими же характеристиками. После этого требование покидает штриховой блок (и систему) с вероятностью P или возвращается в очередь B с вероятностью $(1 - P)$. Предполагая, что в системе существует установившийся режим:

- найти среднее число требований во всей системе (в системе "очередь + штриховой блок");
- найти среднее время \bar{T} пребывания требования во всей системе для значений k , равных 1 и ∞ ;

- в) найти условия стационарности для значений k , равных 1 и ∞ ;
 г) доказать, что $\bar{T}(k=1) \geq \bar{T}(k=\infty)$ для общей области стационарности.

Задача 2. Задана сеть массового обслуживания с M узлами и матрицей передач R , ($i, j = \overline{1, M}$). Требование из внешнего источника с интенсивностью λ_0 поступает в узел 1 и покидает сеть с вероятностью $1 - \alpha_M$.

Найти следующие показатели системы:

- а) трафик λ_i , $i = \overline{1, M}$, каждого узла сети в явном виде;

$$R = \begin{pmatrix} \alpha_1 & 1 - \alpha_1 & 0 & 0, & \dots, & 0 \\ \alpha_2 & 0 & 1 - \alpha_2 & 0, & \dots, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha_{M-1} & 0 & 0 & 0, & \dots, & 1 - \alpha_{M-1} \\ \alpha_M & 0 & 0 & 0, & \dots, & 0 \end{pmatrix};$$

- б) предполагая, что все узлы являются одноканальными с интенсивностями обслуживания μ_i , $i = \overline{1, M}$, определить:

условия установившегося режима;

соотношение между μ_i , обеспечивающее один и тот же показатель загрузки обслуживающего прибора;

- в) характеристики сети при условии

$$\alpha_i = \alpha, \quad \mu_i = \mu, \quad i = \overline{1, M}.$$

Задача 3. Задана замкнутая сеть массового обслуживания с пятью узлами, $M = 5$, где i -й узел имеет m_i обслуживающих приборов, каждый

из которых требует $\frac{1}{\mu_i}$ с обслуживания при показательном законе распределения. Пусть $m_5 = M$, $m_1 = m_2 = m_3 = m_4 = 1$ и пятый узел является терминальным. Сеть содержит M требований, циркулирующих в соответствии с матрицей передач

$$R = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Пусть $\mu_1 = 3c^{-1}$, $\mu_2 = 2c^{-1}$, $\mu_3 = \mu_4 = 3c^{-1}$, $\mu_5 = (2/17)c^{-1}$.

Требуется:

- а) построить граф сети;
- б) определить отношения λ_i / λ_5 для $i = \overline{1, 4}$;
- в) определить наиболее нагруженный узел;
- г) определить среднее время T ответа (от момента ухода требования из пятого узла до момента возвращения).

Задача 4. Задана разомкнутая сеть массового обслуживания, включающая четыре узла и источник с интенсивностью $\lambda_0 = 0,25 c^{-1}$. В дальнейшем источник будем считать нулевым узлом. Матрица передач

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0,02 & 0 & 0,18 & 0,8 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix};$$

$$m_1 = 1, \quad m_2 = m_3 = m_4 = 2,$$

где m_i — число каналов в i -м узле, а интенсивности обслуживания в узлах:

$$\mu_1 = 20 \text{ c}^{-1}, \quad \mu_2 = 1 \text{ c}^{-1}, \quad \mu_3 = 7 \text{ c}^{-1}, \quad \mu_4 = 9 \text{ c}^{-1}.$$

Требуется:

- а) выяснить, может ли данная сеть работать в установившемся режиме;
- б) определить наиболее нагруженный узел;
- в) определить среднее время реакции (среднее время пребывания требования в сети).

Задача 5. Задана разомкнутая сеть массового обслуживания (рис. 4.10).

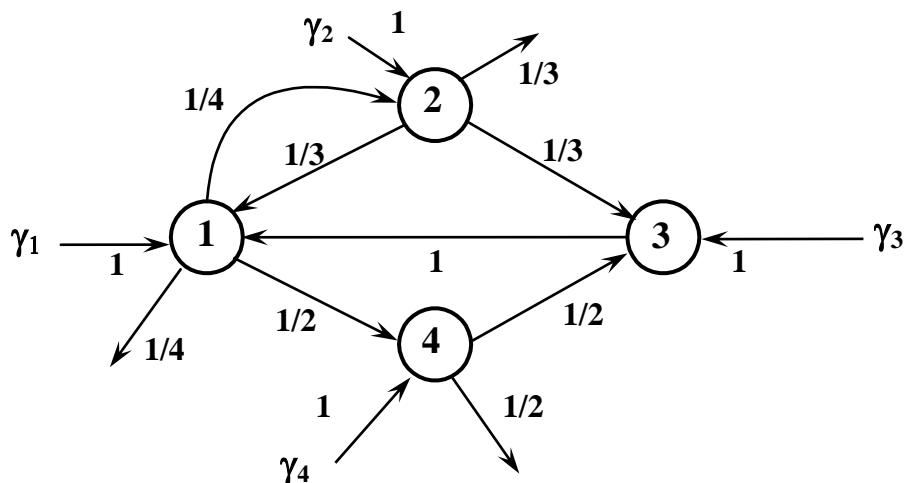


Рис. 4.10

Интенсивности γ_i внешних источников требований, поступающих в сеть, равны: $\gamma_1 = 1, \quad \gamma_2 = 0, \quad \gamma_3 = 5, \quad \gamma_4 = 2$.

Требуется:

- а) найти интенсивность требований для i -го узла;

б) найти характеристики узлов при следующих условиях:

$$m_1 = m_3 = 1, \quad m_2 = m_4 = 2;$$

$$\mu_1 = 20 \text{ c}^{-1}, \quad \mu_2 = 2 \text{ c}^{-1}, \quad \mu_3 = 15 \text{ c}^{-1}, \quad \mu_4 = 5 \text{ c}^{-1};$$

в) определить среднее время пребывания требований в сети для различных внешних источников.

Задача 6. Задана разомкнутая неоднородная сеть с числом узлов, равным трем, $M = 3$ и таким же числом $L = 3$ внешних источников требо-

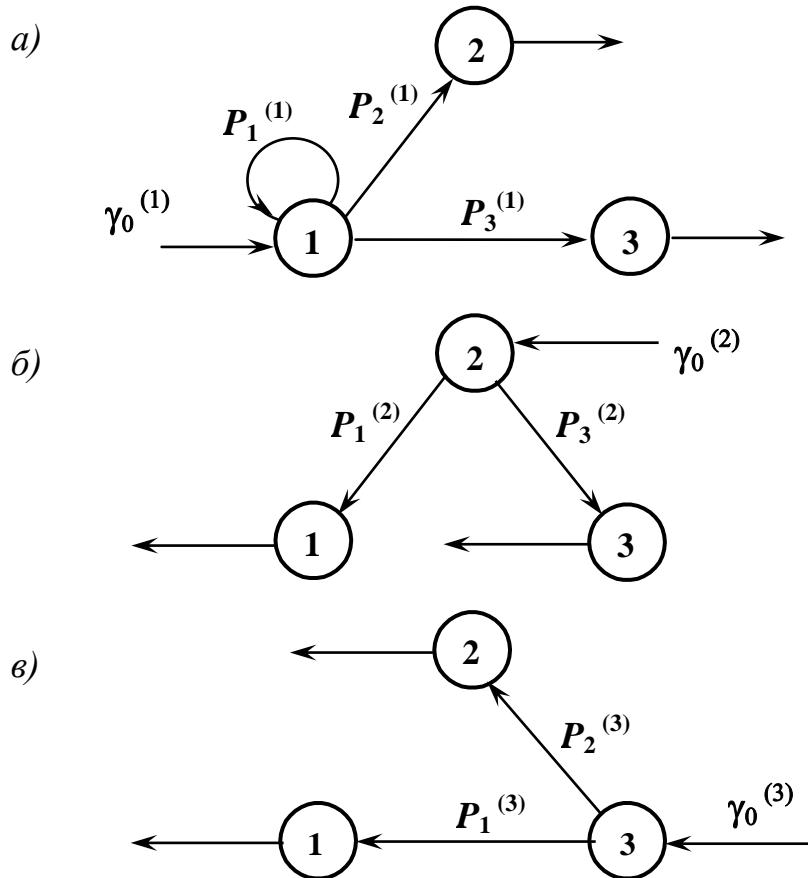


Рис. 4.11

ваний, трафик которых представлен на рис. 4.11.

Предполагается равновероятное распределение требований в трафике:

$$P_j^{(1)} = 1/3, \quad j = \overline{1, 3} — (\text{вариант } a, \text{ см. рис. 4.11, } a);$$

$$P_j^{(2)} = P_j^{(3)} = 1/2, \quad j=1, 2 \quad (\text{варианты } \delta \text{ и } \varepsilon, \text{ см. рис. 4.11,} \delta \text{ и } \varepsilon).$$

Выполнить следующее задание:

- 1) найти интенсивность неоднородных потоков заявок в узлах сети и сформулировать условия установившегося режима в сети с одноканальными узлами;
- 2) для заданных векторов интенсивностей обслуживания в узлах:

$$[\mu^{(1)}]^T = (1, 2, 3);$$

$$[\mu^{(2)}]^T = (3, 2, 1);$$

$$[\mu^{(3)}]^T = (1, 3, 2)$$

найти интенсивности входных потоков $\gamma_0^{(i)}, i = \overline{1, 3}$, максимизирующих суммарную пропускную способность при установленном режиме в сети;

- 3) проанализировать и сравнить следующие дисциплины обслуживания в узлах сети:

бесприоритетное обслуживание в порядке поступления,

обслуживание с относительными приоритетами для всех **3!** возможных вариантов их назначения.

Задача 7. Задана сеть массового обслуживания, включающая три узла, $M = 3$. Число каналов обслуживания в узлах определяется вектором $m^T = (1 \ 1 \ 1)$, интенсивности обслуживания — вектором

$$\mu^T = (2 \text{ c}^{-1}; \ 0,8 \text{ c}^{-1}; \ 0,1 \text{ c}^{-1}).$$

В сети циркулируют N заявок в соответствии с матрицей передач

$$R = \begin{pmatrix} 0,1 & 0,4 & 0,5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Требуется:

- определить характеристики узлов и сети в целом ($N = 3$);
- сопоставить характеристики узлов указанной сети с сетью, где узлы 2 и 3 объединены в один узел с числом каналов, равным двум и усредненной интенсивностью $0,45 \text{ c}^{-1}$.

Задача 8. Рассматривается сетевая модель вычислительной системы, представленная на рис. 4.12, где узел **1** моделирует обслуживание пакета N программ в процессоре, а узлы **2** и **3** — обслуживание в различных ступенях внешней памяти. Параметры узлов:

$$\begin{aligned} m_1 &= 1, \quad m_2 = 1, \quad m_3 = 1; \\ \mu_1 &= 0,5 \text{ c}^{-1}; \quad \mu_2 = 0,6 \text{ c}^{-1}; \quad \mu_3 = 1 \text{ c}^{-1}. \end{aligned}$$

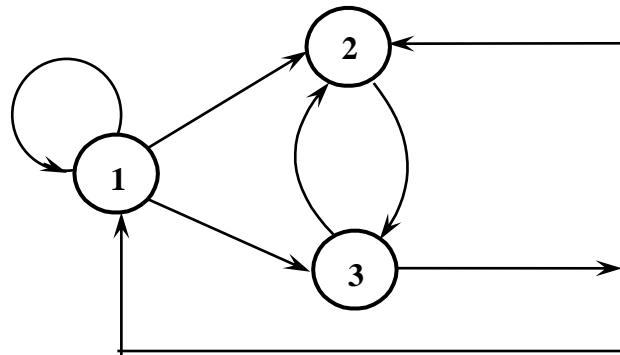


Рис. 4.12

Число программ N равно **2**. Заданы следующие варианты структуры сети с соответствующими матрицами передач:

Вариант 1

Вариант 2

$$P = \begin{pmatrix} 0,2 & 0,3 & 0,5 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; \quad P = \begin{pmatrix} 0,2 & 0,8 & 0 \\ 0,37 & 0 & 0,63 \\ 0 & 1 & 0 \end{pmatrix};$$

Вариант 3

$$P = \begin{pmatrix} 0 & 0,5 & 0,5 \\ 1 & 0 & 0 \\ 0,4 & 0,6 & 0 \end{pmatrix};$$

Вариант 4

$$P = \begin{pmatrix} 0,35 & 0,55 & 0,1 \\ 0,5 & 0 & 0,5 \\ 0,5 & 0,5 & 0 \end{pmatrix}.$$

Требуется провести сравнительный анализ различных структур:

- a) по следующей системе показателей: коэффициент загрузки узла и среднее время пребывания требования в узле;
- b) при условии, что $N = 0$, а в 1-й узел поступает поток внешних заявок с интенсивностью λ_0 ; значение λ_0 выбрать исходя из условия равенства **0,8** коэффициента загрузки 1-го узла при наличии установившегося режима в сети. Матрицу передач разомкнутой сети сформировать так, чтобы там, где имел место возврат заявок в 1-й узел замкнутой сети, теперь был бы обеспечен возврат заявок в 0-й узел разомкнутой сети.

Задача 9. Рассматривается вычислительная система с перекрестными связями, состоящая из равного числа n процессоров и модулей оперативной памяти (рис. 4.13).

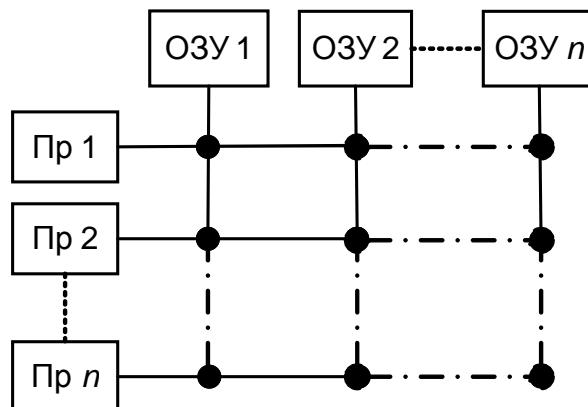


Рис. 4.13

Обращение процессора к любому модулю памяти равновероятно. Потоки запросов на ОЗУ имеют интенсивность λ_i , обслуживание проис-

ходит с интенсивностью μ_i , ($i = \overline{1, n}$).

Выполнить следующее задание:

определить характеристики занятости магистралей и задержки в выполнении программ;

определить зависимость вероятности простоя из-за занятости магистралей в зависимости от изменения параметра λ_i на $\pm 50\%$ при постоянных значениях μ_i , $i = \overline{1, n}$, и от изменения параметра μ_i на $\pm 50\%$ при постоянных значениях λ_i , $i = \overline{1, n}$;

сравнить указанные зависимости с такими же при наличии одной магистрали обмена при тех же условиях, исходные данные приведены в табл. 4.4.

Т а б л и ц а 4. 4

Вариант	n	λ_1	λ_2	λ_3	μ_1	μ_2	μ_3
1	2	2	3	–	2,5	3	–
2	2	10	15	–	20	30	–
3	3	2	3	5	3	4	8
4	3	20	30	25	50	30	40

Задача 10. Оценка эффективности управления трафиком в виртуальном канале с помощью механизма "скользящего окна". Виртуальный канал представляет собой последовательное объединение нескольких каналов с разной пропускной способностью.

Пакеты передаются от источника получателю по виртуальному каналу. Пакет считается успешно принятым на стороне получателя при условии доставки источнику подтверждающей квитанции.

Как только число неподтвержденных пакетов становится равным

пределльному значению N (размеру "окна"), виртуальный канал блокируется для передачи новых пакетов. Методология моделирования трафика в виртуальном канале рассмотрена в [10].

Требуется:

- предложить модель трафика в виртуальном канале;
- определить вероятность блокировки виртуального канала и среднее время передачи пакета;
- определить "узкое место" в виртуальном канале.

Варианты заданий приведены в табл. 4.5.

Т а б л и ц а 4.5

Номер задания	Интенсивность источника λ_0	Размер "окна" N	Состав виртуального канала		
			Канал 1 μ_1	Канал 2 μ_2	Канал 3 μ_3
1	40	15	70	120	130
2	50	20	90	130	110
3	60	10	160	150	140
4	70	25	180	170	150
5	80	30	200	210	190

5. ОПТИМИЗАЦИЯ СЕТЕЙ СМО

5.1. Оптимизация разомкнутых сетей СМО

5.1.1. Постановка задачи оптимизации разомкнутой сети СМО

Сеть СМО — формальная модель, адекватно представляющая функционирование дискретных систем различного назначения: вычислительных, производственных, технических и других. Сеть СМО, объединяющую некоторое число узлов, можно интерпретировать как Исполнителя задачий, поступающих на вход сети, маршрут прохождения заявок по узлам сети — как реализуемую последовательность выполнения работ, а интегральные показатели, такие как среднее число заявок в сети, — как количественную оценку эффективности использования ресурсов Исполнителя.

В любой сфере деятельности задача оптимизации системы является более сложной, чем задача анализа — оценки её эффективности. При этом обе задачи взаимосвязаны, поскольку каждая из них опирается на использование единой формальной модели системы. Это утверждение в полной мере относится к проблеме оптимизации сети СМО.

Рассмотренные ранее методы расчета сетей СМО (раздел 4) были ориентированы на решение задач их анализа. В то же время полученные результаты, намечают пути решения задачи оптимизации сети СМО. Поиск конструктивных методов оптимального выбора параметров сети СМО является актуальной задачей именно потому, что эти методы могут быть распространены на содержательные задачи оптимизации параметров реальных систем.

Рассмотрим формальные постановки задач оптимизации разомкнутых сетей СМО. Воспользуемся предложенной в [22] интерпретацией сети

СМО с тем, чтобы придать содержательный смысл возможным постановкам задач их оптимизации. Сеть СМО — производственный комплекс, узел сети — отдельное предприятие комплекса, характеризуемое производительностью и номенклатурой (типами) выпускаемой продукции. Предприятие (узел) объединяет некоторое подразделений (каналов) с одинаковой производительностью. Существует цикл выпуска продукции соответствующего типа, заданный в виде маршрута прохождения требования на выпуск единицы продукции, поступающего на вход производственной системы. На выходе системы появляются экземпляры готовой продукции. Эффективность производственного комплекса тем выше, чем меньше объем незавершенной работы, присутствующий в нем. Стоимость комплекса является аддитивной функцией стоимостей входящих в нее предприятий. Стоимость предприятия является функцией его производительности. Аналогичную интерпретацию рассматриваемой задачи несложно предложить и для других объектов, например, для вычислительной системы, моделируемой на системном уровне, или для компьютерной сети.

Начнем с пары очевидных двойственных задач оптимизации, в которых целевая функция и ограничение меняются местами. Каждую из них сформулируем сначала как задачу оптимизации производственной системы, а затем, как формальную задачу оптимизации сети СМО. Кроме того, снабдим постановки задач (табл. 5.1 и 5.2), предложенной в [22] кодировкой. Условные обозначения параметров оптимизационных задач, приведенных в табл. 5.1, 5.2 поясняются в табл. 5.3.

Задачи SP.1.1 и SP.2.1 сведены к задачам оптимизации однородных разомкнутых сетей СМО, что применительно к производственному комплексу означает его ориентацию на выпуск однотипных изделий. Узлы сети СМО, как уже отмечалось, являются подразделениями предприятий

комплекса. Каждое из подразделений j -го предприятия характеризуется одинаковой производительностью μ_j .

Таблица 5.1

Содержательные постановки задач оптимизации

производственного комплекса

Задача SP.1.1	
Минимизируемая целевая функция	Стоимость производственного комплекса
Переменные задачи	производительность каждого предприятия, входящего в комплекс
Ограничение	допустимое значение величины незавершенной работы
Задача SP.2.1	
Минимизируемая целевая функция	значение величины незавершенной работы
Переменные задачи	производительность каждого предприятия, входящего в комплекс
Ограничение	Стоимость производственного комплекса
Задача SP.3.1	
Минимизируемая целевая функция	Стоимость производственного комплекса
Переменные задачи	Число предприятий, номенклатура изделий, выпускаемых каждым предприятием, производительность каждого предприятия, входящего в комплекс по каждому типу производимых изделий
Ограничение	Ограничения по номенклатуре выпускаемых изделий для каждого предприятия, допустимое значение величины незавершенной работы

Таблица 5.2

Постановки задачи оптимизации сети СМО

Задача SP.1.1	
Минимизируемая целевая функция	$\min \sum_{j=1}^M F_j(\mu_j, m_j)$
Переменные задачи	$\{m_j\}, \{\mu_j\}, j = \overline{1, M}$
Ограничение	$\sum_{j=1}^M v_j L_j(\mu_1, m_1; \mu_2, m_2; \dots; \mu_M, m_M) \leq L_T$
Задача SP.2.1	
Минимизируемая целевая функция	$\min \sum_{j=1}^M v_j L_j(\mu_1, m_1; \mu_2, m_2; \dots; \mu_M, m_M)$
Переменные задачи	$\{m_j\}, \{\mu_j\}, j = \overline{1, M}$
Ограничение	$\sum_{j=1}^M F_j(\mu_j, m_j) \leq F_T$
Задача SP.3.1	
Минимизируемая целевая функция	$\min \sum_{j=1}^M \left(F_j \left\{ \mu_{jr}, m_j \right\}_{r \in PM_j} \right)$
Переменные задачи	$M, \left\{ \mu_{jr}, m_j \right\}_{\substack{j=\overline{1, M} \\ r \in PM_j}}, \{PM_j\}, j = \overline{1, M}$
Ограничение	$\sum_{j=1}^M \sum_{r \in PM_j} v_{jr} L_{jr} \leq L_T$

Стоимость производственного комплекса есть сумма стоимостей его предприятий. За каждую находящуюся в сети заявку комплексу приходится платить штраф. Если последний равен 1, то величина суммарного штрафа становится равной среднему числу заявок в сети. Определенное упрощение исходных формулировок задач SP.1.1 и SP.2.1 следует рассматривать как плату за построение конструктивной модели исследуемого объекта.

та в виде однородной сети СМО.

Таблица 5.3

Условные обозначения параметров задач SP.1.1, SP.2.1, SP.3.1

M	Число узлов сети
μ_j	Интенсивность обслуживания в j -м узле
m_j	Число каналов в j -м узле
$F_j(\mu_j, m_j)$	Стоимость j -го узла, зависящая от интенсивности обслуживания в канале и числа каналов в узле
L_j	Среднее число заявок в j -м узле
L_T	Предельно допустимое значение штрафа за наличие заявок в сети
F_T	Верхнее граничное значение стоимости сети
v_j	Штраф за наличие одной заявки
μ_{jr}	Интенсивность обслуживания заявок r -го класса в j -м узле
PM_j	Множество классов заявок, обрабатываемых j -м узлом
v_{jr}	Штраф за наличие одной заявки r -го класса в j -м узле
L_{jr}	Среднее число заявок r -го класса в j -м узле

Задача SP.3.1 значительно сложнее, чем две предыдущие. Здесь гораздо большее число степеней свободы — переменных решения, выбор которых определяет значение целевой функции. Смысл этой задачи состоит в том, чтобы найти оптимальные значения следующих параметров сети СМО. Во-первых, числа M узлов сети (числа отдельных предприятий, входящих в производственный комплекс). Во-вторых, каждому j -му предприятию требуется определить номенклатуру выпускаемой продукции, то есть каждому j -му узлу — назначить множество PM_j классов обрабатываемых узлом заявок. В-третьих, найти число m_j подразделений (каналов) каждого j -го предприятия (j -го узла) и назначить им производительности μ_{jr} . Выбранные значения параметров неоднородной сети

СМО должны минимизировать целевую функцию — стоимость сети СМО — при ограничении на величину штрафа, зависящего от количества незавершенной работы, присутствующей в производственном комплексе. В итоге получается довольно объемная задача оптимизации, которую при необходимости можно усложнять, увеличивая количество переменных решения, но оставаясь при этом в рамках модели в формате сети СМО.

5.1.2. Классификация задач оптимизации разомкнутых сетей СМО

Переход от содержательных задач оптимизации систем к задачам оптимизации сетей СМО, может приводить к формальным моделям, существенно различающимся по типу сети СМО и набору оптимизируемых переменных. Поэтому целесообразно провести классификацию задач оптимизации сетей СМО с тем, чтобы каждой классификационной группе подобных задач подобрать наилучший метод решения.

Система условных обозначений оптимационных моделей, дополненная рядом признаков, служит основой классификации задач оптимизации сетей СМО [22].

Произвольная задача оптимизации сети СМО формально может быть задана кортежем из четырех параметров:

$$\alpha/\beta/\chi/\delta,$$

где $\alpha \in \{SP1.1, SP2.1, SP3.1\}$,

$$\beta \in \{J, G\},$$

$$\chi \in \{S, M\},$$

$$\delta \in \{R, N\}$$

Параметр α обозначает тип задачи оптимизации, β — тип сети СМО: сеть Джексона (J) или немарковская сеть (G), χ — количество каналов в

узлах: одноканальная сеть (S) или многоканальная (M), δ — переменные задачи (оптимизируемые параметры сети): интенсивности обслуживания в узлах (R) или числа каналов в узлах (N).

5.1.3. Оптимизация сетей Джексона

В соответствии с введенной классификацией задача относится к типу $./J./..$. Точка в обозначении параметра означает, что он не определен, и может принимать любое из допустимых значений. В данном случае задается только тип сети СМО, используемой в задаче оптимизации. Однородные одноканальные сети Джексона приводят к алгебраическим выражениям для ряда показателей их эффективности, что открывает возможности получения точных аналитических решений задач SP.1.1 и SP.2.1.

5.1.3.1. Точные методы решения задач SP.1.1/J/S/R и SP.2.1/J/S/R

Начнем с задачи SP.2.1/J/S/R, которая впервые получила аналитическое решение в работах Леонарда Клейнрока [26, 27]. Объектом оптимизации является разомкнутая линейная экспоненциальная одноканальная сеть СМО, расчет показателей которой излагается в разделе 5.2. Формальная постановка задачи оптимизации предполагает выбор интенсивностей обслуживания в узлах сети, минимизирующий среднее число заявок в сети при ограничении на величину стоимости сети:

SP.2.1/J/S/R

$$\left. \begin{aligned} \min_{\{\mu_i\}} & \left(L(\{\mu_i\}, \beta) = \sum_{j=1}^M L_j(\mu_j) = \sum_{j=1}^M \left(\frac{\lambda_j}{\mu_j - \lambda_j} \right) \right) \\ & \sum_{j=1}^M v_j \mu_j = F_T \end{aligned} \right\}, \quad (5.1)$$

где

$\{\mu_i\} = (\mu_1, \mu_2, \dots, \mu_M)$ — вектор интенсивностей обслуживания в узлах

сети,

λ_j — интенсивность потока на входе в j -й узел,

$L_j(\mu_j)$ — среднее число заявок в j -м узле,

$v_j \mu_j$ — стоимость j -го узла (линейная функция интенсивности μ_j),

F_T — стоимость сети СМО.

Для решения задачи (5.1) используется метод неопределенных множителей Лагранжа [28]. Функция Лагранжа приобретает следующий вид:

$$L(\{\mu_i\}, \beta) = \sum_{j=1}^n \left(\frac{\lambda_j}{\mu_j - \lambda_j} \right) + \beta \left(\sum_{j=1}^M v_j \mu_j - F_T \right), \quad (5.2)$$

а решением задачи (5.1) становится особая точка функции Лагранжа, если

матрица вторых производных $H(L)_\mu = \left\{ \frac{\partial^2 L(\{\mu_i\}, \beta)}{\partial \mu_i \partial \mu_j} \right\}_{\substack{i=1, M \\ j=1, M}}$ этой функции

положительно определена [28].

Метод Лагранжа приводит к следующей системе уравнений.

$$\left. \begin{array}{l} \frac{\partial L}{\partial \mu_j} = -\lambda_j (\mu_j - \lambda_j)^{-2} + \beta v_j = 0, \quad j = \overline{1, M} \\ \sum_{j=1}^M v_j \mu_j = F_T \end{array} \right\}, \quad (5.3)$$

и к решению

$$\mu_j^* = \lambda_j + \frac{\sqrt{v_j \lambda_j}}{\sum_{i=1}^M \sqrt{v_i \lambda_i}} \cdot \frac{F_T - \sum_{i=1}^n v_i \lambda_i}{v_j}, \quad j = \overline{1, M}, \quad (5.4)$$

которое действительно является оптимальным, поскольку положительная

определенность матрицы $\mathbf{H}(\mathbf{L})_{\mu}$ следует из очевидного условия (5.5).

$$\frac{\partial^2 \mathbf{L}}{\partial \mu_j^2} = 2\lambda_j (\mu_j - \lambda_j)^{-3} > 0 . \quad (5.5)$$

Перейдем к двойственной задаче минимизации стоимости сети с заданным значением среднего числа заявок в сети

SP.1.1/J/S/R

$$\left. \begin{array}{l} \min_{\{\mu_i\}} \sum_{j=1}^M v_j \mu_j = F_T \\ \sum_{j=1}^M \left(\frac{\lambda_j}{\mu_j - \lambda_j} \right) = L_T \end{array} \right\} . \quad (5.6)$$

Использование метода Лагранжа позволяет найти оптимальное решение и в этом случае [5]:

$$\mu_j^* = \lambda_j + \frac{\sum_{i=1}^M \sqrt{v_i \lambda_i}}{L_T} \cdot \sqrt{\frac{\lambda_j}{v_j}}, \quad j = 1, M . \quad (5.7)$$

5.1.3.2. Алгоритмы поиска оптимальных решений

задач SP.1.1/J/M/N и SP.2.1/J/M/N

Оптимизация однородных многоканальных сетей Джексона с использованием выпуклого программирования уже не приводит к конструктивным результатам по причине существенно более сложных конструкций выражений для показателей эффективности. При этом остается формальная возможность построения соответствующей функции Лагранжа, например, для задач SP.1.1/J/M/N и SP.2.1/J/M/N. Можно далее построить систему уравнений, вытекающую (как и ранее в задачах SP.1.1 и SP.2.1) из условия равенства нулю градиента функции Лагранжа по переменным за-

дачи и множителю Лагранжа. Проблема состоит в том, что не удается найти аналитического решения этой системы уравнений и приходится обращаться к численным методам, основанным на использовании подходов, приводящих к эвристическим алгоритмам оптимизации [22].

Для удобства ниже повторно приводится формула для среднего числа заявок в узле разомкнутой многоканальной однородной сети из табл. 5.1, активно используемая в алгоритмах оптимизации рассматриваемых задач.

Приведенные ниже алгоритмы 1 и 2 являются интерпретацией приведенных в [22] алгоритмов (Algorithm 1, Algorithm 2).

$$L_j(m_j, \mu_j, \lambda_j) = \frac{\frac{\lambda_j}{m_j \mu_j} \left(\frac{\lambda_j}{\mu_j} \right)^{m_j} \cdot P_0}{m_j \cdot m_j! \left(1 - \frac{\lambda_j}{m_j \mu_j} \right)^2}, \quad (5.8)$$

где

$$P_0 = \left[1 + \sum_{i=1}^{m_j} \frac{\left(\frac{\lambda_j}{\mu_j} \right)^i}{i!} + \frac{\left(\frac{\lambda_j}{\mu_j} \right)^{m_j+1}}{m_j! \left(m_j - \frac{\lambda_j}{m_j \mu_j} \right)} \right]^{-1}, \text{ а}$$

m_j, μ_j, λ_j — число каналов, интенсивность обслуживания и интенсивность входного потока j -го узла типа M/M/m соответственно.

Определим предпосылки решения задач SP.1.1/J/M/N и SP.2.1/J/M/N.

1. Сеть СМО находится в установившемся режиме и найдены интенсивности потоков $\{\lambda_j\}, j = \overline{1, M}$ в узлах сети.

2. Для заданных интенсивностей $\{\mu_j\}, j = \overline{1, M}$ определены минимально возможные числа $m_j^0 = \left| \frac{\lambda_j}{\mu_j} \right| + 1, j = \overline{1, M}$ каналов в узлах, при условии

$\frac{\lambda_j}{\mu_j} < 1$ наличия установившегося режима в каждом узле.

3. Для каждого узла j -го узла формируются два показателя его реакции на единичное увеличение числа каналов: приращение стоимости $\Delta F_j(m_j + I)$ узла и уменьшение числа заявок $\Delta L_j(m_j + I)$ в узле.

$$\Delta F_j(m_j + I) = F_j(m_j + I) - F_j(m_j) \geq 0$$

$$\Delta L_j(m_j + I) = L_j(m_j + I) - L_j(m_j) < 0$$

Алгоритм 1 решения задачи SP.1.1/J/M/N

Для каждого j -го узла рассчитывается индекс приоритета $PI_j(m_j)$ на добавление канала по следующей формуле

$$PI_j(m_j) = \frac{\Delta F_j(m_j + I)}{-v_j \Delta L_j(m_j + I)} . \quad (5.9)$$

SP.1.1/J/M/N

$$\left. \begin{array}{l} \min_{\{m_j\}} F(\{m\}) = F(\{m_j\}) = \sum_{j=1}^M F_j(m_j) \\ L(\{m\}) = L(\{m_j\}) = \sum_{j=1}^M v_j L_j(m_j) \leq L_T \\ m_j \geq m_j^0, \quad j = \overline{1, M}; \quad m_j - \text{integer} \end{array} \right\} \quad (5.10)$$

Шаг 1. Найти $\{m_j^0\}, j = \overline{1, M}$, идти к шагу 2

Шаг 2. Если $L(\{m_j^0\}) \leq L_T$, вектор $\{m_j^0\}, j = \overline{1, M}$ есть оптимальное решение,

Идти к шагу 9, иначе идти к шагу 3.

Шаг 3. Найти $PI(\{m\}) = \{PI_j(m_j)\}, j = \overline{1, M}$, идти к шагу 4.

Шаг 4. Найти $PI_{j^*} = \min_{(j)} PI_j(m_j)$, идти к шагу 5.

Шаг 5. $m_{j^*} = m_{j^*} + 1$, идти к шагу 6.

Шаг 6. Пересчитать $L(j^*)$, $F_j(m_{j^*})$, $PI_{j^*}(m_{j^*})$, идти к шагу 7.

Шаг 7. Коррекция m , $L(m)$, $F(m)$, $PI(m)$, идти к шагу 8.

Шаг 8. Если $L(m) \leq L_T$, то вектор $m = \{m_j\}, j = \overline{1, M}$ есть оптимальное решение,

Идти к шагу 9, иначе идти к шагу 4.

Шаг 9. ОСТАНОВ.

Алгоритм 2 решения задачи SP.2.1/J/M/N

SP.2.1/J/M/N

$$\left. \begin{array}{l} \min_{\{m_j\}} L(\{m_j\}) = \sum_{j=1}^M v_j L_j(m_j) \\ \sum_{j=1}^M m_j = M \mathbf{1} \\ m_j \geq m_j^0, \quad j = \overline{1, M}; \quad m_j - \text{integer} \end{array} \right\} \quad (5.11)$$

Для каждого j -го узла рассчитывается индекс приоритета $PI_j(m_j)$ на добавление канала по следующей формуле

$$PI_j(m_j) = -v_j \Delta L_j(m_j + 1) \quad . \quad (5.12)$$

Шаг 1. Найти $\{m_j^0\}, j = \overline{1, M}$, идти к шагу 2

Шаг 2. Если $\sum_{j=1}^M m_j^0 = M \mathbf{1}$, вектор $\{m_j^0\}, j = \overline{1, M}$ есть оптимальное

решение,

Идти к шагу 7, иначе идти к шагу 3.

Шаг 3. Найти $\{PI_j(m_j)\}, j = \overline{1, M}$, идти к шагу 4.

Шаг 4. Найти $PI_{j^*} = \max_{(j)} PI_j(m_j)$, идти к шагу 5.

Шаг 5. $m_{j^*} = m_{j^*} + 1$, идти к шагу 6.

Шаг 6. Если $\sum_{j=1}^n m_j = M\mathbf{1}$, вектор $\{m_j\}, j = \overline{1, M}$ есть оптимальное

решение,

Идти к шагу 8, иначе идти к шагу 7.

Шаг 7. Пересчет $PI_j(m_j^*)$,

идти к шагу 3.

Шаг 8. ОСТАНОВ.

5.1.4. Оптимизация разомкнутых немарковских сетей СМО

5.1.4.1. Оптимизация разомкнутых однородных немарковских сетей СМО

Рассмотрим два алгоритма оптимизации разомкнутых однородных одноканальных немарковских сетей СМО, SP.1.1/G/S/R и SP.2.1/G/S/R, реализующие принцип декомпозиции сети и основанные на использовании приведенных в разделе 4.4.1. расчетных соотношений для определения среднего времен пребывания заявки в сети.

Предпосылки решения задач SP.1.1/G/S/R и SP.2.1/G/S/R следующие.

1. Сеть СМО находится в установившемся режиме и найдены интенсивности потоков $\{\lambda_j\}, j = \overline{1, M}$ в узлах сети.

2. С учетом найденных интенсивностей $\{\lambda_j\}, j = \overline{1, M}$ определены допустимые значения интенсивностей $\mu_j^0, j = \overline{1, M}$ обслуживания в узлах из условия стационарности $\mu_j^0 > \lambda_j$.

3. Среднее число заявок в j -м узле $L_j(\mu_j)$ определяется по правилу Литтла с учетом (4.79):

$$L_j(\mu_j) = \lambda_j E(W_j) + \frac{\lambda_j}{\mu_j} = \frac{\left(\frac{\lambda_j}{\mu_j}\right)^2 (ca_j + cs_j) g(\rho_j, ca_j, cs_j)}{2(1 - \rho_j)} + \frac{\lambda_j}{\mu_j} \quad (5.13)$$

4. Параметры $ca_j, j = \overline{1, M}$ определяются в результате решения системы уравнений (4.78).

Представленные ниже пошаговые процедуры (алгоритмы 3 и 4) являются интерпретацией приведенных в [22] алгоритмов (Algorithm 3, Algorithm 4).

Алгоритм 3 решения задачи SP.1.1/G/S/R

SP.1.1/G/S/R

$$\left. \begin{array}{l} \min_{\{\mu_j\}} F(\mu) = F(\{\mu_j\}) = \sum_{j=1}^M F_j(\mu_j) \\ L(\mu) = L(\{\mu_j\}) = \sum_{j=1}^M v_j L_j(\mu_j) \leq L_T \\ \mu_j \geq \mu_j^0, \quad j = \overline{1, M}. \end{array} \right\} \quad (5.14)$$

Для каждого j -го узла рассчитывается индекс приоритета $PI_j(\mu_j)$ на увеличение интенсивности обслуживания μ_j

$$PI_j(\mu_j) = \frac{\partial F_j(\mu_j)/\partial \mu_j}{-\nu_j \partial L_j(\mu_j)/\partial \mu_j}, \quad (5.15)$$

где функция $L_j(\mu_j)$ задается выражением (5.13), а ν_j — коэффициенты, связывающие число заявок в узле с величиной штрафа за их наличие.

Параметром алгоритма является заданное приращение Δ интенсивности обслуживания в узле.

Шаг 1. Найти допустимое решение $\{\mu_j^0\}, j = \overline{1, M}$,

Найти $\{ca_j\}, j = \overline{1, M}$, $L(\mu^0)$ идти к шагу 2.

Шаг 2. Если $L(\mu^0) \leq L_r$, вектор $\{\mu_j^0\}, j = \overline{1, M}$ есть оптимальное

решение, идти к шагу 9, иначе идти к шагу 3.

Шаг 3. Найти $\{PI_j(\mu_j)\}, j = \overline{1, M}$, идти к шагу 4.

Шаг 4. Найти $PI_{j^*} = \min_{(j)} PI_j(\mu_j)$, идти к шагу 5.

Шаг 5. $\mu_{j^*} = \mu_{j^*} + \Delta$, идти к шагу 6.

Шаг 6. Пересчитать $\rho_{j^*} = \frac{\lambda_{j^*}}{\mu_{j^*}}$ и ca_{j^*} , пересчитать также $\{ca_i\}$ всех узлов, связанных с выходом узла j^* : $\{ca_i, i \in I_{j^*}\}, I_{j^*} = \{i : q_{ij^*} \neq 0\}$,

Шаг 7. Пересчитать $L(j^*), F_j(\mu_{j^*}), PI_j(\mu_{j^*})$, идти к шагу 8.

Шаг 8. Коррекция $\mu, L(\mu), F(\mu), PI(\mu)$, идти к шагу 9.

Шаг 9. Если $L(\mu) \leq L_r$, то вектор $\mu = \{\mu_j\}, j = \overline{1, M}$ есть оптимальное решение,

Идти к шагу 10, иначе идти к шагу 3.

Шаг 10. ОСТАНОВ.

Алгоритм 4 решения задачи SP.2.1/G/S/R

SP.2.1/G/S/R

$$\left. \begin{array}{l} \min_{\{\mu_j\}} L(\mu) = L(\{\mu_j\}) \\ \sum_{j=1}^n \mu_j = M1 = Const \\ \mu_j \geq \mu_j^\theta, \quad j = \overline{1, M} \end{array} \right\} \quad (5.16)$$

Пояснения к алгоритму 4.

Алгоритм реализует перераспределение интенсивностей обслуживания в узлах, увеличивая их в одних узлах за счет такого же уменьшения в других узлах с учетом постоянно корректируемого множества $\{PI_j(\mu_j)\}$ индексов приоритета узлов сети СМО.

Для каждого j -го узла рассчитывается индекс приоритета $PI_j(\mu_j)$ на увеличение интенсивности обслуживания μ_j

$$PI_j(\mu_j) = -v_j \frac{\partial L_j(\mu_j)}{\partial \mu_j}. \quad (5.17)$$

Параметром алгоритма является приращение Δ интенсивности обслуживания в узле. Расчет Δ_1 с учетом $\{\epsilon_j\}, j \in J_\theta$, полученных в результате решения уравнения $PI_j(\mu_j, (\lambda_j + \epsilon_j)) = \max_{\{j\}} PI_j(\mu_j), j \in J_\theta$, гарантирует выполнение условия $\mu_j > \lambda_j, j = \overline{1, M}$ (см. шаг 3).

Шаг 1.

1. Найти допустимое решение $\{\mu_j^I\} \geq \{\mu_j^\theta\}, j = \overline{1, M}, \sum_{j=1}^M \mu_j^I = M1$.

2. Найти $\{ca_j\}, j = \overline{I, M}, L(\mu^1)$.
3. Сформировать множество $J_0 = \{1, 2, \dots, M\}$, включающее в себя все узлы сети СМО;
 сформировать пустое множество J_1 узлов, получающих добавление интенсивности обслуживания;
 сформировать пустое множество J_2 узлов, получающих уменьшение интенсивности обслуживания, идти к шагу 2.

Шаг 2. Найти $\{PI_j(\mu_j)\}, j \in J_0$, определить $\{\varepsilon_j\}, j \in J_0$

из условия $PI_j(\mu_j, (\lambda_j + \varepsilon_j)) = \max_{\{j\}} PI_j(\mu_j), j \in J_0$,

идти к шагу 3.

Шаг 3. Найти $j_2 = \arg \max_{(j)} PI_j(\mu_j)$;

найти $j_1 = \arg \min_{(j)} PI_j(\mu_j)$;

Найти $\Delta_1 = \min \{\Delta, \mu_{j_1} - \lambda_{j_1} - \varepsilon_{j_1}\}$ и

выполнить $\mu_{j_1} \leftarrow \mu_{j_1} - \Delta_1, \mu_{j_2} \leftarrow \mu_{j_2} + \Delta_1$,

если $j_1 \in J_1$, тогда выполнить $J_0 \leftarrow J_0 - \{j_1\}$;

если $j_2 \in J_2$, тогда выполнить $J_0 \leftarrow J_0 - \{j_2\}$;

если $j_1 \notin J_1$, и $j_2 \notin J_2$, тогда выполнить $J_1 \leftarrow J_1 \cup \{j_2\}$ и $J_2 \leftarrow J_2 \cup \{j_1\}$;

идти к шагу 4.

Шаг 4. Если J_0 есть пустое (или унитарное) множество или

$$PI_{j_1} = PI_{j_2},$$

идти к шагу 8, иначе идти к шагу 5.

Шаг 5. Пересчитать $\rho_{j_2} = \frac{\lambda_{j_2}}{\mu_{j_2}}$ и ca_{j_2} , $\rho_{j_1} = \frac{\lambda_{j_1}}{\mu_{j_1}}$ и ca_{j_1} , пересчитать

$\{ca_i\}$ всех узлов, связанных с выходами узлов j_1 и j_2 , идти к шагу 6.

Шаг 6. Пересчитать $L_{j_1}(\mu_{j_1})$, $L_{j_2}(\mu_{j_2})$, $PI_{j_1}(\mu_{j_1})$, $PI_{j_2}(\mu_{j_2})$, идти к шагу 7.

Шаг 7. Коррекция μ , $L(\mu)$, $\{PI_j(\mu_j)\}$, идти к шагу 3.

Шаг 8. ОСТАНОВ.

5.1.4.2. Оптимизация разомкнутых неоднородных немарковских сетей СМО

Объектами оптимизации рассмотренных выше задач являются исключительно однородные сети СМО. Для оптимизации неоднородных сетей СМО последние приводятся к однородным с использованием изложенного в разделе 4.4.3. подхода. Далее проводится оптимизация сконструированной однородной сети, а затем осуществляется обратный переход от однородной сети к неоднородной. Использование алгоритмов SP.1.1/G/S/R и SP.2.1/G/S/R для оптимизации неоднородной сети СМО поясняется приведенной ниже последовательностью шагов.

Шаг 1. Задание неоднородной сети СМО типа G/G/1:

$$\left\{ M, \left\{ \lambda_0^{(r)} \right\}_{r=1,\overline{R}}, \left\{ ca_0^{(r)} \right\}_{r=1,\overline{R}}, \left\{ Q^{(r)} \right\}_{r=1,\overline{R}}, \left\{ \mu_{ir} \right\}_{\substack{i=1,\overline{M} \\ r=1,\overline{R}}}, \left\{ cs_{ir} \right\}_{\substack{i=1,\overline{M} \\ r=1,\overline{R}}} \right\}.$$

Шаг 2. Приведение неоднородной сети к однородной типа G/G/1:

$$\left\{ M, \lambda_0, ca_0, Q, \left\{ \mu_i \right\}_{i=1,\overline{M}}, \left\{ cs_i \right\}_{i=1,\overline{M}} \right\}.$$

Шаг 3. Переход к оптимизированной однородной сети СМО типа

G/G/1 — оптимальный выбор интенсивностей обслуживания $\{\mu_i^*\}$ в узлах

сети: $\left\{M, \lambda_0, ca_0, Q, \{\mu_i^*\}_{i=1,M}, \{cs_i\}_{i=1,M}\right\}.$

Шаг 4. Возврат к неоднородной сети СМО с расчетом интенсивностей $\{\mu_{ir}^*\}$ и откорректированных параметров $\{cs_{ir}^*\}$:

$$\left\{M, \left\{\lambda_0^{(r)}\right\}_{r=1,R}, \left\{ca_0^{(r)}\right\}_{r=1,R}, \left\{Q^{(r)}\right\}_{r=1,R}, \left\{\mu_{ir}^*\right\}_{\substack{i=1,M \\ r=1,R}}, \left\{cs_{ir}^*\right\}_{\substack{i=1,M \\ r=1,R}}\right\}$$

Обратный переход от оптимизированной осредненной сети к исходной неоднородной сети СМО

Для завершения процедуры оптимизации неоднородной сети необходимо осуществить переход от множеств $\{\mu_i^*\}$ и $\{cs_i\}$ осредненной оптимизированной однородной сети к множествам $\{\mu_{ir}^*\}$ и $\{cs_{ir}^*\}$ неоднородной сети. Элементы последних двух множеств необходимо сформировать таким образом, чтобы при их осреднении в соответствии с формулами (4.84) и (4.85) было обеспечено получение найденных на шаге 3 элементов множеств $\{\mu_i^*\}$ и $\{cs_i\}$ однородной оптимизированной осредненной сети. Начнем с формирования элементов множества $\{\mu_{ir}^*\}$. Значения этим элементам необходимо назначать с использованием уже упомянутой

формулы (4.84):
$$\frac{1}{\mu_i^*} = \frac{\sum_{r=1}^R \left(\frac{1}{\mu_{ir}^*} \right) \lambda_i^{(r)}}{\lambda_i}, \quad i = \overline{1, M}.$$

Введем обозначение $\tau_{ir}^* = \frac{1}{\mu_{ir}^*}$, тогда для расчета τ_{ir}^* получаем систему

$$\text{му линейных уравнений} \quad \sum_{r=1}^R \tau_{ir}^* \lambda_i^{(r)} \mu_i^* = \lambda_i, \quad i = \overline{1, M}. \quad (5.18)$$

Очевидно, что при заданных значениях элементов множества $\{\mu_i^*\}$

этая система уравнений удовлетворяет бесконечному множеству решений $\{\mu_{ir}^*\}$. Чтобы найти единственное решение $\{\mu_{ir}^*\}$, систему (5.18) необходимо дополнить. Потребуем, чтобы соотношения между интенсивностями обслуживания заявок разных классов в каждом узле после оптимизации сети оставались неизменными, то есть выполнялось бы следующее равенство:

$$\frac{\mu_{i(r+1)}^*}{\mu_{ir}^*} = \frac{\mu_{i(r+1)}}{\mu_{ir}}, \quad r = \overline{1, R-1}. \quad (5.19)$$

Это условие означает, что все интенсивности обслуживания заявок разных классов в узле неоднородной оптимизированной сети изменяются синхронно в одинаковое число раз.

Преобразуем равенство (5.19) с учетом введенных ранее обозначений:

$$\frac{\tau_{ir}}{\tau_{i(r+1)}} = \frac{\mu_{i(r+1)}}{\mu_{ir}} \Leftrightarrow \tau_{ir} - \tau_{i(r+1)} \frac{\mu_{i(r+1)}}{\mu_{ir}} = 0. \quad \square$$

В конечном итоге формируется следующая система уравнений для назначения интенсивностей обслуживания заявок разных классов в i -м

$$\text{узле: } \begin{cases} \sum_{r=1}^R \tau_{ir} \lambda_i^{(r)} \mu_i^* = \lambda_i \\ \tau_{ir} - \tau_{i(r+1)} \frac{\mu_{i(r+1)}}{\mu_{ir}} = 0, \quad r = \overline{1, (R-1)}. \end{cases} \quad (5.20)$$

Для того чтобы предложить способ коррекции элементов множества $\{cs_{ir}^*\}$ при переходе к оптимизированным значениям интенсивностей $\{\mu_{ir}^*\}$, проведем дополнительный анализ формулы (4.85).

$$cs_i^* = \frac{\sum_{r=1}^R \lambda_i^{(r)} \left(\frac{1}{\mu_{ir}^*} \right)^2 (cs_{ir}^* + 1)}{\lambda_i \left(\frac{1}{\mu_i^*} \right)^2} - 1, \quad i = \overline{1, M}$$

Синхронное изменение интенсивностей обслуживания заявок разных классов в узле сети при обратном переходе к неоднородной сети приводит к тому, что отношение

$$\frac{\sum_{r=1}^R \lambda_i^{(r)} \left(\frac{1}{\mu_{ir}^*} \right)^2}{\lambda_i \left(\frac{1}{\mu_i^*} \right)^2} \text{ остается неизменным, то есть}$$

$$\frac{\sum_{r=1}^R \lambda_i^{(r)} \left(\frac{1}{\mu_{ir}^*} \right)^2}{\lambda_i \left(\frac{1}{\mu_i^*} \right)^2} = \frac{\sum_{r=1}^R \lambda_i^{(r)} \left(\frac{1}{\mu_{ir}} \right)^2}{\lambda_i \left(\frac{1}{\mu_i} \right)^2}, \quad \text{а значит, для обеспечения равенства}$$

$cs_i^* = cs_i, \quad i = \overline{1, M}$ следует и элементы множества $\{cs_{ir}^*\}$ также оставить неизменными, назначив $\{cs_{ir}^*\} = \{cs_{ir}\}$.

Пример 5.1.

Оптимизация неоднородной разомкнутой одноканальной немарковской сети СМО

Постановка задачи

$$\min_{\{\mu_{ir}\}} \sum_{i=1}^M \sum_{r=1}^R \mu_{ir}$$

$$L = \sum_{i=1}^M \sum_{r=1}^R L_{ir} \leq L_T$$

Таблица 5.4. Исходные данные

M	$\{\lambda_0^{(r)}\}$	$\{ca_0^{(r)}\}$	$\{Q^{(r)}\}_{r=1,3}$	$\{\mu_{ir}\}$	$\{cs_{ir}\}$	L_T
3	3 4 2	0,16 0,09 0,16		7 8 9 9 9 10 10 9 10	0,04 0,16 0,09 0,16 0,09 0,09 036 0,36 0,16	2

Сеть СМО состоит из трех одноканальных узлов. На входе — неоднородный поток заявок $\{\lambda_0^{(r)}\}, r = 1, 2, 3$ с заданными значениями $\{ca_0^{(r)}\}, r = 1, 2, 3$ квадратов коэффициентов вариации длительностей интервалов поступления заявок в сеть. Матрицы передач $\{Q^{(r)}\}_{r=1,2,3}$ заявок каждого класса задают их маршрутизацию.

$$Q^{(1)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}, Q^2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, Q^{(3)} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

В матрице $Q^{(3)}$ присутствуют нулевые строка и столбец потому, что заявки 3-го класса не появляются в 1-м узле, то есть не могут входить в 1-й узел и не могут, естественно, из него выходить.

Данные о законах обслуживания представлены в форме значений $\{\mu_{ir}\}_{i=1,2,3}, r=1,2,3$, $\{cs_{ir}\}_{i=1,2,3}, r=1,2,3$ интенсивностей и квадратов коэффициентов вариации длительностей обслуживания в узлах. Необходимо найти интенсивности $\{\mu_{ir}\}_{i=1,2,3}, r=1,2,3$ неоднородного обслуживания в узлах, обеспечивающих выполнения условия $L = \sum_{i=1}^M \sum_{r=1}^R L_{ir} \leq L_T$, ограничивающего число L заявок в сети величиной L_T .

Задача решается в 3 этапа.

Этап 1. Приведение неоднородной сети СМО к однородной

Этап 2. Оптимизация однородной сети СМО в соответствии с алгоритмом 3 (раздел 5.1.4.1).

Этап 3. Возврат к неоднородной сети — переход от неоднородной оптимизированной сети СМО к неоднородной сети.

Этап 1

В соответствии с методикой, изложенной в разделе 4.4.3, осуществляется следующая последовательность действий.

$$\lambda_j^{(r)} = \sum_{i=0}^M \lambda_i^{(r)} q_{ij}^{(r)}, \quad j = \overline{0, M}$$

$$\lambda_{ij}^{(r)} = \lambda_i^{(r)} q_{ij}^{(r)}, \quad i, j = \overline{0, M}$$

$$\lambda_{ij} = \sum_{r=1}^R \lambda_{ij}^{(r)}, \quad i, j = \overline{0, M}$$

В итоге для данной задачи получена матрица интенсивностей

$$\Lambda = \left\{ \lambda_{ij} \right\}_{\substack{i=\overline{0,M} \\ j=\overline{0,M}}} = \begin{pmatrix} 0 & 3 & 2 & 4 \\ 4 & 0 & 3 & 0 \\ 0 & 4 & 0 & 5 \\ 5 & 0 & 4 & 0 \end{pmatrix}, \quad \text{на основе которой получена матрица}$$

$$Q = \left\{ q_{ij} \right\}_{\substack{i=\overline{0,M} \\ j=\overline{0,M}}} = \begin{pmatrix} 0,0000 & 0,3333 & 0,2222 & 0,4445 \\ 0,5714 & 0,0000 & 0,4286 & 0,0000 \\ 0,0000 & 0,4444 & 0,0000 & 0,5556 \\ 0,5556 & 0,0000 & 0,4444 & 0,0000 \end{pmatrix} \quad \text{однородной сети СМО,}$$

где $q_{ij} = \frac{\lambda_{ij}}{\lambda_i}$. По формулам (4.84) – (4.87) найдены значения параметров

ca_{0j} : $ca_{01} = 0,5272$; $ca_{02} = 0,2187$; $ca_{03} = 0,6377$, интенсивностей $\{\mu_j\}$:

$\mu_1 = 7,5385$; $\mu_2 = 9,2045$; $\mu_3 = 9,5294$ и параметров $\{cs_j\}$:

$cs_1 = 0,1055$; $cs_2 = 0,1164$; $cs_3 = 0,3234$.

Отметим, что осредненная однородная сеть СМО находится в уставновившемся режиме, т. к. полученные результаты подтверждают выполнение условия $\lambda_j < \mu_j$ в каждом узле ($\lambda_1 = 7$; $\lambda_2 = 9$; $\lambda_3 = 9$).

Этап 2

Оптимизация однородной сети осуществляется в соответствии с алгоритмом 3 (раздел 5.1.4.1.). Параметры оптимизации: инкремент интенсивности $\Delta = 1$. Найденный выше вектор осредненных значений интенсивностей обслуживания, удовлетворяет условию $\lambda_j < \mu_j$ и принимается

$$\text{в качестве вектора } \mu^0 = \begin{pmatrix} 7,5385 \\ 9,2045 \\ 9,5294 \end{pmatrix}.$$

Оптимизация однородной сети потребовала 32 итерации. В таблицах 5.5. и 5.6 представлены показатели ССМО, получаемые на каждой

итерации.

Таблица 5.5. Изменение интенсивностей обслуживания заявок в процессе оптимизации сети

№	μ_1	μ_2	μ_3	$\text{№ узла, получающий инкремент интенсивности } \Delta$
1	7.5385	9.2045	9.5294	2
2	7.5385	10.2045	9.5294	3
3	7.5385	10.2045	10.5294	1
4	8.5385	10.2045	10.5294	2
5	8.5385	11.2045	10.5294	3
6	8.5385	11.2045	11.5294	1
7	9.5385	11.2045	11.5294	3
8	9.5385	11.2045	12.5294	2
9	9.5385	12.2045	12.5294	1
10	10.5385	12.2045	12.5294	3
11	10.5385	12.2045	13.5294	2
12	10.5385	13.2045	13.5294	1
13	11.5385	13.2045	13.5294	3
14	11.5385	13.2045	14.5294	2
15	11.5385	14.2045	14.5294	3
16	11.5385	14.2045	15.5294	1
17	12.5385	14.2045	15.5294	2
18	12.5385	15.2045	15.5294	3
19	12.5385	15.2045	16.5294	1
20	13.5385	15.2045	16.5294	2
21	13.5385	16.2045	16.5294	3
22	13.5385	16.2045	17.5294	1
23	14.5385	16.2045	17.5294	2
24	14.5385	17.2045	17.5294	3
25	14.5385	17.2045	18.5294	2
26	14.5385	18.2045	18.5294	1
27	15.5385	18.2045	18.5294	3
28	15.5385	18.2045	19.5294	2
29	15.5385	19.2045	19.5294	1
30	16.5385	19.2045	19.5294	3
31	16.5385	19.2045	20.5294	2
32	16.5385	20.2045	20.5294	-

Таблица 5.6. Изменение характеристик сети в ходе оптимизации

№	$\sum_{i=1}^3 \mu_i$	L	ρ_1	ρ_2	ρ_3
1	26.2724	33.6488	0.93	0.98	0.94
2	27.2724	20.2199	0.93	0.88	0.94
3	28.2724	13.6378	0.93	0.88	0.85
4	29.2724	9.8323	0.82	0.88	0.85
5	30.2724	8.5428	0.82	0.80	0.85
6	31.2724	7.1380	0.82	0.80	0.78
7	32.2724	6.2924	0.73	0.80	0.78
8	33.2724	5.6695	0.73	0.80	0.72
9	34.2724	5.1638	0.73	0.74	0.72
10	35.2724	4.7763	0.66	0.74	0.72
11	36.2724	4.4200	0.66	0.74	0.67
12	37.2724	4.1377	0.66	0.68	0.67
13	38.2724	3.9096	0.61	0.68	0.67
14	39.2724	3.6768	0.61	0.68	0.62
15	40.2724	3.4923	0.61	0.63	0.62
16	41.2724	3.3271	0.61	0.63	0.58
17	42.2724	3.1748	0.56	0.63	0.58
18	43.2724	3.0427	0.56	0.59	0.58
19	44.2724	2.9189	0.56	0.59	0.54
20	45.2724	2.8086	0.52	0.59	0.54
21	46.2724	2.7084	0.52	0.56	0.54
22	47.2724	2.6118	0.52	0.56	0.51
23	48.2724	2.5276	0.48	0.56	0.51
24	49.2724	2.4484	0.48	0.52	0.51
25	50.2724	2.3707	0.48	0.52	0.49
26	51.2724	2.3042	0.45	0.52	0.49
27	52.2724	2.2397	0.45	0.49	0.49
28	53.2724	2.1757	0.45	0.49	0.46
29	54.2724	2.1220	0.45	0.47	0.46
30	55.2724	2.0678	0.42	0.47	0.46
31	56.2724	2.0141	0.42	0.47	0.44
32	57.2724	1.9686	0.42	0.44	0.44

Результаты оптимизации однородной сети СМО:

1) Суммарная интенсивность обслуживания в узлах

$$\sum_{i=1}^3 \mu_i = 57,2724$$

2) Достигнутое значение $L = 1,9686$

3) Вектор оптимальных интенсивностей $\mu^* = \begin{pmatrix} 16,5385 \\ 20,2045 \\ 20,5294 \end{pmatrix}$.

Этап 3

Обратный переход от однородной оптимизированной сети к неоднородной сети СМО осуществлен в соответствии с разделом 5.1.4.2. Интенсивности $\{\mu_{ir}^*\}$ назначены с использованием формул (5.18) – (5.20).

$$\{\mu_{ir}^*\} = \begin{pmatrix} 15,3571 & 17,5510 & 19,7449 \\ 19,7556 & 19,7556 & 21,9506 \\ 21,5432 & 19,3889 & 21,5432 \end{pmatrix}.$$

Квадраты коэффициентов вариации длительностей обслуживания в узлах остаются неизменными.

5.2. Оптимизация замкнутых сетей СМО

Пусть объектом оптимизации является замкнутая линейная экспоненциальная одноканальная или многоканальная сеть СМО, расчет показателей которой излагается в разделе 4.2.2. Возможные постановки задач оптимизации для подобных сетей, а также подходы к их решению рассмотрены, например, в работе [18]. В частности, показано что решение некоторых типовых задач оптимизации может быть сведено к решению сис-

тем нелинейных уравнений. Если использовать введенную выше в разделе 5.1.2. классификацию оптимизационных задач, то речь идет об аналогах задач типа SP.1.1 и SP.2.1, но с некоторыми изменениями формулировок. Изменения связаны именно с замкнутостью сетевых моделей. Самое существенное изменение состоит в том, что вместо такой характеристики, как среднее число заявок в разомкнутой сети, в нашем случае следует ввести другую характеристику – производительность замкнутой сети. Соответственно, в задаче SP.2.1 следует отыскивать максимум производительности, а в задаче SP.1.1 производительность выступает в качестве ограничения. Проиллюстрируем упомянутый подход на примере аналога задачи SP.2.1.

Пусть стоимость сети S как функция вектора $\mu = \{\mu_1, \mu_2, \dots, \mu_M\}$ задается следующим образом:

$$S = \sum_{j=1}^M c_j \mu_j^{\beta_j} \quad (5.21)$$

Здесь c_j – стоимостные коэффициенты; β_j – некоторые неотрицательные коэффициенты нелинейности. Коэффициенты c_j можно для многоканальной сети СМО интерпретировать, как числа каналов в j -м узле. Данная формула предложена в [18] и представляется достаточно общей, логичной и адекватной реальным ситуациям.

Производительности узлов обслуживания (интенсивности выходящих из узлов заявок) в замкнутых экспоненциальных сетях пропорциональны друг другу, поэтому под производительностью сети λ можно понимать производительность одного из узлов сети (например, первого, т.е. λ_1). Задача оптимизации данной замкнутой однородной сети СМО состоит в максимизации производительности сети при стоимости, не превосходящей заданной.

Из результатов работы [20], а также из теорем, доказанных в [18] следует, что максимум производительности сети не может достигаться внутри области ограничения на стоимость, и, следовательно, должно выполняться равенство $S(\mu) = S^*$, где S^* – ограничение на стоимость сети СМО. Таким образом, в данном случае задача оптимизации замкнутой экспоненциальной сети массового обслуживания может быть сформулирована в следующей постановке – найти:

$$\max_{\mu} \lambda_1 = \omega_1 G_M(N-1)/G_M(N) \quad (5.22)$$

при ограничении

$$S = \sum_{j=1}^M c_j \mu_j^{\beta_j} = S^*, \quad \mu > \mathbf{0}. \quad (5.23)$$

В выражении (5.22) использованы обозначения, введенные в разделе 4.2. Соответственно, $G_M(N)$ есть нормирующая константа замкнутой сети СМО с M узлами и N заявками, а ω_i – относительная вероятность прохождения заявки через i -й узел сети.

Как показано в [18], функция $\lambda(\mu)$ – выпуклая, поэтому любой локальный максимум задачи является также и глобальным максимумом.

Решение задачи (5.22, 5.23) целесообразно искать методом неопределенных множителей Лагранжа. Составим функцию Лагранжа $Q = \lambda_1 + \gamma(S - S^*)$, где γ – множитель Лагранжа. Взяв все частные производные функции Лагранжа и приравняв их к нулю, получим

$$\frac{\partial \lambda_1}{\partial \mu_i} + \gamma \frac{\partial S}{\partial \mu_i} = \mathbf{0} \quad , \quad i = \overline{1, M} \quad . \quad (5.24)$$

Далее будем использовать важные расчетные соотношения, полученные в [18] в результате доказательства соответствующих теорем:

$$\frac{\partial \lambda_i(N)}{\partial \mu_i} = \frac{1}{\mu_i} \lambda_i(N) [\bar{n}_i(N) - \bar{n}_i(N-1)] \quad (5.25)$$

и

$$\frac{\partial \lambda_i(N)}{\partial \mu_j} = \frac{1}{\mu_j} \lambda_i(N) [\bar{n}_j(N) - \bar{n}_j(N-1)] , \quad (5.26)$$

где $\bar{n}_i(N)$ и $\bar{n}_i(N-1)$ – средние числа заявок (с учетом находящихся в очереди) в узле номер i сети СМО, содержащей N и $N-1$ заявок, соответственно. Для уточнения смысла величины $\bar{n}_i(N)$ укажем, что в замкнутой сети СМО с числом заявок N и числом узлов M справедливо соотношение $\sum_{i=1}^M \bar{n}_i(N) = N$.

Подставляя (5.25) и (5.26) в (5.24), а также используя (5.23), получим

$$\frac{1}{\mu_i} [\bar{n}_i(N) - \bar{n}_i(N-1)] = -\gamma c_i \beta_i \mu_i^{\beta_i-1}, \quad i = \overline{1, M} . \quad (5.27)$$

С целью исключения множителя γ разделим каждое i -е уравнение, начиная со второго, на первое. А учитывая еще и ограничение (5.23) получим в итоге совокупность из M уравнений:

$$\frac{c_i \beta_i \mu_i^{\beta_i}}{c_1 \beta_1 \mu_1^{\beta_1}} = \frac{\bar{n}_i(N) - \bar{n}_i(N-1)}{\bar{n}_1(N) - \bar{n}_1(N-1)} , \quad i = \overline{2, M} ; \quad (5.28)$$

$$\sum_{i=1}^M c_i \mu_i^{\beta_i} = S^* . \quad (5.29)$$

Для получения итоговой системы нелинейных уравнений относительно искомых значений μ_i ($i = \overline{1, M}$) подставим значение $\mu_i^{\beta_i}$, полученное из (5.28) в (5.29). Система уравнений примет вид

$$\mu_1^{\beta_1} = \frac{S^*}{c_1 \left[1 + \frac{\beta_1}{\bar{n}_1(N) - \bar{n}_1(N-1)} \sum_{i=2}^M \frac{\bar{n}_i(N) - \bar{n}_i(N-1)}{\beta_i} \right]} \quad (5.30)$$

$$\mu_i^{\beta_i} = \frac{c_1 \beta_1}{c_i \beta_i} \mu_1^{\beta_1} \frac{\bar{n}_i(N) - \bar{n}_i(N-1)}{\bar{n}_1(N) - \bar{n}_1(N-1)} , \quad i = \overline{2, M} . \quad (5.31)$$

Если решить систему уравнений (5.30), (5.31), то найденные значения μ_i ($i = \overline{1, M}$) обеспечат максимум целевой функции $\lambda(\mu)$ при выполнении ограничения $S(\mu) = S^*$.

Перейдем теперь к двойственной задаче минимизации стоимости сети в условиях заданного значения производительности, то есть к аналогу задачи оптимизации SP.1.1. Пусть по-прежнему стоимость сети S как функция вектора $\mu = \{\mu_1, \mu_2, \dots, \mu_M\}$ задается аддитивным выражением (5.21). Как и в предыдущем случае, можно показать, что минимум стоимости сети не может достигаться внутри области ограничения на производительность, и, следовательно, должно выполняться равенство $\lambda_I = \lambda^*$, где λ^* – ограничение на производительность сети СМО. Таким образом, в данном случае задача оптимизации состоит в нахождении:

$$\min_{\mu} S = \sum_{j=1}^M c_j \mu_j^{\beta_j} \quad (5.32)$$

при ограничении

$$\lambda_1 = \omega_1 G_M(N-1)/G_M(N) = \lambda^* . \quad (5.33)$$

Аналогично предыдущему случаю функция $S(\mu)$ обладает тем свойством, что любой локальный минимум задачи является также и глобальным минимумом. Как и ранее, решение будем искать методом неопределенных множителей Лагранжа. Составим функцию Лагранжа

$Q = S + \gamma(\lambda_1 - \lambda^*)$, где γ – множитель Лагранжа. Возьмем все частные производные функции Лагранжа и приравняем их к нулю

$$\frac{\partial S}{\partial \mu_i} + \gamma \frac{\partial \lambda_1}{\partial \mu_i} = 0 \quad , \quad i = \overline{1, M} . \quad (5.34)$$

Подставляя (5.25) и (5.26) в (5.34), а также используя (5.21), получим

$$c_i \beta_i \mu_i^{\beta_i-1} = -\gamma \frac{1}{\mu_i} \lambda_1(N) [\bar{n}_i(N) - \bar{n}_i(N-1)], \quad i = \overline{1, M} \quad (5.35)$$

С целью исключения множителя γ разделим каждое i -е уравнение, начиная со второго, на первое. Кроме того используем известную формулу $\lambda_i = \mu_i \bar{n}_i^{обсл}(N)$, где $\bar{n}_i^{обсл}(N)$ – среднее число занятых каналов (приборов обслуживания) в узле i . В итоге получим совокупность из M уравнений:

$$\mu_1 = \lambda^* / \bar{n}_1^{обсл}(N) \quad (5.36)$$

$$\mu_i^{\beta_i} = \frac{c_1 \beta_1}{c_i \beta_i} \mu_1^{\beta_1} \frac{\bar{n}_i(N) - \bar{n}_i(N-1)}{\bar{n}_1(N) - \bar{n}_1(N-1)} , \quad i = \overline{2, M} . \quad (5.37)$$

Если решить систему уравнений (5.36), (5.37), то найденные значения μ_i ($i = \overline{1, M}$) обеспечат минимум целевой функции $S(\mu)$ при выполнении ограничения $\lambda_1 = \lambda^*$.

Таким образом, и при первой постановке задачи оптимизации, и при второй удалось показать, что задача сводится к решению систем нелинейных уравнений: (5.30), (5.31) – в первом случае и (5.36), (5.37) – во втором. Однако получить решения данных систем уравнений аналитическим способом затруднительно, а потому на практике обычно приходится использовать численные методы. Остановимся теперь на вычислительных аспектах решения рассмотренных задач оптимизации.

В большинстве известных публикацийделено мало внимания изучению возможности применения тех или иных численных методов для реше-

ния систем нелинейных уравнений, которые получаются в результате применения метода неопределенных множителей Лагранжа к рассмотренным выше задачам оптимизации замкнутых сетей СМО. Поэтому при подготовке настоящего учебного пособия авторами совместно с несколькими студентами было осуществлено экспериментальное испытание различных численных методов.

В частности, были испытаны следующие итерационные методы: метод простых итераций, метод Ньютона, метод Бройдена, модифицированный метод Ньютона. Алгоритмы и характеристики перечисленных методов описаны в литературе по вычислительной математике, например, в [21]. Здесь приведем лишь самые общие выводы и рекомендации, полученные экспериментально.

Как известно, метод простых итераций наиболее прост для реализации, но применительно к рассмотренным задачам оптимизации оказалось, что в большинстве случаев он не обеспечивает сходимости итерационного процесса. Следовательно, метод простых итераций не может быть рекомендован к применению. Существенно лучшие результаты показали три оставшиеся из перечисленных методов. Все они могут быть рекомендованы для решения рассмотренных выше систем нелинейных уравнений. При этом следует отметить, что методы Ньютона и Бройдена наиболее надежны с точки зрения сходимости. Модифицированный метод Ньютона показал самую высокую скорость сходимости итерационного процесса, но он более критичен к выбору начальных приближений.

В порядке подведения итога опишем последовательность шагов (этапов) решения задачи оптимизации с использованием модели в виде замкнутой сети СМО.

Шаг 1. Для заданного реального объекта (системы) построение струк-

туры замкнутой сетевой модели массового обслуживания.

Шаг 2. Определение основных параметров, описывающих замкнутую сеть СМО: M — число узлов сети; N — постоянное число заявок в сети; $\pi = \{p_{ij}\}$, $i, j = \overline{1, M}$ — стохастическая матрица передач, $\mu = \{\mu_1, \dots, \mu_M\}$ — вектор, i -м компонентом которого является интенсивность обслуживания заявки в канале i -го узла сети, $m = \{m_1, \dots, m_M\}$ — вектор, i -м компонентом которого является число каналов в i -м узле сети.

Шаг 3. Решение системы линейных алгебраических уравнений вида

$$(4.22) \quad \omega_j = \sum_{i=1}^m \omega_i p_{ij}, \quad i = \overline{1, M}, \quad \text{с целью определения вектора } \omega = \{\omega_1, \dots, \omega_M\} \text{ относительных частот прохождения заявки через узел } i.$$

Шаг 4. Составление (или заимствование) расчетных соотношений для определения следующих важнейших показателей качества обслуживания в произвольном узле j замкнутой сети СМО при условии наличия в ней N заявок:

$\overline{n_j}(N)$ — среднее число заявок в узле;

$\overline{n_j^{обсл}}(N)$ $\overline{n_j^{обсл}}(N)$ — среднее число занятых приборов;

$\lambda_j(N)$ — интенсивность прохождения заявок через узел и т.п.

При необходимости составление (или заимствование) компьютерных программ (или подпрограмм) для расчета указанных характеристик.

Шаг 5. Постановка задачи оптимизации, то есть выбор целевой функции и ограничений, например, $\max \lambda(\mu)$ при ограничении на стоимость сети S^* или $\min S(\mu)$ при ограничении на производительность λ^* . Определение вида формулы для стоимости сети и входящих в нее коэффи-

циентов. Рекомендуемый вид формулы $S = \sum_{j=1}^M c_j \mu_j^{\beta_j}$.

Шаг 6. Аналитическое решение задачи оптимизации, например, методом неопределенных множителей Лагранжа. В последнем случае необходимо построить функцию Лагранжа и после приравнивания нулю всех частных производных функции Лагранжа, а также необходимых подстановок получить систему нелинейных уравнений относительно искомых параметров.

Шаг 7. Выбор численного метода для решения полученной на Шаге 6 системы нелинейных уравнений. Составление (или заимствование) программы реализации выбранного численного метода.

Шаг 8. Численное решение системы нелинейных уравнений с помощью составленной (заимствованной) программы. Анализ полученных результатов и планирование дальнейших действий по оптимизации сети СМО.

При реализации части из перечисленных шагов требуется составление или заимствование компьютерных программ. В настоящее время широкие возможности предоставляют существующие специализированные пакеты прикладных программ. Так, например, в нашем случае может быть с успехом использован пакет прикладных программ MATLAB для решения задач технических вычислений и одноимённый язык программирования, используемый в этом пакете.

Пример 5.2.

Пусть для некоторой исходной системы построена модель в виде замкнутой сети СМО с количеством узлов $M = 4$ и постоянным количеством заявок, циркулирующих по сети, $N = 5$. Все узлы сети считаются од-

ноканальными. При этом стохастическая матрица передач определена как

$$\pi = \{p_{ij}\} = \begin{pmatrix} 0 & 0.2 & 0.3 & 0.5 \\ 0.5 & 0 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0 & 0.5 \\ 0.6 & 0.2 & 0.2 & 0 \end{pmatrix}. \quad (5.38)$$

Прототипом такой сетевой модели может быть какая-либо телекоммуникационная или вычислительная система.

Для определения вектора $\omega = \{\omega_1, \dots, \omega_M\}$ относительных частот прохождения заявки через каждый i -й узел составим систему уравнений $\omega_j = \sum_{i=1}^m \omega_i p_{ij}$, $i = \overline{1, M}$, которая в данном случае примет вид

$$\begin{cases} \omega_1 = 0.0\omega_1 + 0.2\omega_2 + 0.3\omega_3 + 0.5\omega_4 \\ \omega_2 = 0.5\omega_1 + 0.0\omega_2 + 0.3\omega_3 + 0.2\omega_4 \\ \omega_3 = 0.4\omega_1 + 0.1\omega_2 + 0.0\omega_3 + 0.5\omega_4 \\ \omega_4 = 0.6\omega_1 + 0.2\omega_2 + 0.2\omega_3 + 0.0\omega_4 \end{cases}. \quad (5.39)$$

В системе уравнений (5.39) всего три независимых уравнений, поэтому без введения каких-либо дополнительных условий она может быть решена только с точностью до мультипликативной константы. Положим $\omega_1 = 1$, тогда для остальных неизвестных получим следующие значения: $\omega_2 = 0.44$, $\omega_3 = 0.61$, $\omega_4 = 0.89$.

Расчетные соотношений для определения интересующих нас показателей качества обслуживания в произвольном узле i одноканальной замкнутой сети СМО с числом заявок N , заимствованные из раздела 4.2.5., приведены ниже.

Маргинальное распределение числа заявок в одноканальном узле однородной замкнутой сети СМО:

$$P_i(n, N) = \left(\frac{\omega_i}{\mu_i} \right)^n \left[G_M(N-n) - \left(\frac{\omega_i}{\mu_i} \right) G_M(N-n-1) \right] / G_M(N) . \quad (5.40)$$

Интенсивность на выходе i -го узла однородной замкнутой сети СМО:

$$\lambda_i(N) = \omega_i G_M(N-1) / G_M(N) . \quad (5.41)$$

Среднее число заявок в одноканальном узле однородной замкнутой сети СМО:

$$\bar{n}_i(N) = \left[\sum_{n=1}^N \left(\frac{\omega_i}{\mu_i} \right)^n G_M(N-n) \right] / G_M(N) . \quad (5.42)$$

Перейдем теперь к постановке задачи оптимизации. Пусть необходимо найти максимум производительности сети при ограничении на стоимость, то есть $\max \lambda(\mu)$ при $S = S^*$. Кроме того, пусть в данном случае стоимость сети линейно зависит от интенсивности обслуживания в каждом узле сети, то есть используемые в формуле для расчета стоимости сети (5.21) коэффициенты β_i все равны единице. Так как рассматриваемая сеть одноканальная, то и все коэффициенты c_i тоже равны единице. Таким образом, при сделанных допущениях формула стоимости сети примет вид:

$$S = \sum_{i=1}^M \mu_i . \quad (5.43)$$

Преобразуем поставленную оптимизационную задачу к системе уравнений. Составим функцию Лагранжа $Q = \lambda_1 + \gamma(S - S^*)$, где γ – множитель Лагранжа. Приравняем нулю все частные производные функции Лагранжа:

$$\frac{\partial \lambda_1}{\partial \mu_i} + \gamma \frac{\partial S}{\partial \mu_i} = 0 , \quad i = \overline{1, M} . \quad (5.44)$$

С учетом (5.25) и (5.26) получим систему уравнений:

$$\frac{1}{\mu_i} \lambda_1 [\bar{n}_i(N) - \bar{n}_i(N-1)] + \gamma = 0 , \quad i = \overline{1, M} . \quad (5.45)$$

Для того чтобы избавиться от множителя Лагранжа, разделим уравнения с номерами $i = \overline{2, M}$ на первое. С учетом (5.43) получим итоговую систему уравнений:

$$\mu_i = \mu_1 \frac{\bar{n}_i(N) - \bar{n}_i(N-1)}{\bar{n}_1(N) - \bar{n}_1(N-1)} , \quad i = \overline{2, M} \quad (5.46)$$

$$S^* = \sum_{i=1}^M \mu_i . \quad (5.47)$$

Осуществим некоторые дальнейшие преобразования с целью упрощения вида выражений для искомых величин μ_i . Подставим (5.46) в (5.47), тогда

$$S^* = \mu_1 + \mu_1 \sum_{i=2}^M \frac{\bar{n}_i(N) - \bar{n}_i(N-1)}{\bar{n}_1(N) - \bar{n}_1(N-1)} . \quad (5.48)$$

Заметим, что, поскольку изучаемая система замкнута, $\sum_{i=1}^M \bar{n}_i(N) = N$.

Преобразуем (5.48) с учетом последнего соотношения:

$$\begin{aligned} S^* &= \mu_1 \left(1 + \frac{1}{\bar{n}_1(N) - \bar{n}_1(N-1)} \left[\sum_{i=2}^M \bar{n}_i(N) - \sum_{i=2}^M \bar{n}_i(N-1) \right] \right) = \\ &= \mu_1 \left(1 + \frac{1}{\bar{n}_1(N) - \bar{n}_1(N-1)} [N - \bar{n}_1(N) - (N-1) + \bar{n}_1(N-1)] \right) = \\ &= \mu_1 \left(1 + \frac{1 - \bar{n}_1(N) + \bar{n}_1(N-1)}{\bar{n}_1(N) - \bar{n}_1(N-1)} \right) = \frac{\mu_1}{\bar{n}_1(N) - \bar{n}_1(N-1)} . \end{aligned}$$

Тогда:

$$\mu_1 = S^* [\bar{n}_1(N) - \bar{n}_1(N-1)] . \quad (5.49)$$

Подставим (5.49) в (5.46) и получим окончательную систему нелинейных уравнений:

$$\mu_i = S^* \left[\bar{n}_i(N) - \bar{n}_i(N-1) \right], \quad i = \overline{1, M}. \quad (5.50)$$

Для численного решения системы уравнений (5.41) может быть применен один из рекомендованных выше численных методов. Для $S^* = 12$ ниже приведен результат решения данной системы методом Ньютона:

$$\mu_1 = 3.784773; \quad \mu_2 = 2.095212; \quad \mu_3 = 2.644322; \quad \mu_4 = 3.475693.$$

При этом вычисления проводились с погрешностью до единицы шестого после десятичной точки знака, а в качестве начального приближения были выбраны равные друг другу значения интенсивностей обслуживания в узлах сети: $\mu_i = 3.0$, $i = \overline{1, M}$.

Соответственно, значения составляющих целевой функции, определяемые как $\lambda_i(N) = \omega_i G_M(N-1)/G_M(N)$, $i = \overline{1, M}$, получились следующие:

$$\lambda_1 = 2.585434; \quad \lambda_2 = 1.137591; \quad \lambda_3 = 1.577115; \quad \lambda_4 = 2.301036.$$

Задачи

Задача 1. Оптимизация многоканальной сети Джексона в соответствии с алгоритмом 1 (задача SP.1.1/J/M/N, п. 5.1.3.2):

$$\min \sum_{j=1}^M m_j;$$

$$L(\mathbf{m}) = \sum_{j=1}^M L_j \leq L_T.$$

Дано: $\left\{ M - \text{число узлов}; \quad Q = \left\{ q_{ij} \right\}_{\substack{i=\overline{0, M} \\ j=\overline{0, M}}}; \quad \vec{\mu} = (\mu_1, \mu_2, \dots, \mu_M) \right\}$.

В предположении установившегося режима в сети СМО самостоя-

тельно подобрать λ_0 , чтобы $\min_{(j)} \left\{ \overline{n_{обсл_j}} = \frac{\lambda_j}{\mu_j} \right\} \geq 2$.

Рекомендации по выбору λ_0

Сначала необходимо найти частоты $\{\alpha_i\}$, потом найти $\min_{(i)} \left\{ \frac{\alpha_i}{\mu_i} \right\} = \frac{\alpha_s}{\mu_s}$ и далее выбрать $\lambda_0 \leq \frac{2\mu_s}{\alpha_s}$. Граничное значение L_T числа заявок в сети также следует подобрать самостоятельно.

Рекомендации по выбору значения L_T

Очевидно, что минимальное значение числа заявок в сети (при сколь угодно большом, даже бесконечном, числе каналов в узле) равно

$\sum_{j=1}^M \frac{\lambda_j}{\mu_j}$. Таким образом, $L_{T \min} = \sum_{j=1}^M \frac{\lambda_j}{\mu_j}$. Верхняя граница $L_{T \max}$ может

быть рассчитана при начальном распределении числа каналов по узлам в соответствии с очевидным правилом: $m_j = \left[\frac{\lambda_0 \alpha_j}{\mu_j} \right] + 1$, $j = \overline{1, M}$, обеспечивающим установившийся режим в сети СМО, где $[.]$ — целая часть числа. Остается лишь самостоятельно выбрать L_T в диапазоне $(L_{T \min}, L_{T \max})$.

Варианты формулировок задачи 1 сведены в табл. 5.8.

Таблица 5.8

Номер варианта	$Q = \{q_{ij}\}_{\substack{i=0, M \\ j=0, M}}$					$\rightarrow \mu$			
1	0	0,2	0,3	0,5	0	1	2	3	2
	0,1	0	0,4	0,3	0,2				
	0,1	0,5	0	0,3	0,1				
	0,2	0,2	0,1	0	0,5				
	0,3	0	0,6	0,1	0				
2	0	0,1	0,7	0,1	0,1	2	4	2	1
	0,1	0	0,5	0,2	0,2				
	0,2	0,1	0	0,3	0,4				
	0,6	0,2	0,1	0	0,1				
	0,3	0,1	0,2	0,4	0				
3	0	0,3	0,2	0,1	0,4	3	2	1	1
	0,4	0	0,1	0,2	0,3				
	0,5	0	0	0,3	0,2				
	0,4	0,3	0	0	0,3				
	0,8	0	0,1	0,1	0				
4	0	0,3	0,3	0,2	0,2	5	4	3	2
	0,6	0	0,2	0,1	0,1				
	0,2	0,4	0	0,2	0,2				
	0,1	0,2	0,3	0	0,4				
	0,4	0,3	0,1	0,2	0				

Задача 2. Оптимизация многоканальной сети Джексона в соответствии с алгоритмом 2 (задача SP.2.1/J/M/N, п. 5.1.3.2):

$$\min \left(L(\mathbf{m}) = \sum_{j=1}^M L_j \right);$$

$$\sum_{j=1}^M m_j = M\mathbf{1}.$$

Дано: $\left\{ \begin{array}{l} M - \text{число узлов}; \quad Q = \left\{ q_{ij} \right\}_{\substack{i=0, M \\ j=0, M}}; \quad \vec{\mu} = (\mu_1, \mu_2, \dots, \mu_M) \end{array} \right\}.$

Для установившегося режима в сети СМО самостоятельно подобрать λ_0 , чтобы $\min_{(j)} \left\{ \overline{n_{обсл}}_j = \frac{\lambda_j}{\mu_j} \right\} \geq 2$.

Рекомендации по выбору λ_0 те же, что и для задачи 1.

Варианты формулировок задачи 2 представлены в табл. 5.9.

Т а б л и ц а 5.9

Номер варианта	$Q = \left\{ q_{ij} \right\}_{\substack{i=0, M \\ j=0, M}}$					$\vec{\mu}$				
1	0	0,1	0,5	0,1	0,3	<table border="1"> <tr><td>5</td><td>1</td><td>3</td><td>1</td></tr> </table>	5	1	3	1
5	1	3	1							
0	0	0,2	0,7	0,1						
0,6	0,2	0	0,1	0,1						
0,3	0,4	0,2	0	0,1						
0,1	0,1	0	0,8	0						
2	0	0	0,2	0,5	0,3	<table border="1"> <tr><td>2</td><td>3</td><td>2</td><td>1</td></tr> </table>	2	3	2	1
2	3	2	1							
0,2	0	0,3	0,1	0,4						
0	0,3	0	0,5	0,2						
0,7	0,1	0,1	0	0,1						
0,5	0	0,3	0,2	0						

Окончание табл. 5.9

Номер варианта	$Q = \{q_{ij}\}_{\substack{i=0, M \\ j=0, M}}$								
3	0	0,4	0,2	0,4	0	1	3	1	5
	0,3	0	0,5	0,1	0,1				
	0,2	0,4	0	0,2	0,2				
	0,1	0,1	0,1	0	0,7				
	0,2	0	0,2	0,6	0				
4	0	0,1	0,3	0,3	0,3	2	3	2	1
	0,2	0	0,2	0,2	0,4				
	0	0,5	0	0,4	0,1				
	0	0,6	0,2	0	0,2				
	0,1	0,2	0,3	0,4	0				
5	0	0,3	0,2	0,2	0,3	2	1	3	5
	0,3	0	0,4	0	0,3				
	0,2	0,1	0	0,7	0				
	0,4	0,3	0,2	0	0,1				
	0	0,3	0,2	0,5	0				

Задача 3. Оптимизация однородной одноканальной немарковской сети СМО в соответствии с алгоритмом 3 (задача SP.1.1/G/S/R, подразд. 5.3).

$$\min \sum_{j=1}^M \mu_j;$$

$$L(\mu) = \sum_{j=1}^M L_j \leq L_T.$$

Дано: $\left\{ \lambda_0; ca_0; M - \text{число узлов}; Q = \left\{ q_{ij} \right\}_{\substack{i=\overline{0,M} \\ j=\overline{0,M}}}; \left\{ cs_j \right\}, j = \overline{1, M}, L_T \right\}$,

где ca_0 — квадрат коэффициента вариации входного потока;

$\left\{ cs_j \right\}, j = \overline{1, M}$, — вектор квадратов коэффициентов вариации законов об- служивания в узлах сети.

Варианты условий задачи 3 приведены в табл. 5.10.

Т а б л и ц а 5. 10

Номер варианта	$Q = \left\{ q_{ij} \right\}_{\substack{i=\overline{0,M} \\ j=\overline{0,M}}}$					ca_0	λ_0	L_T	$\left\{ cs_j \right\}$			
1	0	0,6	0,1	0,1	0,2	0,04	10	5	0,64	0,25	0,25	0,25
	0,2	0	0,1	0,1	0,6							
	0,1	0,3	0	0,3	0,3							
	0,1	0,5	0,2	0	0,2							
	0,8	0	0,1	0,1	0							
2	0	0,3	0,2	0,1	0,4	0,09	12	6	0,36	0,36	0,16	0,09
	0,4	0	0,3	0,2	0,1							
	0	0,4	0	0,4	0,2							
	0,1	0,7	0,1	0	0							
	0,2	0,6	0	0,2	0							
3	0	0,6	0,1	0,1	0,2	0,16	8	5	0,04	0,09	0,16	0,16
	0,2	0	0,1	0,1	0,6							
	0,1	0,3	0	0,3	0,3							
	0,1	0,5	0,2	0	0,2							
	0,8	0	0,1	0,1	0							

Окончание табл. 5.10

Номер варианта	$Q = \{q_{ij}\}_{\substack{i=0, M \\ j=0, M}}$					ca_0	λ_0	L_T	$\{cs_j\}$			
4	0	0,6	0,1	0,1	0,2	0,25	6	6	0,09	0,09	0,09	0,09
	0,2	0	0,1	0,1	0,6							
	0,1	0,3	0	0,3	0,3							
	0,1	0,5	0,2	0	0,2							
	0,8	0	0,1	0,1	0							
5	0	0,6	0,1	0,1	0,2	0,36	11	7	2	2	2	2
	0,2	0	0,1	0,1	0,6							
	0,1	0,3	0	0,3	0,3							
	0,1	0,5	0,2	0	0,2							
	0,8	0	0,1	0,1	0							

Задача 4. Оптимизация однородной одноканальной немарковской сети СМО в соответствии с алгоритмом 4 (задача SP.2.1/G/S/R, подразд. 5.3):

$$\min L(\mu) = \sum_{j=1}^M L_j;$$

$$\sum_{j=1}^M \mu_j = \mu.$$

Дано: $\left\{ \lambda_0; ca_0; M - \text{число узлов}; Q = \{q_{ij}\}_{\substack{i=0, M \\ j=0, M}}; \{cs_j\}; j = \overline{1, M}; \mu \right\}$.

Варианты формулировок задачи 4 представлены в табл. 5.11.

Т а б л и ц а 5. 11

Номер варианта	$Q = \{q_{ij}\}_{\substack{i=0, M \\ j=0, M}}$					ca_0	λ_0	μ	$\{cs_j\}$			
1	0	0,1	0,5	0	0,4	0,49	23	4	1,96	1,96	1,96	1,96
	0,3	0	0,2	0,4	0,1							
	0,4	0,1	0	0,3	0,2							
	0,5	0	0,1	0	0,4							
	0	0,2	0,2	0,6	0							

Окончание табл. 5.11

Номер варианта	$Q = \{q_{ij}\}_{\substack{i=0, M \\ j=0, M}}$					ca_0	λ_0	L_T	$\{cs_j\}$			
2	0	0,1	0,3	0,3	0,3	0,64	12	8	0,36	0,36	0,36	0,36
	0,2	0	0,3	0,3	0,2							
	0,2	0,2	0	0,5	0,1							
	0	0,3	0,5	0	0,2							
	0,1	0,2	0,3	0,4	0							
3	0	0,4	0,3	0,3	0	0,09	13	6	0,16	0,16	0,09	0,09
	0	0	0,3	0,6	0,1							
	0,5	0,2	0	0,1	0,2							
	0,1	0,5	0,2	0	0,2							
	0	0,2	0,2	0,6	0							
4	0	0,2	0,3	0,2	0,3	0,16	8	9	0,04	0,04	0,04	0,04
	0,1	0	0,2	0,6	0,1							
	0,6	0,2	0	0,1	0,1							
	0	0,5	0,1	0	0,4							
	0,5	0,3	0,1	0,1	0							

Задача 5. Оптимизация неоднородной одноканальной немарковской сети СМО с приведением к однородной сети и с использованием алгоритма 3 (задача SP.1.1/G/S/R, подразд. 5.3):

$$\min \sum_{i=1}^M \mu_i;$$

$$L(\mu) = \sum_{i=1}^M L_i \leq L_r .$$

Дано:

$$\left\{ M; \left\{ \lambda_k' \right\}; \left\{ ca_k' \right\}_{k=\overline{1, r}}; \left\{ n_k \right\}_{k=\overline{1, r}}; \vec{n}_k = (n_{k1}, n_{k2}, \dots, n_{kn_k}); \left\{ \mu_{kl} \right\}; \left\{ cs_{kl} \right\}; \right.$$

где M — число узлов;

$\left\{ \lambda_k' \right\}$ — вектор интенсивностей входных потоков заявок разных классов, $k = \overline{1, r}$;

r — число классов заявок, равное числу детерминированных маршрутов;

n_k — число последовательно посещаемых узлов для k -го класса заявок (k -го маршрута);

$\vec{n}_k = (n_{k1}, n_{k2}, \dots, n_{kn_k})$ — цепочка номеров узлов, последовательно посещаемых на k -м маршруте;

μ_{kl} — интенсивность обслуживания заявок в l -м по счету узле на k -м маршруте; сочетание индексов kl соответствует узлу с номером n_{kl} , т. е. l -я работа на k -м маршруте выполняется узлом n_{kl} ;

$\{\mu_{kl}\}$ — интенсивность выполнения l -й работы на k -м маршруте (для всех пар $\{kl\}$);

$\{cs_{kl}\}$ — квадрат коэффициента вариации длительности l -й работы на k -м маршруте (для всех пар $\{kl\}$);

L_r — предельное число заявок в сети.

Варианты формулировок задачи 5 представлены в табл. 5.12.

Т а б л и ц а 5. 12

Номер варианта	M	$\{\lambda'_k\}$	$\{ca'_k\}$	$\{n_k\}$	\vec{n}_k			$\{cs_{kl}\}$			L_r
1	4	1	0,04	3	1	2	3	0,04	0,16	0,36	2
		2	0,09	3	4	3	2	0,09	0,36	0,09	
	4	3	1,44	3	1	3	4	1,44	1,96	4,0	3
		4	4,0	3	2	4	3	0,36	0,36	0,49	
3	4	1	0,25	3	1	3	2	0,09	0,16	0,04	2
		3	0,09	3	2	4	1	0,36	0,49	0,25	

Окончание табл. 5.12

Номер варианта	M	$\{\lambda_k'\}$	$\{ca_k'\}$	$\{n_k\}$	\vec{n}_k			$\{cs_{kl}\}$			L_r
4	4	4	0,16	3	4	2	1	0,25	0,25	0,36	3
		5	0,36	3	3	2	3	0,09	0,04	0,16	
	4	3	0,49	3	3	1	2	0,09	0,36	0,49	3
5	4	1	0,36	3	2	3	4	0,64	0,25	0,25	
		3	0,25	3	1	3	2	0,09	0,16	0,04	6
	4	3	0,09	3	2	4	1	0,36	0,49	0,25	

Рекомендации по решению задачи 5

Оптимизация неоднородной сети СМО в соответствии с алгоритмом 3 предполагает ее предварительное приведение к однородной сети. Тем не менее, на выходе задачи требуется получить интенсивности обслуживания заявок соответствующих классов в узлах, т. е. множество $\{\mu_{ik}^*\}, i = \overline{1, M}$.

Анализ метода приведения неоднородной сети СМО к однородной, а именно содержания формулы (4.84), позволяет сделать вывод о целесообразности распределения оптимальной интенсивности μ_i^* i -го узла приведенной однородной сети СМО пропорционально интенсивностям потоков разных классов на входе в узел:

$$\mu_{ik}^* = \mu_i^* \frac{\lambda^{(k)}}{\lambda_i}, \quad i = \overline{1, M}, \quad (5.47)$$

где μ_{ik}^* — оптимальная интенсивность обслуживания заявки r -го класса в i -м узле.

Эту процедуру нужно провести однократно при одинаковых коэф-

фициентах вариации длительностей обслуживания заявок разных классов в узле. Если это условие не выполняется, процедура проводится много-кратно с использованием формул (5.49) и (5.50) для коррекции $E(s_{kl})$ и cs_{kl} . Для формирования обобщенной матрицы передач приведенной однородной сети СМО следует использовать формулу (5.48), а остальные параметры приведенной однородной сети СМО вычислять по формулам (5.49) и (5.50):

$$q_{ij} = \frac{\sum_{k=1}^r \sum_{l=1}^{n_k-1} \lambda_k' \mathbf{1}\{(k, l) : n_{kl} = i, n_{k, l+1} = j\}}{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda_k' \mathbf{1}\{(k, l) : n_{kl} = i\}}; \quad (5.48)$$

$$E(s_j) = \frac{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda_k' E(s_{kl}) \mathbf{1}\{(k, l) : n_{kl} = j\}}{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda_k' \mathbf{1}\{(k, l) : n_{kl} = j\}}; \quad (5.49)$$

$$cs_j = \frac{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda_k' E(s_{kl})^2 (cs_{kl} + 1) \mathbf{1}\{(k, l) : n_{kl} = j\}}{\sum_{k=1}^r \sum_{l=1}^{n_k} \lambda_k' \mathbf{1}\{(k, l) : n_{kl} = j\} E(s_j)^2} - 1, \quad (5.50)$$

где $E(s_{kl}) = \frac{1}{\mu_{kl}}$, а $\mathbf{1}\{(k, l) : n_{kl} = j\}$ и $\mathbf{1}\{(k, l) : n_{kl} = i, n_{k, l+1} = j\}$ — индикаторные функции аргументов k, l , принимающие значение 1 при выполнении соответствующих условий: $n_{kl} = j$ и $(n_{kl} = i, n_{k, l+1} = j)$ соответственно, и значение 0 во всех остальных случаях.

Последовательность шагов при решении задачи 5

1. Найти матрицу Q по формуле (5.48).
2. Найти $\{\lambda_i^k\}$ — интенсивности потоков разных классов на входе

i -го узла ($i = \overline{1, M}$).

3. Просуммировать потоки на входах в узлы $\sum_{k=1}^r \lambda_i^k$ и назначить начальные значения μ_i^0 интенсивностей обслуживания приведенной однородной сети в узлах ($i = \overline{1, M}$) по условию $\mu_i^0 > \sum_{k=1}^r \lambda_i^k$.

4. Найти оптимальные значения $\{\mu_i^*\}, i = \overline{1, M}$ интенсивностей обслуживания заявок в узлах приведенной однородной сети.

5. Осуществить возврат к неоднородной сети СМО, распределив интенсивности обслуживания в узлах по классам в соответствии с рекомендациями разд. 5.1.4.2.

Задача 6. Оптимизация неоднородной одноканальной немарковской сети СМО с приведением к однородной сети и с использованием алгоритма 4 (задача SP.2.1/G/S/R, подразд. 5.3):

$$\min L(\mu) = \sum_{j=1}^n L_j;$$

$$\sum_{j=1}^n \mu_j = \mu.$$

Дано:

$$\left\{ M; \{\lambda_k'\}; \{ca_k'\}_{k=\overline{1,r}}; \{n_k\}_{k=\overline{1,r}}; \vec{n}_k = (n_{k1}, n_{k2}, \dots, n_{kn_k}); \{\mu_{kl}\}; \{cs_{kl}\}; \mu \right\},$$

где M — число узлов;

μ — предельная суммарная интенсивность обслуживания во всех узлах сети СМО;

$\{\lambda_k'\}$ — вектор интенсивностей входных потоков заявок разных

классов ($k = \overline{1, r}$);

r — число классов заявок, равное числу детерминированных маршрутов;

n_k — число последовательно посещаемых узлов (работ-operations) для k -го класса заявок (k -го маршрута);

$\vec{n}_k = (n_{k1}, n_{k2}, \dots, n_{kn_k})$ — цепочка номеров узлов, последовательно посещаемых на k -м маршруте;

μ_{kl} — интенсивность обслуживания заявок в l -м по счету узле на k -м маршруте; сочетание индексов kl соответствует узлу с номером n_{kl} , т. е. l -я работа на k -м маршруте выполняется узлом n_{kl} ;

$\{\mu_{kl}\}$ — интенсивность выполнения l -й работы на k -м маршруте (для всех пар $\{kl\}$);

$\{cs_{kl}\}$ — квадрат коэффициента вариации длительности l -й работы на k -м маршруте (для всех пар $\{kl\}$).

Варианты формулировок задачи 6 представлены в табл. 5.13.

Т а б л и ц а 5. 13

Номер варианта	M	$\{\lambda_k^*\}$	$\{ca_k^*\}$	$\{n_k\}$	\vec{n}_k	$\{cs_{kl}\}$	μ
1	4	3	1,44	3	1 3 4	1,44 1,96 4,0	10
		4	4,0	3	2 4 3	0,36 0,36 0,49	
2	4	1	0,25	3	1 3 2	0,09 0,16 0,04	8
		3	0,09	3	2 4 1	0,36 0,49 0,25	
3	4	4	0,16	3	4 2 1	0,25 0,25 0,36	15
		5	0,36	3	3 2 3	0,09 0,04 0,16	

Окончание табл. 5.13

Номер варианта	M	$\{\lambda_k^*\}$	$\{ca_k^*\}$	$\{n_k\}$	\vec{n}_k	$\{cs_{kl}\}$	μ
4	4	3	0,49	3	3 1 2	0,09 0,36 0,49	9
		1	0,36	3	2 3 4	0,64 0,25 0,25	
5	4	3	1,44	3	1 3 4	1,44 1,96 4,0	7
		4	4,0	3	2 4 3	0,36 0,36 0,49	
6	4	1	0,25	3	1 3 2	0,09 0,16 0,04	6
		3	0,09	3	2 4 1	0,36 0,49 0,25	

Рекомендации по решению задачи б

См. рекомендации к решению задачи 5.

Последовательность шагов при решении задачи б

1. Найти матрицу Q по формуле (5.48).
2. Найти $\{\lambda_i^k\}$ — интенсивности потоков заявок разных классов на входе i -го узла ($i = \overline{1, M}$).
3. Просуммировать потоки на выходах в узлы $\sum_{k=1}^r \lambda_i^k$ и назначить начальные значения μ_i^0 интенсивностей обслуживания приведенной однородной сети в узлах ($i = \overline{1, M}$), удовлетворяющие условиям:

$$\mu_i^0 > \sum_{k=1}^r \lambda_i^k; \quad \sum_{i=1}^M \mu_i^0 = \mu.$$

4. Найти оптимальные значения $\{\mu_i^*\}$, $i = \overline{1, M}$, интенсивностей обслуживания заявок в узлах приведенной однородной сети.

5. Осуществить возврат к неоднородной сети СМО, распределив интенсивности обслуживания в узлах по классам в соответствии с рекомен-

дациями разд. 5.1.4.2.

ПРИЛОЖЕНИЕ

ПОСТРОЕНИЕ МАТРИЦ A_i, F_i, B_i, D_i ДЛЯ МАРКОВСКОГО ГРАФА СОСТОЯНИЙ

Введем векторные состояния $P_i = \{P_{i0}, P_{i1}, \dots, P_{in}\}, i = \overline{0, m}$, объединяющие скалярные состояния $(i0, i1, \dots, in)$ регулярного марковского графа типа «решетка», представленного на рис. П.1.

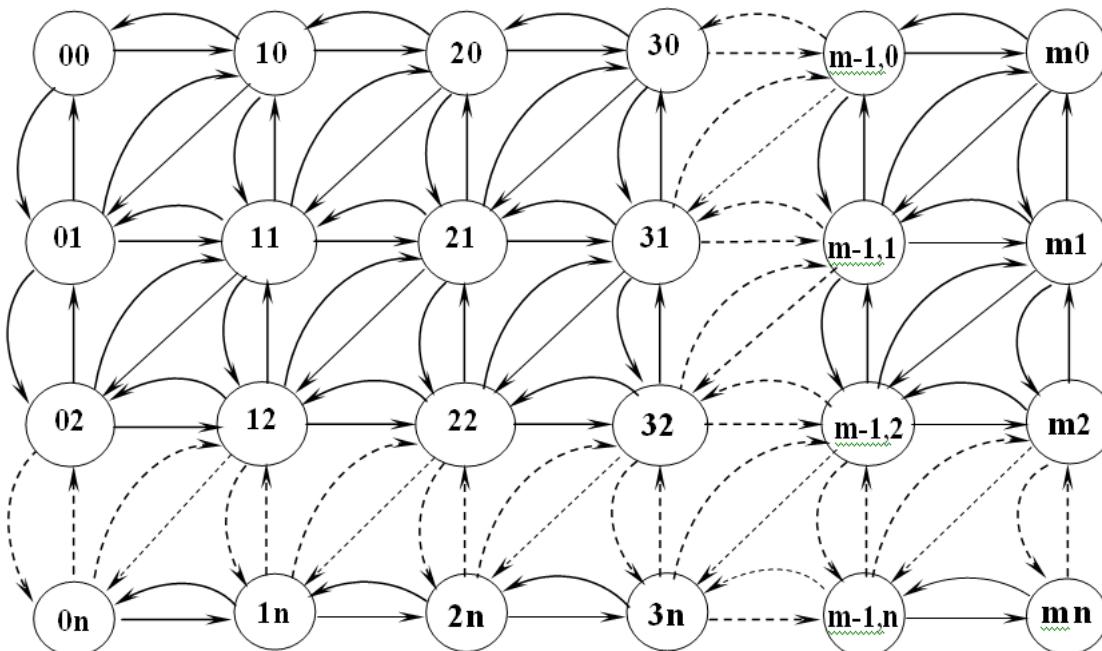


Рис. П. 1

Матрицы A_i, F_i, B_i, D_i объединяют соответствующие интенсивности переходов между векторными состояниями марковского графа.

Матрица A_i объединяет интенсивности прибытия заявки из векторного состояния i в соседнее векторное состояние $(i + 1)$.

Матрица F_i объединяет интенсивности переходов из состояния в состояние в пределах элементов векторного состояния i .

Матрица B_i объединяет интенсивности переходов из векторного состояния i в соседнее векторное состояние $(i - 1)$.

Матрица D_i является диагональной и описывает уходы заявки из векторного состояния i в соседние состояния $(i+1)$ и $(i-1)$.

Для графа состояний с векторными состояниями, содержащими одинаковое число $(n+1)$ компонентов, A_i , F_i , B_i , D_i являются квадратными матрицами размерности $(n+1) \times (n+1)$.

В общем случае каждое векторное состояние i может иметь разное число σ_i компонентов: $P_i = \{P_{i0}, P_{i1}, \dots, P_{i\sigma_i}\}$, тогда матрицы A_i , F_i , B_i , D_i приобретают следующую размерность:

$$A_i [\sigma_i \times \sigma_{i+1}], F_i [\sigma_i \times \sigma_i], B_i [\sigma_i \times \sigma_{i-1}], D_i [\sigma_i \times \sigma_i].$$

Свойства матриц A_i , F_i , B_i , D_i схематично поясняются на рис. П.2, П.3, П.4, П.5. На этих рисунках стрелками схематично показаны возможные переходы из состояния в состояние, задаваемые матрицами A_i , F_i , B_i , D_i . Так матрица A_i представляет переходы из микросостояний вектора состояний P_i в микросостояния вектора P_{i+1} при поступлении заявки в систему, рис. П.2. На рис. П.3 ориентация стрелок означает, что матрица F_i описывает переходы в пределах микросостояний вектора P_i . Стрелки на рис. П.4 направлены справа налево, так как матрица B_i описывает переходы из микросостояний вектора P_i в микросостояния вектора P_{i-1} при убытии заявки из системы в связи с завершением обслуживания.

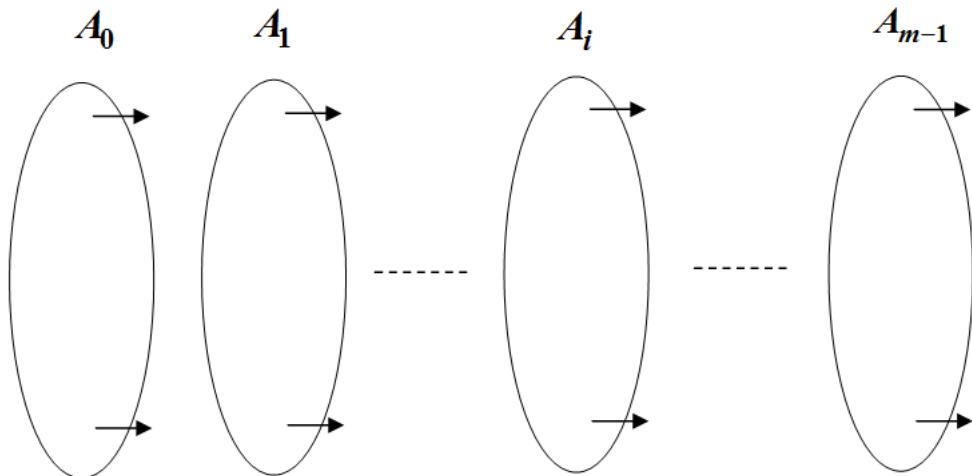


Рис. П.2

Двунаправленная ориентация стрелок на рис. П.5 означает, что матрица D_i описывает уходы заявки из состояния P_i по любой причине: в случае прибытия заявки (стрелка слева направо к состоянию P_{i+1}) или в случае

убытия заявки (стрелка справа налево к состоянию P_{i-1}). Матрицам D_0 и

D_m на рис. П5 соответствуют односторонние стрелки по вполне

очевидной причине: если в системе нет заявок, то уход из векторного состояния P_0 возможен только в микросостояния вектора P_1 в связи с прибытием заявки (стрелка справа налево), если все места в системе заняты (состояние P_m), то уход из этого векторного состояния возможен только в микросостояния вектора P_{m-1} в связи с убытием заявки из системы (стрелка слева направо).

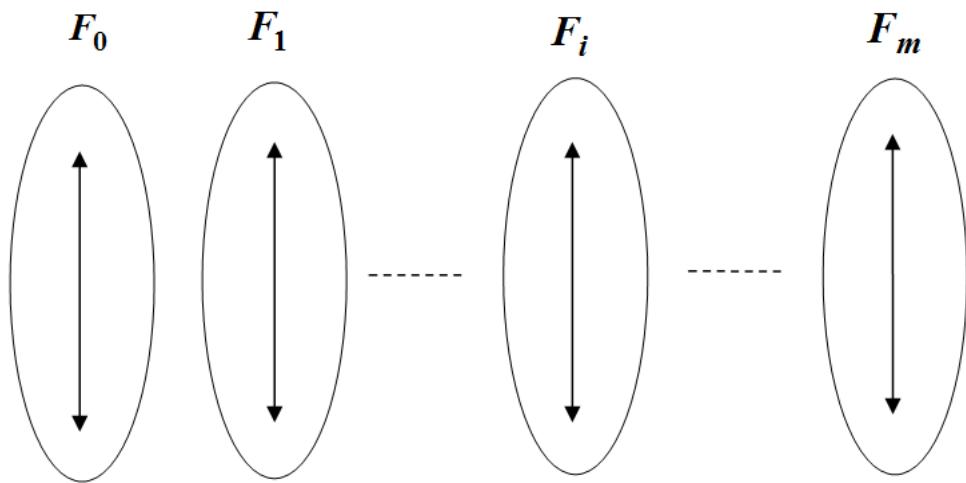


Рис. П.3

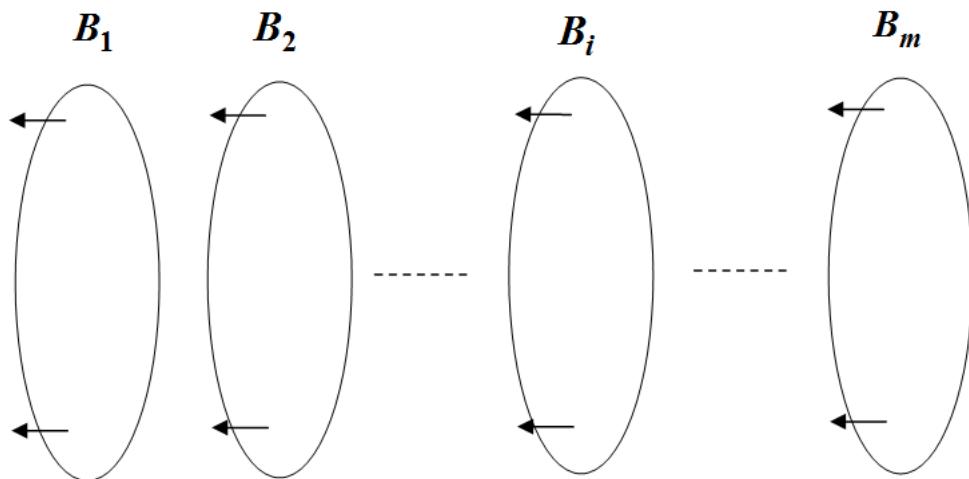


Рис. П.4

Для графа, представленного на рис. 2.5 и продублированного в данном приложении (рис. П.5), в соответствии с предложенной схемой полу-

чены следующие матрицы A_0 , F_0 , B_1 , D_0 .

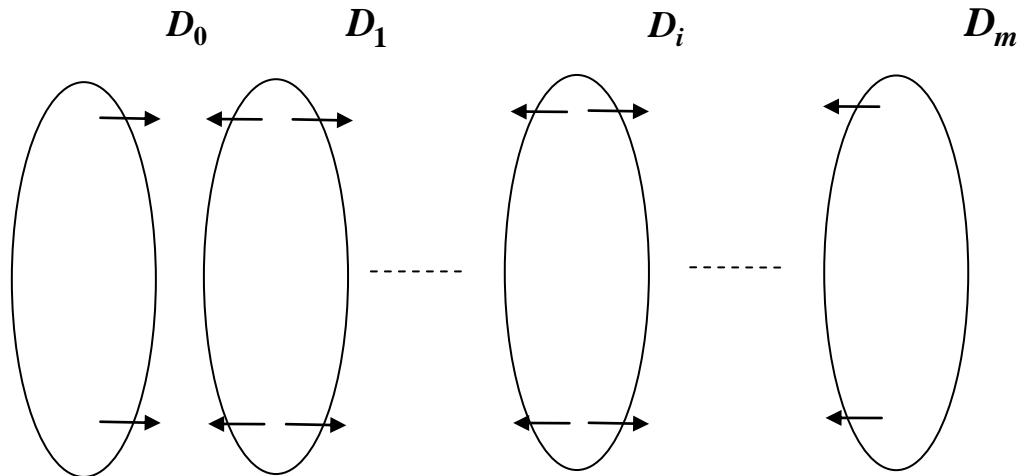


Рис. П.5

$$A_0 = \begin{cases} \overbrace{\begin{bmatrix} 10 & 11 & 12 \\ \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}}^{\text{00}} \\ \overbrace{\begin{bmatrix} 00 & 01 & 02 \\ 0 & 0 & 0 \\ \mu_2 & 0 & 0 \\ 0 & \mu_2 & 0 \end{bmatrix}}^{\text{01}} \\ \overbrace{\begin{bmatrix} 10 & 01 & 02 \\ 0 & \mu_1 & 0 \\ 0 & 0 & \mu_1 \\ 0 & 0 & 0 \end{bmatrix}}^{\text{02}} \end{cases}, \quad F_0 = \begin{cases} \overbrace{\begin{bmatrix} 00 & 01 & 02 \\ 0 & 0 & 0 \\ \mu_2 & 0 & 0 \\ 0 & \mu_2 & 0 \end{bmatrix}}^{\text{00}} \\ \overbrace{\begin{bmatrix} 01 & 02 \\ \mu_2 & 0 \end{bmatrix}}^{\text{01}} \\ \overbrace{\begin{bmatrix} 02 \\ 0 \end{bmatrix}}^{\text{02}} \end{cases}, \quad B_1 = \begin{cases} \overbrace{\begin{bmatrix} 00 & 01 & 02 \\ 0 & 0 & 0 \\ \mu_1 & 0 & 0 \\ 0 & \mu_1 & 0 \end{bmatrix}}^{\text{10}} \\ \overbrace{\begin{bmatrix} 11 \\ 0 \end{bmatrix}}^{\text{11}} \\ \overbrace{\begin{bmatrix} 12 \\ 0 \end{bmatrix}}^{\text{12}} \end{cases},$$

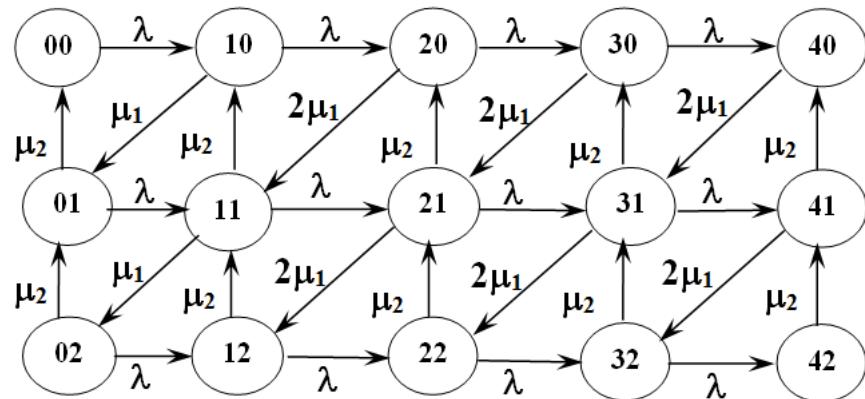


Рис. П.6

$$D_0 = \begin{cases} \overbrace{\begin{bmatrix} 00 & 01 & 02 \\ \lambda & 0 & 0 \\ 0 & (\lambda + \mu_2) & 0 \\ 0 & 0 & (\lambda + \mu_2) \end{bmatrix}}^{\text{00}} \\ \overbrace{\begin{bmatrix} 01 & 02 \\ (\lambda + \mu_2) & 0 \end{bmatrix}}^{\text{01}} \\ \overbrace{\begin{bmatrix} 02 \\ 0 \end{bmatrix}}^{\text{02}} \end{cases}.$$

Библиографический список

1. **Клейнрок Л.** Теория массового обслуживания / Л. Клейнрок. — М. : Машиностроение, 1979. — 431 с.
2. **Гнеденко Б. В.** Введение в теорию массового обслуживания / Б. В. Гнеденко, И. Н. Коваленко. — М. : Наука, 1987. — 336 с.
3. **Клейнрок Л.** Вычислительные системы с очередями / Л. Клейнрок. — М. : Мир, 1979. — 598 с.
4. **Денисов А. А.** Теория больших систем управления / А. А. Денисов, Д. Н. Колесников. — Л. : Энергоиздат, 1982. — 286 с.
5. **Вишневский В. М.** Теоретические основы проектирования компьютерных сетей / В. М. Вишневский. — М. : Техносфера, 2003. — 512 с.
6. **Bitran, G., Morabito, R.** Open Queueing Networks: Optimization and Performance Evaluation Models for Discrete Manufacturing Systems. Technical Report WP#3743–94, Sloan School of Management, MIT, Cambridge, MA, EUA. 1994. 45 p.
7. **Башарин Г. П.** Анализ очередей в вычислительных сетях. Теория и методы расчета / Г. П. Башарин, П. П. Бочаров, Я. А. Коган. — М. : Наука, 1989. — 336 с.
8. **Рыжиков Ю. И.** Расчет разомкнутых немарковских сетей с преобразованием потоков / Ю. И. Рыжиков, А. Д. Хомоненко // АВТ. — 1989. — № 3. — С. 15–24.
9. **Смагин В. А.** Аппроксимационный метод расчета разомкнутых сетей массового обслуживания / В. А. Смагин, Г. В. Филимонихин // АВТ. — 1986. — № 4. С. 28–33.
10. **Рыжиков Ю. И.** Численные методы теории очередей: учеб. пособие / Ю. И. Рыжиков. — СПб. : Лань, 2019. — 512 с.

11. **Клейнрок Л.** Коммуникационные сети (стохастические потоки и задержки сообщений) / Л. Клейнрок. — М. : Наука, 1970. — 256 с.
12. Системный анализ и принятие решений : учеб. пособие / Д. Н. Колесников [и др.] ; под ред. Д. Н. Колесникова. — СПб. : Изд-во Политехн. ун-та, 2008. — 468 с.
13. **Каминский В. Н.** Оптимизация замкнутых стохастических сетей с экспоненциальным обслуживанием / В. Н. Каминский. Известия АН СССР. Техническая кибернетика. — 1980. — № 6. С. 68–76.
14. **Устинов С. М.** Вычислительная математика / С. М. Устинов, В. А. Зимницкий. — СПб. : БХВ-Петербург, 2009. — 336 с.
15. Решение задач автоматики и вычислительной техники методами теории массового обслуживания : учеб. пособие / под ред. Д. Н. Колесникова ; ЛПИ. — Л., 1987. — 77 с.
16. **Рыжиков Ю. И.** Вычислительные методы / Ю. И. Рыжиков. — СПб. : БХВ-Петербург, 2007. — 400 с.
17. Моделирование сложных систем с использованием сетей массового обслуживания / Дурандин К. П. [и др.]; ЛПИ. — Л., 1981 – 83 с.
18. Моделирование систем с использованием теории массового обслуживания: учеб. пособие / сост. Д. Н. Колесников [и др.] ; под ред. Д. Н. Колесникова. — СПб. : Изд-во Политехн. ун-та, 2009. — 228 с.
19. **Башарин Г. П.** Матричный метод нахождения стационарного распределения для некоторых нестандартных систем массового обслуживания / Г. П. Башарин, А. И. Громов // Автоматика и телемеханика — 1978. — выпуск 1. С. 29–38.
20. **Рыжиков Ю. И.** Полный расчет системы обслуживания с распределениями Кокса / Ю. И. Рыжиков // Информационные каналы и среды — 2006. — № 2. С. 38–46.
21. **Рыжиков Ю. И.** Теория очередей и управление запасами: учеб. пособие /Ю. И. Рыжиков. — СПб. : Питер, 2001. – 384 с.