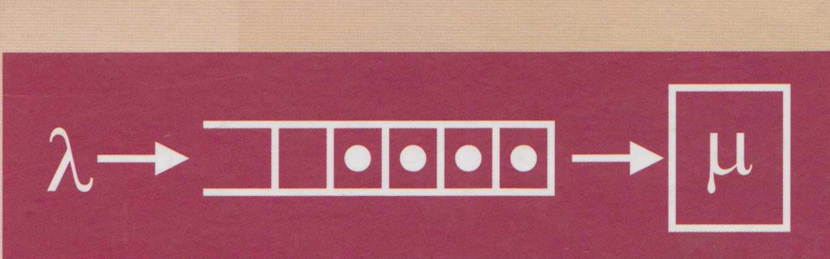
ТОМСКОГО УЧЕБНИКИ УНИВЕРСИТЕТА

А.А. Назаров

А.Ф. Терпугов

ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

А.А. Назаров, А.Ф. Терпугов

ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ

Рекомендовано Сибирским региональным учебно-методическим центром высшего профессионального образования для межвузовского использования в качестве учебного пособия для студентов, обучающихся по специальностям 010200 (010501) «Прикладная математика и информатика», 061800 (080116) «Математические методы в экономике»

ИЗДАТЕЛЬСТВО НАУЧНО-ТЕХНИЧЕСКОЙ ЛИТЕРАТУРЫ

Томск – 2010

УДК 519.248
ББК 22.171
Н 192

Н 192 **Назаров А.А., Терпугов А.Ф.** Теория массового обслуживания: учебное пособие. – 2-е изд., испр. – Томск: Изд-во НТЛ. 2010. – 228 с.

ISBN 5–89503–233–8

Материал, изложенный в учебном пособии, соответствует программе курса «Теория массового обслуживания». Он содержит изложение необходимого для дальнейшего математического аппарата – уравнений в конечных разностях и преобразования Лапласа – Стилтьеса. Рассматривается теория случайных потоков. Подробно изложены все основные модели теории массового обслуживания и методы их исследования – марковские и полумарковские модели, эргодические методы исследования, асимптотические методы. Изложено также применение моделей теории массового обслуживания к исследованию компьютерных сетей связи с маркерным и случайным множественным доступом.

От читателя требуется математическая подготовка в объеме курса теории вероятностей и случайных процессов.

Для студентов, специализирующихся по прикладной математике и математическим методам в экономике. Книга будет полезна аспирантам, научным работникам, инженерам, экономистам и представителям других специальностей, занимающихся приложением математических методов и, в частности, моделей и методов теории массового обслуживания.

УДК 519.248
ББК 22.171

ISBN 5–89503–233–8

Содержание

Предисловие

Теория массового обслуживания входит как основной либо как специальный курс в учебные планы различных специальностей. Особенно полное ее изложение требуется для студентов специальностей 010200 – «Прикладная математика» и 061800 – «Математические методы в экономике», которые изучают теорию вероятностей, теорию случайных процессов, математическую статистику, а также некоторые специальные разделы современной прикладной математики.

С позиций математика теория массового обслуживания представляет собой своеобразную задачу теории случайных процессов. Действительно, случайный поток, являющийся не чем иным, как целочисленным монотонно возрастающим случайным процессом, подвергается некоторой трансформации системой обслуживания. Требуется найти характеристики результата воздействия этой трансформации. Такими характеристиками в зависимости от поставленной задачи исследования могут быть число заявок в системе, время ожидания начала обслуживания, вероятность потери заявки, продолжительность периода занятости и т.д.

Это замечание поясняет определяющее значение теории вероятностей и теории случайных процессов для теории массового обслуживания, изучение которой невозможно без знания основ перечисленных дисциплин.

При отборе материала авторы руководствовались стремлением изложить, прежде всего, методы исследования систем массового обслуживания. Поэтому в первой главе рассмотрены линейные уравнения в конечных разностях, которые широко используются для анализа марковских моделей, а также преобразование Лапласа – Стилтьеса, применяемое при анализе временных характеристик систем обслуживания. Вторая глава посвящена теории случайных потоков однородных событий – основного элемента систем массового обслуживания. Здесь рассмотрены как потоки без последействия (пуассоновский и простейший), так и потоки с ограниченным последействием (рекуррентные потоки).

В третьей главе рассмотрены марковские системы обслуживания и показано, что их анализ определяется исследованием цепей Маркова с непрерывным временем, то есть дискретных марковских процессов, описывающих процесс изменения во времени состояний таких систем.

Четвертая глава посвящена полумарковским моделям, в которых либо длительности интервалов входящего потока, либо продолжительности времени обслуживания имеют произвольную функцию распределения. Поэтому процесс изменения состояний в таких системах является полумарковским, и для его исследования применяются как методы вложенных цепей Маркова, так и методы дополнительной переменной, то есть методы марковизации полумарковских моделей.

Перечисленные главы составляют основное содержание классической теории массового обслуживания, изучение которой необходимо студентам всех специальностей, чьи учебные планы включают названную дисциплину.

В следующих трех главах рассмотрены более специальные системы обслуживания: немарковские, приоритетные и циклические, для исследования которых применяется значительное количество разнообразных оригинальных методов, решающих определенные классы задач для таких моделей.

Последняя, восьмая, глава посвящена исследованию систем обслуживания с повторными вызовами, которые являются математическими моделями компьютерных сетей связи, управляемых протоколами случайного множественного доступа. Здесь для их исследования применяется метод асимптотического анализа, рассмотренный для марковских моделей.

Излагаемый в книге материал читается как специальный курс для студентов факультета прикладной математики и кибернетики Томского госуниверситета. Книга будет полезна аспирантам и специалистам, занимающимся приложениями математических методов и, в частности, методов теории массового обслуживания.

Глава 1. МАТЕМАТИЧЕСКИЙ АППАРАТ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

§ 1.1. Уравнения в конечных разностях – общие замечания

Теория массового обслуживания использует специфический математический аппарат, обычно не излагаемый в курсах математического анализа. К числу специфических уравнений, которые очень часто встречаются в теории массового обслуживания, относятся уравнения в конечных разностях.

Пусть функция $y(i)$ определена лишь для **целочисленных** значений аргумента i, скажем, для $i = 0,\ 1,\ 2,\ 3,\ \ldots$. Уравнение вида

$$\Phi(i,\ y(i),\ y(i+1),\ y(i+2),\ \ldots,\ y(i+n)) = 0 \qquad (1.1)$$

называется уравнением в конечных разностях. Решением этого уравнения называется любая функция $y(i)$, которая при подстановке ее в левую часть (1.1) обращает уравнение в тождество.

Порядок уравнения

Если в левой части (1.1) присутствует $y(i)$ и $y(i+n)$, то уравнение (1.1) называется уравнением в конечных разностях n-го порядка. Если мы будем иметь дело с уравнением вида

$$\Phi\big(i,\ y(i+k),\ y(i+k+1),\ \ldots,\ y(i+n)\big) = 0,$$

то, вводя функцию $z(i) = y(i+k)$, мы приведем его к виду

$$\Phi(i,\ z(i),\ z(i+1),\ \ldots,\ z(i+n-k)) = 0,$$

откуда видно, что в данном случае мы имеем дело с уравнением $(n-k)$-го порядка. Вообще, порядок уравнения, есть разность между максимальным и минимальным значениями аргумента функции y, которые присутствуют в уравнении. Например, уравнение

$$\Phi(i,\ y(i-1),\ y(i),\ y(i+1)) = 0$$

имеет второй порядок, так как $(i+1)-(i-1) = 2$.

Вид решения уравнения

Представим себе, что уравнение (1.1) можно разрешить относительно $y(i+n)$:

$$y(i+n) = f(i,\ y(i),\ y(i+1),\ ...,\ y(i+n-1)), \tag{1.2}$$

где $f(\cdot)$ – однозначная функция.

Тогда, задавая (произвольно) значения $y(0),\ y(1),\ ...,\ y(n-1)$, можно, беря $i=0$, получить $y(n)$; затем, так как $y(n)$ известно, беря $i=1$, получить $y(n+1)$. Продолжая эту процедуру, можно получить единственную функцию $y(i)$ для любых $i>0$. Таким образом, решение уравнения (1.2) зависит от n произвольных констант

$$y(i) = \varphi(i,\ C_1,\ C_2,\ ...,\ C_n), \tag{1.3}$$

которыми, в частности, могут быть $y(0),\ y(1),\ ...,\ y(n-1)$.

Заметим, что верно и обратное. Если мы имеем функцию $y(i)$, зависящую от n произвольных констант (1.3), то, беря $i=i,\ i+1,\ ...,\ i+n-1$, получим

$$\begin{cases} y(i) = \varphi(i,\ C_1,\ C_2,\ ...,\ C_n), \\ y(i+1) = \varphi(i+1,\ C_1,\ C_2,\ ...,\ C_n), \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ y(i+n-1) = \varphi(i+n-1,\ C_1,\ C_2,\ ...,\ C_n). \end{cases}$$

В данной системе n уравнений. Из нее можно выразить $C_1,\ C_2,\ ...,\ C_n$ через $y(i),\ y(i+1),\ ...,\ y(i+n-1)$. Подставляя эти выражения в (1.3) при $i=i+n$, получим, что

$$y(i+n) = f(i,\ y(i),\ y(i+1),\ ...,\ y(i+n-1)),$$

то есть уравнение вида (1.2).

Таким образом, общее решение уравнения в конечных разностях n-го порядка зависит от n произвольных констант. Наоборот, всякая функция дискретного аргумента, зависящая от n произвольных констант, является решением некоторого уравнения в конечных разностях n-го порядка.

§ 1.2. Линейные однородные уравнения в конечных разностях

Уравнение вида

$$a_0 y(i+n) + a_1 y(i+n-1) + \ldots + a_n y(i) = 0\,, \qquad (1.4)$$

где $a_0 \neq 0$ и $a_n \neq 0$, называется линейным однородным уравнением в конечных разностях n-го порядка с постоянными коэффициентами. Уравнения такого вида в теории массового обслуживания встречаются очень часто.

Любая функция $y(i)$, которая обращает (1.4) в тождество, называется частным решением уравнения (1.4).

Теорема 1. Пусть $y_1(i)$, $y_2(i) \ldots y_k(i)$ есть некоторые частные решения уравнения (1.4). Тогда функция

$$y(i) = C_1 y_1(i) + C_2 y_2(i) + \ldots + C_k y_k(i)\,,$$

где C_1, $C_2, \ldots$, C_k – произвольные константы, есть также решение уравнения (1.4).

Доказательство

Пусть $y_s(i)$ есть решение уравнения (1.4), то есть

$$\sum_{l=0}^{n} a_{n-l} y_s(i+l) = 0\,.$$

Поэтому

$$\sum_{l=0}^{n} a_{n-l} y(i+l) = \sum_{l=0}^{n} a_{n-l} \sum_{s=1}^{k} C_s y_s(i+l) = \sum_{s=1}^{k} C_s \left(\sum_{l=0}^{n} a_{n-l} y_s(i+l) \right) = 0\,,$$

откуда и следует доказываемое.

Прежде чем формулировать и доказывать следующую теорему, напомним понятие линейной независимости функций.

Определение. Функции $y_1(i)$, $y_2(i) \ldots y_k(i)$ называются линейно независимыми, если условие

$$C_1 y_1(i) + C_2 y_2(i) + \ldots + C_k y_k(i) \equiv 0$$

выполняется только при $C_1 = C_2 = \ldots = C_k = 0$. Если это условие выполняется, когда хотя бы одна из $C_s \neq 0$, то функции $y_1(i)$, $y_2(i), \ldots, y_k(i)$ называются линейно зависимыми.

Теорема 2.1. Если функции $y_1(i)$, $y_2(i), \ldots, y_n(i)$ линейно зависимы, то определитель

$$D\left[y_1(i),\ y_2(i),\ \ldots\ ,y_n(i)\right]=\begin{vmatrix} y_1(i) & y_2(i) & \ldots & y_n(i) \\ y_1(i+1) & y_2(i+1) & \ldots & y_n(i+1) \\ \ldots & \ldots & \ldots & \ldots \\ y_1(i+n-1) & y_2(i+n-1) & \ldots & y_n(i+n-1) \end{vmatrix}\equiv 0 .$$

Доказательство

Если функции линейно зависимы, то один из столбцов определителя является линейной комбинацией остальных столбцов, следовательно, определитель равен нулю.

Следствие. Если $D\neq 0$, то $y_n(i)$ линейно независимы.

Теорема 2.2. Если линейно независимые функции $y_1(i)$, $y_2(i),\ldots,y_n(i)$ являются решениями линейного однородного в конечных разностях уравнения

$$y(i+n)=b_1y(i+n-1)+b_2y(i+n-2)+\ldots+b_ny(i), \qquad (1.4а)$$

то определитель $D\left[y_1(i),\ y_2(i),\ \ldots\ ,y_n(i)\right]$ не может обращаться в нуль ни в одной точке $i=0,\ 1,\ 2,\ \ldots$

Доказательство

Пусть для некоторого i_0 $D(i_0)=0$, тогда существует набор постоянных α_1, $\alpha_2,\ldots,$ α_n, не равных нулю одновременно, таких, что система уравнений

$$\alpha_1y_1(i_0)+\alpha_2y_2(i_0)+\ldots+\alpha_ny_n(i_0)=0,$$

$$\alpha_1y_1(i_0+1)+\alpha_2y_2(i_0+1)+\ldots+\alpha_ny_n(i_0+1)=0,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\alpha_1y_1(i_0+n-1)+\alpha_2y_2(i_0+n-1)+\ldots+\alpha_ny_n(i_0+n-1)=0 \qquad (1.4б)$$

относительно α_1, $\alpha_2,\ldots,$ α_n имеет нетривиальное решение.

При таком выборе констант α_1, $\alpha_2,\ldots,$ α_n функция

$$y(i)=\alpha_1y_1(i)+\alpha_2y_2(i)+\ldots+\alpha_ny_n(i)$$

будет также решением уравнения (1.4), удовлетворяющим нулевым условиям (1.4б), но этим условиям удовлетворяет и решение $y(i)\equiv 0$, в силу единственности решения

$$\alpha_1y_1(i_0)+\alpha_2y_2(i_0)+\ldots+\alpha_ny_n(i_0)\equiv 0,$$

где не все $\alpha_1, \alpha_2, \ldots, \alpha_n$ равны нулю. Следовательно, функции $y_1(i)$, $y_2(i), \ldots, y_n(i)$ линейно зависимы. Полученное противоречие доказывает теорему.

Теорема 3. Пусть $y_1(i)$, $y_2(i) \ldots y_n(i)$ есть n линейно независимых решений уравнения $\sum_{l=0}^{n} a_{n-l} y(i+l) = 0$. Тогда общее решение этого уравнения имеет вид

$$y(i) = C_1 y_1(i) + C_2 y_2(i) + \ldots + C_n y_n(i).$$

Доказательство

Как уже говорилось ранее, решение уравнения в конечных разностях n-го порядка полностью определяется заданием $y(0)$, $y(1)$, …, $y(n-1)$.

По теореме 1 $y(i)$ есть решение нашего уравнения. Попробуем подобрать константы C_1, $C_2, \ldots$, C_n так, чтобы выполнялись начальные условия

$$\begin{cases} C_1 y_1(0) + C_2 y_2(0) + \ldots + C_n y_n(0) = y(0), \\ C_1 y_1(1) + C_2 y_2(1) + \ldots + C_n y_n(1) = y(1), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ C_1 y_1(n-1) + C_2 y_2(n-1) + \ldots + C_n y_n(n-1) = y(n-1) \end{cases}$$

для заданного частного решения $y(i)$.

В силу линейной независимости решений $y_s(i)$ детерминант этой системы, равный $D\left[y_1(0),\ y_2(0), \ldots, y_n(0)\right]$, отличен от нуля. Поэтому эта система имеет единственное решение и константы C_1, $C_2, \ldots$, C_n находятся однозначно.

Так как $y(i)$ есть решение и выполнены начальные условия, то оно и есть то решение, которое удовлетворяет нашему уравнению при заданных начальных условиях.

Таким образом, для решения однородного линейного уравнения в конечных разностях n-го порядка надо найти n его линейно независимых решений, линейная комбинация которых и даст его общее решение.

§ 1.3. Решение однородных линейных уравнений в конечных разностях с постоянными коэффициентами

Рассмотрим уравнение

$$a_0 y(i+n) + a_1 y(i+n-1) + \ldots + a_n y(i) = \sum_{l=0}^{n} a_{n-l} y(i+l) = 0 . \qquad (1.5)$$

Попробуем найти его решение в виде $y(i) = z^i$. Подставляя это решение в уравнение, запишем

$$a_0 z^{i+n} + a_1 z^{i+n-1} + \ldots + a_n z^i = 0 ,$$

откуда, сокращая на z^i, получим уравнение

$$a_0 z^n + a_1 z^{n-1} + \ldots + a_n = 0 . \qquad (1.6)$$

Полином

$$P(z) = a_0 z^n + a_1 z^{n-1} + \ldots + a_n$$

называется характеристическим полиномом уравнения (1.5), а уравнение (1.6) – характеристическим уравнением.

Рассмотрим подробно возникающие здесь ситуации.

А. Все корни уравнения (1.6) вещественны и различны

Как известно, алгебраическое уравнение степени *n* имеет всего *n* корней z_1, $z_2, \ldots$, z_n. Если все они различны, то мы имеем *n* частных решений вида

$$y_s(i) = z_s^i .$$

Проверим их линейную независимость. Определитель $D\left[y_1(i),\ y_2(i),\ \ldots\ , y_n(i)\right]$ примет вид

$$D\left[y_1(i),\ y_2(i),\ \ldots\ , y_n(i)\right] = \begin{vmatrix} z_1^i & z_2^i & \ldots & z_n^i \\ z_1^{i+1} & z_2^{i+1} & \ldots & z_n^{i+1} \\ \ldots & \ldots & \ldots & \ldots \\ z_1^{i+n-1} & z_2^{i+n-1} & \ldots & z_n^{i+n-1} \end{vmatrix} =$$

$$= z_1^i z_2^i \cdot \ldots \cdot z_n^i \cdot \begin{vmatrix} 1 & 1 & \ldots & 1 \\ z_1^1 & z_2^1 & \ldots & z_n^1 \\ \ldots & \ldots & \ldots & \ldots \\ z_1^{n-1} & z_2^{n-1} & \ldots & z_n^{n-1} \end{vmatrix} .$$

Здесь $z_1 z_2 \cdot \ldots \cdot z_n = (-1)^n a_n \neq 0$.

Заметим, что все z_i отличны от нуля, иначе было бы $a_n = 0$. Сам оставшийся определитель называется определителем Вандермонда. Для него есть явное выражение, используя которое, получим

$$D\left[y_1(i),\ y_2(i),\ \ldots\ , y_n(i)\right] = \prod_{s=1}^{n} z_s^i \cdot \prod_{i \neq j} \left(z_i - z_j\right).$$

Так как, по предположению, все z_i различны, то $D\left[y_1(i),\ y_2(i),\ \ldots\ , y_n(i)\right] \neq 0$. По теореме 3 отсюда следует, что общее решение (1.5) имеет вид

$$y(i) = C_1 z_1^i + C_2 z_2^i + \ldots + C_n z_n^i.$$

Б. Все корни характеристического уравнения (1.6) различны, но среди них есть комплексные

Как известно, если все коэффициенты a_n вещественны, то комплексные корни всегда встречаются парами, то есть, если есть корень $z = \alpha + j\beta$, то есть и корень $\overline{z} = \alpha - j\beta$, где $j = \sqrt{-1}$. Соответствующая пара слагаемых в общем решении имеет вид

$$\overline{C}_1 (\alpha + j\beta)^i + \overline{C}_2 (\alpha - j\beta)^i. \tag{1.7}$$

Если представить комплексное число z в тригонометрической форме $z = r[\cos\varphi + j\sin\varphi]$, то $z^i = r^i[\cos i\varphi + j\sin i\varphi]$ и $\overline{z}^i = r^i[\cos i\varphi - j\sin i\varphi]$, и пара слагаемых (1.7) примет вид

$$\overline{C}_1 r^i [\cos i\varphi + j\sin i\varphi] + \overline{C}_2 r^i [\cos i\varphi - j\sin i\varphi] =$$

$$= \left(\overline{C}_1 + \overline{C}_2\right) r^i \cos i\varphi + j\left(\overline{C}_1 - \overline{C}_2\right) r^i \sin i\varphi.$$

Обозначив $\left(\overline{C}_1 + \overline{C}_2\right) = C_1$, $j\left(\overline{C}_1 - \overline{C}_2\right) = C_2$, получим, что паре комплексно сопряженных простых корней соответствует пара слагаемых

$$C_1 r^i \cos i\varphi + C_2 r^i \sin i\varphi,$$

где $r = |z|$, $\varphi = \arg(z) \neq k\pi$, так как

$$\begin{vmatrix} \cos i\varphi & \sin i\varphi \\ \cos(i+1)\varphi & \sin(i+1)\varphi \end{vmatrix} = \sin\varphi \neq 0.$$

В. Все корни характеристического уравнения (1.6) вещественны, но среди них есть кратные

Пусть z – корень уравнения $P(z)=0$ кратности s. Как известно, в этом случае выполняются условия

$$P(z)=0\,,\quad P'(z)=0\,,\quad P''(z)=0\,,\ \ldots,\ P^{(s-1)}(z)=0\,.$$

Заметим, что при этом будут выполнены также и условия

$$z^iP(z)=0\,,\ \left[z^iP(z)\right]'=0\,,\ \left[z^iP(z)\right]''=0\,,\ldots,\ \left[z^iP(z)\right]^{(s-1)}=0\,,\qquad(1.8)$$

так как по формуле Лейбница

$$\left[z^iP(z)\right]^{(k)}=\sum_{r=0}^{k}C_k^r(z^i)^{k-r}P^{(r)}(z)=0$$

и при $k\le s-1$ все сомножители $P^{(r)}(z)$ в стоящим под знаком суммы выражении равны нулю.

В явном виде условия (1.8) примут вид

$$a_0z^{i+n}+a_1z^{i+n-1}+\ldots+a_nz^i=0\,,$$

$$a_0(i+n)z^{i+n-1}+a_1(i+n-1)z^{i+n-2}+\ldots+a_niz^{i-1}=0\,,$$

$$a_0(i+n)(i-1+n)z^{i+n-2}+a_1(i+n-1)(i-1+n-1)z^{i+n-3}+\ldots+a_ni(i-1)z^{i-2}=0\,,$$

....

$$a_0(i+n)(i-1+n)\ldots(i-(s-1)-n)z^{i+n-(s-1)}+\ldots+a_ni(i-1)\ldots(i-(s-1))z^{i-(s-1)}=0.$$

Если второе равенство умножить на z, третье на z^2, ..., последнее на z^{s-1}, то получим

$$a_0z^{i+n}+a_1z^{i+n-1}+\ldots+a_nz^i=0\,,$$

$$a_0(i+n)z^{i+n}+a_1(i+n-1)z^{i+n-1}+\ldots+a_niz^i=0\,,$$

$$a_0(i+n)(i-1+n)z^{i+n}+a_1(i+n-1)(i-1+n-1)z^{i+n-1}+\ldots+a_ni(i-1)z^i=0\,,$$

....

$$a_0(i+n)(i-1+n)\ldots(i-(s-1)-n)z^{i+n}+\ldots+a_ni(i-1)\ldots(i-(s-1))z^i=0\,.$$

Отсюда видно, что исходному уравнению удовлетворяют следующие функции:

$$y_1(i)=z^i\,,\ y_2(i)=iz^i\,,\ y_3(i)=i(i-1)z^i\,,\ \ldots,\ y_{s-1}(i)=i(i-1)\ldots(i-(s-1))z^i\,.$$

Легко видеть, что вместо этой системы функций можно взять систему

$$y_1(i)=z^i,\quad y_2(i)=iz^i,\quad y_3(i)=i^2z^i,\quad \ldots,\quad y_{s-1}(i)=i^{s-1}z^i.$$

Таким образом, от корня z кратности s в общем решении идет следующая группа слагаемых:

$$z^i\left(C_0+C_1i+C_2i^2+\ldots+C_{s-1}i^{s-1}\right).$$

§ 1.4. Решение линейных неоднородных уравнений в конечных разностях с постоянными коэффициентами

Перейдем теперь к решению уравнений вида

$$a_0y(i+n)+a_1y(i+n-1)+\ldots+a_ny(i)=q(i), \tag{1.9}$$

в котором $a_0\neq 0$, $a_n\neq 0$ и $q(i)$ – некоторая произвольная функция от i.

Теорема 1. Общее решение неоднородного уравнения (1.9) можно представить в виде суммы частного решения уравнения (1.9) и общего решения соответствующего однородного уравнения.

Доказательство

Пусть имеем решение $y(i)=\varphi(i)$ уравнения (1.9) с начальными условиями $y(l)=\varphi_l$, $l=\overline{0,\ n-1}$.

Пусть $y_0(i)$ есть какое-то частное решение уравнения (1.9), т.е. $y_0(i)$ удовлетворяет уравнению

$$\sum_{l=0}^{n}a_{n-l}y_0(i+l)=q(i).$$

Представим решение $\varphi(i)$ уравнения (1.9) в виде $\varphi(i)=y_0(i)+\eta(i)$. Тогда

$$\sum_{l=0}^{n}a_{n-l}\varphi(i+l)=\sum_{l=0}^{n}a_{n-l}y_0(i+l)+\sum_{l=0}^{n}a_{n-l}\eta(i+l)=q(i).$$

Отсюда следует, что $\eta(i)$ удовлетворяет уравнению

$$\sum_{l=0}^{n}a_{n-l}\eta(i+l)=0. \tag{1.10}$$

Как решать такие уравнения, подробно разобрано выше. Записывая общее решение уравнения (1.10) в виде

$$\eta(i) = \sum_{k=1}^{n} C_k y_k(i),$$

получим

$$\varphi(i) = y_0(i) + \sum_{k=1}^{n} C_k y_k(i).$$

Используя начальные условия, получим

$$\sum_{k=1}^{n} C_k y_k(i) = \varphi_l - y_0(l), \qquad l = \overline{0,\ n-1},$$

и, в силу линейной независимости решений $y_k(l)$ линейного однородного уравнения, константы C_k определяются однозначно.

Таким образом, все дело сводится к нахождению частного решения неоднородного уравнения (1.9)

Теорема 2 (*Принцип суперпозиции частных решений линейных неоднородных уравнений в конечных разностях*). Если $y_k(i)$ есть решение линейного неоднородного уравнения в конечных разностях вида

$$a_0 y_k(i+n) + a_1 y_k(i+n-1) + \ldots + a_n y_k(i) = q_k(i),$$

то $y(i) = \sum_{k \in K} y_k(i)$ есть решение уравнения вида

$$a_0 y(i+n) + a_1 y(i+n-1) + \ldots + a_{n-1} y(i+1) + a_n y(i) = \sum_{k \in K} q_k(i).$$

Доказательство (самостоятельно)

Есть различные способы нахождения частного решения линейного неоднородного уравнения. В этом параграфе мы рассмотрим лишь простейший из них, который заключается в том, что в некоторых случаях вид этого частного решения заранее известен и дело лишь в подборе некоторых коэффициентов.

Метод неопределенных коэффициентов

А. Пусть правая часть уравнения (1.9) имеет вид $q(i) = Az^i$, где z **не является** корнем характеристического полинома $P(z)$. Тогда $y(i)$ следует искать в виде $y(i) = Bz^i$. Подставляя это решение в уравнение (1.9), получим

$$B\left[a_0 z^{i+n} + a_1 z^{i+n-1} + \ldots + a_n z^i\right] = Az^i,$$

откуда, сокращая на z^i, получим явное выражение для коэффициента B:

$$B = \frac{A}{a_0 z^n + a_1 z^{n-1} + \ldots + a_n} = \frac{A}{P(z)}.$$

Так как по условию $P(z) \neq 0$, то B определяется однозначно.

Б. Аналогично, если $q(i) = Ar^i \cos i\varphi + Br^i \sin i\varphi$ и комплексное число $z = re^{i\varphi}$ **не является** корнем характеристического полинома, то $y(i)$ следует искать в виде $q(i) = C\ r^i \cos i\varphi + D\ r^i \sin i\varphi$. Подставляя это решение в уравнение (1.9), сокращая на r^i, найдем C и D, приравнивая выражения, стоящие перед $\cos i\varphi$ и $\sin i\varphi$, в левой и правой частях уравнения.

В. Пусть правая часть уравнения (1.9) имеет вид $q(i) = Az^i$, где z **является** корнем полинома $P(z)$, т.е. $P(z) = 0$.

1. Пусть для начала z есть **простой корень** полинома $P(z)$, то есть **корень кратности 1**. Тогда $P(z) = 0$, но $P'(z) \neq 0$.

Будем искать решение $y(i)$ в виде $y(i) = Biz^i$. Подставляя это решение в уравнение (1.9), получим

$$B\left[a_0(i+n)z^{i+n} + a_1(i+n-1)z^{i+n-1} + \ldots + a_n iz^i\right] = Az^i.$$

Раскрывая скобки и сократив на z^i, получим

$$Bi\left[a_0 z^n + a_1 z^{n-1} + \ldots + a_n\right] + B\left[a_0 nz^n + a_1(n-1)z^{n-1} + \ldots + a_{n-1}z\right] = A.$$

Но при Bi сомножителем стоит $P(z) = 0$. При B сомножителем стоит, как легко видеть, выражение $zP'(z) \neq 0$. Поэтому B находится однозначно:

$$B = A/zP'(z),$$

и тем самым найдено частное решение уравнения (1.9).

2. Пусть теперь z – корень характеристического уравнения **кратности** s, то есть $P(z) = P'(z) = \ldots = P^{(s-1)}(z) = 0$, но $P^{(s)}(z) \neq 0$. Тогда решение следует искать в виде $y(i) = Bi^s z^i$. Действительно, подставляя это решение в уравнение (1.9), получим

$$B\left[a_0(i+n)^s z^{i+n} + a_1(i+n-1)^s z^{i+n-1} + \ldots + a_n i^s z^i\right] = Az^i .$$

Сокращая на z^i и собирая коэффициенты при одинаковых степенях i, получим, что коэффициенты при i, i^2, ..., i^s обратятся в нуль, так как они выражаются через $P(z)$, $P'(z)$, ..., $P^{(s-1)}(z)$. Останется лишь выражение

$$B\left[a_0 n^s z^n + a_1(n-1)^s z^{n-1} + \ldots + a_{n-s}\ \right] = A ,$$

которое и определит однозначно коэффициент B.

Г. Наконец, пусть $q(i) = A_m(i)z^i$, где $A_m(i)$ – полином по i степени m, а z – корень полинома $P(z)$ кратности s. Тогда $y(i)$ следует искать в виде

$$y(i) = B_m(i) i^s z^i ,$$

где $B_m(i)$ – полином по i степени m. Коэффициенты этого полинома находятся подстановкой этого решения в исходное уравнение с последующим приравниванием коэффициентов при одинаковых степенях i.

Пример

Пусть $i(t)$ – цепь Маркова с непрерывным временем, инфинитезимальные характеристики которой имеют вид

$$q_{i,i+1} = \lambda , \quad q_{i,i-1} = \mu , \quad q_{i\,i} = -(\lambda + \mu) , \quad q_{i\,j} = 0 , \quad \forall\ j \neq i,\ i+1,\ i-1 .$$

Пусть $T(t,i)$ есть среднее значение длины интервала от момента t, когда $i(t) = i$, до момента попадания цепи в некоторое состояние или множества состояний рассматриваемой цепи. Тогда, рассматривая состояния цепи в момент времени $t + \Delta t$, получим соотношение

$$T(t,i) = \lambda\Delta t T(t+\Delta t, i+1) + \mu\Delta t T(t+\Delta t, i-1) +$$
$$+\left[1-(\lambda+\mu)\Delta t\right]\left(\Delta t + T(t+\Delta t,\ i)\right) + o(\Delta t),$$

откуда следует уравнение для $T(t,i)$:

$$-\frac{\partial T(t,i)}{\partial t} = 1 - (\lambda+\mu)T(t,\ i) + \lambda T(t,\ i+1) + \mu T(t,\ i-1) .$$

В силу однородности цепи Маркова выполнено равенство $T(t,i) = T(i)$; следовательно, $T(i)$ удовлетворяет линейному неоднородному уравнению в конечных разностях

$$\lambda T(i+1)-(\lambda+\mu)T(\ i)+\mu T(i-1)=-1 .$$

Общее решение однородного уравнения имеет вид

$$T_{\text{об.о}}(i)=C_1+C_2\left(\frac{\mu}{\lambda}\right)^i ,$$

так как $z_1=1$, $z_2=\frac{\mu}{\lambda}$.

Так как правая часть нашего уравнения есть константа и $z_1=1$ является простым корнем характеристического уравнения, то, применяя алгоритм B, частное решение $T_ч(i)$ будем искать в виде

$$T_ч(i)=B\cdot i(1)^i=B\cdot i .$$

Подставив это решение в уравнение, получим

$$\lambda B(i+1)-(\lambda+\mu)Bi+\mu B(i-1)=-1 ,$$

$$Bi(\lambda-\lambda+\mu-\mu)+B(\lambda-\mu)=-1 ,$$

и поэтому

$$B=\frac{1}{\mu-\lambda}.$$

Следовательно, общее решение $T(i)$ имеет вид

$$T(i)=C_1+C_2\left(\frac{\mu}{\lambda}\right)^i+\frac{i}{\mu-\lambda}.$$

Значения констант C_1 и C_2 определяются из дополнительных условий.

При $\lambda=\mu$ неоднородное уравнение примет вид

$$T(i+1)-2T(\ i)+T(i-1)=\frac{-1}{\lambda} ,$$

его характеристическое уравнение имеет вид $z^2-2z+1=0$, и, следовательно, $z_{1,2}=1$ есть корень характеристического уравнения кратности 2, что соответствует варианту B для однородных уравнений. Поэтому

$$T_{\text{об.о}}(i)=C_1+C_2 i .$$

Так как $z = 1$ является корнем характеристического уравнения кратности 2, то частное решение будем искать в виде $T(i) = Bi^2$:

$$B(i+1)^2 - 2Bi^2 + B(i-1)^2 = \frac{-1}{\lambda},$$

$$B(i^2 + 2i + 1) - 2Bi^2 + B(i^2 - 2i + 1) = \frac{-1}{\lambda}.$$

Отсюда следует, что $2B = \frac{-1}{\lambda}$, $B = \frac{-1}{2\ \lambda}$, и поэтому общее решение исходного уравнения имеет вид $T(i) = C_1 + C_2 i - \frac{i^2}{2\lambda}$.

§ 1.5. Решение линейных уравнений в конечных разностях методом производящих функций (Z-преобразования)

Пусть $y(i)$ – функция дискретного аргумента i, принимающего значения $i = \overline{0,\infty}$. Функция

$$Y(z) = \sum_{i=0}^{\infty} z^i y(i) \tag{1.11}$$

называется **производящей функцией** (другое название – ***Z*-преобразование**) функции $y(i)$.

Отметим, что не только $y(i)$ однозначно определяет $Y(z)$, но и наоборот, $Y(z)$ однозначно определяет $y(i)$. Действительно, дифференцируя (1.11) k раз по z и затем полагая $z = 0$, получим

$$y(k) = \frac{1}{k!} Y^{(k)}(0).$$

Отметим некоторые простейшие свойства Z-преобразования:

1. Если $Y_1(z) \Leftrightarrow y_1(i)$ и $Y_2(z) \Leftrightarrow y_2(i)$, то

$$C_1 Y_1(z) + C_2 Y_2(z) \Leftrightarrow c_1 y_1(i) + c_2 y_2(i).$$

2. Рассмотрим функцию $y(i+k)$, $i = \overline{0,\infty}$. Тогда

$$y(i+k) \Leftrightarrow \sum_{i=0}^{\infty} y(i+k) z^i =$$

$$=\frac{1}{z^k}\sum_{i=0}^{\infty}y(i+k)z^{i+k}=\frac{1}{z^k}\left(\sum_{s=0}^{\infty}y(s)z^s-y(0)-y(1)z-...-y(k-1)z^{k-1}\right)=$$

$$=\frac{Y(z)}{z^k}-\frac{y(0)}{z^k}-\frac{y(1)}{z^{k-1}}-...-\frac{y(k-1)}{z}.$$

Эта формула является основной для решения уравнений в конечных разностях.

3. $y(i-k)\Leftrightarrow z^kY(z)$.

Заметим, что, так как $y(i)$ определена лишь для $i\geq 0$, то $y(i-k)$ определена лишь для $i\geq k$. Поэтому

$$y(i-k)\Leftrightarrow\sum_{i=k}^{\infty}y(i-k)z^i=z^k\sum_{i=k}^{\infty}y(i-k)z^{i-k}=z^kY(z).$$

Эта формула находит свое применение при нахождении $y(i)$ по $Y(z)$.

4. $\sum_{s=0}^{i}y(s)\lambda^{i-s}\Leftrightarrow\frac{Y(z)}{1-\lambda z}$.

Имеем

$$\sum_{s=0}^{i}y(s)\ \lambda^{i-s}\Leftrightarrow\sum_{i=0}^{\infty}z^i\sum_{s=0}^{i}y(s)\ \lambda^{i-s}=\sum_{s=0}^{\infty}y(s)\lambda^{-s}\sum_{i=s}^{\infty}z^i\lambda^i=$$

$$=\sum_{s=0}^{\infty}y(s)\lambda^{-s}\frac{z^s\lambda^s}{1-\lambda z}=\frac{1}{1-\lambda z}\sum_{s=0}^{\infty}y(s)z^s.$$

Эта формула также находит применение при нахождении $y(i)$ по $Y(z)$.

Заметим, что комбинация свойств 3 и 4 приводит к соотношению

$$\frac{z^kY(z)}{1-\lambda z}\Leftrightarrow\sum_{s=0}^{i-k}y(s)\lambda^{i-k-s}\ \text{при}\ i\geq k.$$

При этом правая часть считается равной 0 при $i<k$.

5. $\lambda^ky(k)\Leftrightarrow Y(\lambda\ \ z)$.

$$\sum_{k=0}^{\infty}\lambda^ky(k)\ \ z^k\Leftrightarrow\sum_{k=0}^{\aleph}y(k)\ \ (\lambda z)^k=Y(\lambda\ \ z).$$

6. $\sum_{s=0}^{k}g(s)y(k-s)\ =\sum_{s=0}^{k}y(s)g(k-s)\ \Leftrightarrow Y(z)G(z)$.

7. $z^r Y^{(r)}(z) \Leftrightarrow k(k-1)(k-2)...(k-r+1)y(k)$.

Примеры *Z*-преобразования

$y(i)$	$Y(z)$
$y(i)=1$	$Y(z)=\dfrac{1}{1-z}$
$y(i)=\begin{cases}0, & i\neq j\\ 1, & i=j\end{cases}$	$Y(z)=z^j$
$y(i)=\alpha^i$	$Y(z)=\sum\limits_{i=0}^{\infty} z^j\alpha^i=\dfrac{1}{1-\alpha z}$
$y(i)=i$	$Y(z)=\sum\limits_{i=0}^{\infty} i\ \ z^i=z\left(\sum\limits_{i=0}^{\infty} z^i\right)'_z=z\left(\dfrac{1}{1-z}\right)'=\dfrac{z}{(1-z)^2}$
$y(i)=i^2$	$Y(z)=\sum\limits_{i=0}^{\infty} i^2 z^i=\sum\limits_{i=2}^{\infty} i(i-1)z^i+\sum\limits_{i=0}^{\infty} i\ \ z^i=z^2\left(\sum\limits_{i=0}^{\infty} z^i\right)''_z+\dfrac{z}{(1-z)^2}=$ $=z^2\left(\dfrac{1}{1-z}\right)''+\dfrac{z}{(1-z)^2}=\dfrac{2z^2}{(1-z)^3}+\dfrac{z}{(1-z)^2}$

Возвратимся теперь к линейным уравнениям в конечных разностях

$$a_0 y(i+n)+a_1 y(i+n-1)+...+a_n y(i)=q(i).$$

Применяя к обеим частям этого равенства *Z*-преобразование и используя свойства 1 и 2, получим

$$a_0\left(\frac{Y(z)}{z^n}-\frac{y(0)}{z^n}-\frac{y(1)}{z^{n-1}}-...-\frac{y(n-1)}{z}\right)+$$

$$+a_1\left(\frac{Y(z)}{z^{n-1}}-\frac{y(0)}{z^{n-1}}-\frac{y(1)}{z^{n-2}}-...-\frac{y(n-2)}{z}\right)+...++a_nY(z)=Q(z).$$

Запишем это равенство в виде

$$P\left(\frac{1}{z}\right)Y(z)=Q(z)+P_0\left(\frac{1}{z}\right)y(0)+P_1\left(\frac{1}{z}\right)y(1)+...+P_{n-1}\left(\frac{1}{z}\right)y(n-1),$$

где $P(u)$ – характеристический полином, а $P_k(u)=a_0u^{n-k}+a_1u^{n-k-1}+$ $+...+a_{n-k-1}u$.

Отсюда легко находится $Y(z)$:

$$Y(z)=\frac{Q(z)}{P\left(\frac{1}{z}\right)}+y(0)\frac{P_0\left(\frac{1}{z}\right)}{P\left(\frac{1}{z}\right)}+y(1)\frac{P_1\left(\frac{1}{z}\right)}{P\left(\frac{1}{z}\right)}+\ldots+y(n-1)\frac{P_{n-1}\left(\frac{1}{z}\right)}{P\left(\frac{1}{z}\right)}. \qquad (1.12)$$

Очевидно, что первое слагаемое здесь соответствует частному решению неоднородного уравнения при начальных условиях вида $y(0)=y(1)=\ldots=y(n-1)=0$, а последние слагаемые дают решение однородного уравнения.

В принципе, по $Y(z)$ можно найти $y(i)$.

Рассмотрим два частных случая.

А. Уравнение в конечных разностях первого порядка.

$$a_0y(i+1)+a_1y(i)=q(i). \qquad (1.13)$$

Переходя к $Y(z)$, получим

$$\frac{a_0}{z}\left(Y(z)-y(0)\right)+a_1Y(z)=Q(z),$$

откуда

$$Y(z)=\frac{zQ(z)}{a_0+a_1z}+y(0)\frac{a_0}{a_0+a_1z}. \qquad (1.14)$$

Пусть $z_1=-a_1/a_0$ есть корень характеристического уравнения $a_0z+a_1=0$, тогда (1.14) примет вид

$$Y(z)=\frac{1}{a_0}\frac{zQ(z)}{1-z_1z}+y(0)\frac{a_0}{1-z_1z}.$$

Пользуясь свойствами 3 и 4, запишем решение $y(i)$ уравнения (1.13) в виде

$$y(i)=\frac{1}{a_0}\sum_{s=0}^{i-1}q(s)z_1^{i-1-s}+y(0)z_1^i, \qquad (1.15)$$

причем первое слагаемое считается равным нулю при $i=0$.

Б. Уравнение в конечных разностях второго порядка

$$a_0y(i+2)+a_1y(i+1)+a_2y(i)=q(i). \qquad (1.16)$$

Переходя к Z-преобразованию, получим

$$\left(\frac{a_0}{z^2}+\frac{a_1}{z}+a_2\right)Y(z)=Q(z)+y(0)\left(\frac{a_0}{z^2}+\frac{a_1}{z}\right)+y(1)\frac{a_0}{z}. \qquad (1.17)$$

Пусть z_1 и z_2 есть корни характеристического уравнения $a_0z^2+a_1z+a_2=0$. Пусть оба этих корня вещественны и различны. Тогда (1.17) можно представить в виде

$$Y(z)=\frac{1}{a_0}\frac{z^2Q(z)}{(1-z_1z)(1-z_2z)}+y(0)\frac{a_0+a_1z}{a_0(1-z_1z)(1-z_2z)}+y(1)\frac{z}{a_0(1-z_1z)(1-z_2z)}.$$

Воспользовавшись разложением

$$\frac{z}{(1-z_1z)(1-z_2z)}=\frac{1}{(z_1-z_2)}\left[\frac{1}{1-z_1z}-\frac{1}{1-z_2z}\right]=$$

$$=\frac{1}{(z_1-z_2)}\frac{1-z_2z-1+z_1z}{(1-z_1z)(1-z_2z)}=\frac{z}{(1-z_1z)(1-z_2z)},$$

получим

$$Y(z)=\frac{1}{a_0(z_1-z_2)}\left[\frac{zQ(z)}{1-z_1z}-\frac{zQ(z)}{1-z_2z}\right]+\text{решение однородного уравнения}.$$

$$y(i)=\frac{1}{a_0(z_1-z_2)}\sum_{s=0}^{i-1}q(i)\left(z_1^{i-1-s}-z_2^{i-1-s}\right)+\text{решение однородного уравнения}.$$

Тогда

$$y(i)=\frac{1}{a_0(z_1-z_2)}\sum_{s=0}^{i-1}q(i)\left(z_1^{i-1-s}-z_2^{i-1-s}\right)+\text{решение однородного уравнения}.$$

Выписанное слагаемое дает частное решение неоднородного уравнения. Оно равно 0 при $i=0$ и $i=1$.

Если z_1 и z_2 – комплексно сопряженные корни: $z_{1,2}=r[\cos\omega\pm j\sin\omega]$, то

$$y(i)=\frac{1}{a_0\sin\omega}\sum_{s=0}^{i-1}q(i)r^{i-1-s}\sin(i-1-s)\omega+y(0)\frac{z_2z_1^i-z_1z_2^i}{z_2-z_1}+y(1)\frac{z_2^i-z_1^i}{z_2-z_1}.$$

§ 1.6. Метод вариации произвольных постоянных

Этот метод является самым сильным методом, но и самым сложным в практическом выполнении.

Пусть имеется неоднородное уравнение

$$\sum_{l=0}^{n} a_{n-l} y(i+l) = q(i). \tag{1.18}$$

Общее решение однородного уравнения имеет вид $y(i) = \sum_{s=1}^{n} C_s y_s(i)$, где $y_s(i)$ – линейно независимые решения линейного однородного уравнения. Будем искать частное решение неоднородного уравнения в виде

$$y(i) = \sum_{s=1}^{n} C_s(i) y_s(i). \tag{1.19}$$

Подставляя это решение в уравнение (1.18), получим

$$\sum_{l=0}^{n} a_{n-l} \sum_{s=1}^{n} C_s(i+l) y_s(i+l) = q(i). \tag{1.20}$$

Нам нужно всего одно частное решение уравнения (1.18), а в нашем распоряжении n функций $C_1(i)$, $C_2(i)$,..., $C_n(i)$. Поэтому на них можно произвольно наложить $(n-1)$ ограничение. Выберем эти ограничения в виде

$$\sum_s C_s(i+1) y_s(i+1) = \sum_s C_s(i) y_s(i+1),$$

$$\sum_s C_s(i+2) y_s(i+2) = \sum_s C_s(i) y_s(i+2),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_s C_s(i+n-1) y_s(i+n-1) = \sum_s C_s(i) y_s(i+n-1). \tag{1.21}$$

С учетом этих ограничений уравнение (1.20) приобретет вид

$$\begin{gathered} a_0 \sum_s C_s(i+n) y_n(i+n) + \\ + \sum_s C_s(i) \left[a_1 y_s(i+n-1) + a_2 y_s(i+n-2) + \ldots + a_n y_s(i) \right] = q(i). \end{gathered} \tag{1.22}$$

С другой стороны, вспоминая, что $y_s(i)$ удовлетворяют однородному уравнению

$$a_0 y_s(i+n) + a_1 y_s(i+n-1) + \ldots + a_n y_s(i) = 0,$$

умножая это соотношение на $C_s(i)$ и складывая по s, получим

$$a_0 \sum_s C_s(i) y_s(i+n) + \sum_s C_s(i)\left[a_1 y_s(i+n-1) + a_2 y_s(i+n-2) + a_n y_s(i)\right] = 0. \tag{1.23}$$

Вычитая из (1.22) равенство (1.23), получим

$$a_0 \sum_s C_s(i+n) y_s(i+n) = a_0 \sum_s C_s(i) y_n(i+n) + q(i). \tag{1.24}$$

Преобразуем уравнения (1.21), (1.24). Возьмем первое уравнение (1.21) и заменим в нем i на i+1. Тогда получим

$$\sum_s C_s(i+2) y_s(i+2) = \sum_s C_s(i+1) y_s(i+2).$$

Сравнивая это соотношение со вторым условием системы (1.21), запишем это второе условие в форме

$$\sum_s C_s(i+1) y_s(i+2) = \sum_s C_s(i) y_s(i+2).$$

Аналогично, заменяя во втором условии (1.21) i на i+1

$$\sum_s C_s(i+3) y_s(i+3) = \sum_s C_s(i+1) y_s(i+3)$$

и сравнивая его с третьим, получим

$$\sum_s C_s(i+1) y_s(i+3) = \sum_s C_s(i) y_s(i+3).$$

Аналогично преобразуем и остальные равенства системы (1.21). В последнем из них заменим i на i+1

$$\sum_s C_s(i+n) y_s(i+n) = \sum_s C_s(i+1) y_s(i+n)$$

и, сравнивая с (1.24), получим

$$\sum_s C_s(i+1) y_s(i+n) = \sum_s C_s(i) y_s(i+n) + \frac{1}{a_0} q(i).$$

Таким образом, новая система относительно функций $C_s(i)$ примет вид

$$\sum_s \left[C_s(i+1) - C_s(i)\right] y_s(i+1) = 0,$$

$$\sum_s [C_s(i+1)-C_s(i)]y_s(i+2)=0,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (1.25)$$

$$\sum_s [C_s(i+1)-C_s(i)]y_s(i+n-1)=0,$$

$$\sum_s [C_s(i+1)-C_s(i)]y_s(i+n)=\frac{1}{a_0}q(i).$$

Обозначим

$$D[i]=\begin{vmatrix} y_1(i+1) & y_2(i+1) & \ldots & y_n(i+1) \\ y_1(i+2) & y_2(i+2) & \ldots & y_n(i+2) \\ \ldots & \ldots & \ldots & \ldots \\ y_1(i+n) & y_2(i+n) & \ldots & y_n(i+n) \end{vmatrix}. \qquad (1.26)$$

Этот определитель отличен от нуля в силу линейной независимости решений линейного однородного уравнения $y_s(i)$. Через $D_k[i]$ обозначим определитель, который получается из $D[i]$ вычеркиванием k-го столбца и последней строки. Тогда, рассматривая (1.25) как систему линейных уравнений, по правилу Крамера сможем записать

$$C_s(i+1)-C_s(i)=(-1)^{n+s}\frac{1}{a_0}q(i)\frac{D_s(i)}{D(i)}.$$

Отсюда легко получить, считая, что $C_s(0)=0$,

$$C_s(i)=\sum_{r=0}^{i-1}(-1)^{n+s}\frac{1}{a_0}q(r)\frac{D_s(r)}{D(r)},$$

что и определяет в явном виде $C_s(i)$, а следовательно, и частное решение $y(i)=\sum_s C_s(i)y_s(i)$.

В качестве примера рассмотрим уравнение второго порядка

$$a_0y(i+2)+a_1y(i+1)+a_2y(i)=q(i),$$

когда характеристическое уравнение $a_0z^2+a_1z+a_2=0$ имеет корень z кратности 2. В этом случае $y_1(i)=z^i$, $y_2(i)=iz^i$, и поэтому

$$D[i]=\begin{vmatrix} z^{i+1} & (i+1)z^{i+1} \\ z^{i+2} & (i+2)z^{i+2} \end{vmatrix}=z^{2i+3}.$$

Далее, $D_1(i) = (i+1)z^{i+1}$, $D_2(i) = z^{i+1}$, и поэтому

$$C_1(i) = -\frac{1}{a_0}\sum_{r=0}^{i-1} q(r)\frac{r+1}{z^{r+2}},$$

$$C_2(i) = \frac{1}{a_0}\sum_{r=0}^{i-1} q(r)\frac{1}{z^{r+2}},$$

тогда частное решение неоднородного уравнения имеет вид

$$y(i) = C_1(i)z^i + C_2 iz^i = \frac{1}{a_0}\sum_{r=0}^{i-1} q(r)(i-1-r)z^{i-r-2}. \tag{1.27}$$

Отметим, что в этом частном решении $y(0) = y(1) = 0$.

§ 1.7. Линейные уравнения в конечных разностях с переменными коэффициентами

Всюду выше рассматривались линейные уравнения в конечных разностях с постоянными коэффициентами. Практика решения аналогичных уравнений с коэффициентами, зависящими от i, гораздо сложнее.

Сформулируем без доказательства две теоремы, полезные для анализа вида решения уравнений с переменными коэффициентами.

Теорема Пуанкаре. Если для линейного однородного уравнения в конечных разностях

$$a_0(i)y(i+n) + a_1(i)y(i+n-1) + \ldots + a_{n-1}(i)y(i+1) + a_n(i)y(i) = 0 \tag{1.28}$$

существуют конечные пределы для переменных коэффициентов $a_k(i)$, когда i неограниченно возрастает

$$\lim_{i\to\infty} a_k(i) = \alpha_k, \quad k = 0,\ 1,\ 2,\ \ldots,\ n,$$

и если корни уравнения

$$\alpha_0 z^n + \alpha_1 z^{n-1} + \ldots + \alpha_{n-1}z + \alpha_n = 0 \tag{1.29}$$

все различны по модулю, то каково бы ни было частное решение $y(i)$ уравнения (1.28), предел отношения $y(i+1)$ к $y(i)$, когда i неограниченно возрастает, равен одному из корней z_1, z_2, ..., z_n уравнения (1.29), то есть

$$\lim_{i\to\infty}\frac{y(i+1)}{y(i)} = z_p.$$

Теорема Перрона. Если для уравнения (1.28) условия теоремы Пуанкаре выполнены и, кроме того, $a_0(i)\neq 0$, $a_n(i)\neq 0$ для всех $i=0,\ 1,\ 2,\ \ldots$, то существует n решений $y_1(i)$, $y_2(i),\ldots,\ y_n(i)$ этого уравнения, для каждого из которых

$$\lim_{i\to\infty}\frac{y_k(i+1)}{y_k(i)}=z_k,\quad k=1,\ 2,\ \ldots,\ n.$$

Общее решение уравнения (1.28) имеет вид

$$y(i)=C_1z_1^i(1+\xi_1(i))^i+C_2z_2^i(1+\xi_2(i))^i+\ldots+C_nz_n^i(1+\xi_{n1}(i))^i, \quad (1.30)$$

где все $\xi_k(i)\underset{i\to\infty}{\to}0$.

Выражение (1.30) показывает поведение решения линейного однородного уравнения с переменными коэффициентами.

Теперь рассмотрим два частных случая линейных уравнений в конечных разностях с переменными коэффициентами.

А. Уравнение первого порядка

Линейное разностное уравнение первого порядка с переменными коэффициентами всегда можно привести к виду

$$y(i+1)-y(i)p(i)=q(i). \quad (1.31)$$

Рассмотрим сначала решение соответствующего однородного уравнения

$$y(i+1)-y(i)p(i)=0. \quad (1.32)$$

Полагая $i=0,\ 1,\ 2,\ 3,\ \ldots$, получим

$$y(1)=y(0)p(0),$$

$$y(2)=y(1)p(1)=y(0)p(0)p(1),$$

$$y(3)=y(2)p(2)=y(0)p(0)p(1)p(2).$$

Легко получить, что для произвольного i в этом случае

$$y(i)=y(0)\prod_{s=0}^{i-1}p(s)\text{, причем }\prod_{s=1}^{i-1}p(s)\Big|_{i=0}\equiv 1. \quad (1.33)$$

Решение неоднородного уравнения (1.31) будем искать в виде

$$y(i)=C(i)y(0)\prod_{s=0}^{i-1}p(s). \quad (1.34)$$

Тогда

$$y(i+1) = C(i+1)y(0)\prod_{s=0}^{i} p(s),$$

и подстановка в (1.31) даст

$$C(i+1)y(0)\prod_{s=0}^{i} p(s) - p(i)C(i)y(0)\prod_{s=0}^{i-1} p(s) = q(i),$$

откуда

$$C(i+1) - C(i) = \frac{q(i)}{y(0)\prod\limits_{s=0}^{i} p(s)}.$$

Легко видеть, что

$$C(i) = \sum_{r=0}^{i-1} \frac{q(r)}{y(0)\prod\limits_{s=0}^{r} p(s)} + C(0).$$

При $i = 0$ всю эту сумму следует считать равной нулю. Подставляя это выражение в (1.34), получим

$$y(i) = \sum_{r=0}^{i-1} q(r) \prod_{s=r+1}^{i-1} p(s) + C(0)y(0)\prod_{s=0}^{i-1} p(s),$$

откуда видно, что для выполнения условия $y(0) = y(0)$ следует взять $C(0) = 1$. Окончательно

$$y(i) = y(0)\prod_{s=0}^{i-1} p(s) + \sum_{r=0}^{i-1} q(r) \prod_{s=r+1}^{i-1} p(s). \qquad (1.35)$$

Б. Уравнение второго порядка

Линейное разностное уравнение второго порядка с переменными коэффициентами имеет вид

$$y(i+2) + p_1(i)y(i+1) + p_2(i)y(i) = q(i). \qquad (1.36)$$

Соответствующее однородное уравнение имеет вид

$$y(i+2) + p_1(i)y(i+1) + p_2(i)y(i) = 0. \qquad (1.37)$$

Общее решение этого уравнения неизвестно. Предположим, что мы каким-то образом нашли частное решение этого уравнения $y_1(i)$. Покажем, как можно найти другое частное решение $y_2(i)$, линейно независимое с $y_1(i)$.

Рассмотрим определитель

$$\Delta(i)=\begin{vmatrix} y_1(i) & y_2(i) \\ y_1(i+1) & y_2(i+1)\end{vmatrix}=y_1(i)y_2(i+1)-y_1(i+1)y_2(i). \qquad (1.38)$$

Вычисляем $\Delta(i+1)$

$$\Delta(i+1)=y_1(i+1)y_2(i+2)-y_1(i+2)y_2(i+1),$$

и, учитывая, что $y_1(i)$ и $y_2(i)$ удовлетворяют уравнению (1.37), получим

$$\Delta(i+1)=y_1(i+1)\left[-p_1(i)y_2(i+1)-p_2(i)y_2(i)\right]-$$

$$-y_2(i+1)\left[-p_1(i)y_1(i+1)-p_2(i)y_1(i)\right]=$$

$$=p_2(i)\left[y_2(i+1)y_1(i)-y_1(i+1)y_2(i)\right]=p_2(i)\Delta(i),$$

то есть $\Delta(i)$ удовлетворяют уравнению

$$\Delta(i+1)=p_2(i)\Delta(i).$$

Отсюда

$$\Delta(i)=\Delta(0)\prod_{s=0}^{i-1}p_2(s),$$

и равенство (1.38) дает

$$y_2(i+1)y_1(i)-y_2(i)y_1(i+1)=\Delta(0)\prod_{s=0}^{i-1}p_2(s).$$

Отсюда видно, что $y_2(i)$ удовлетворяет линейному уравнению в конечных разностях уже первого порядка

$$y_2(i+1)=y_2(i)\frac{y_1(i+1)}{y_1(i)}+\frac{\Delta(0)}{y_1(i)}\prod_{s=0}^{i-1}p_2(s), \qquad (1.39)$$

решение которого было разобрано выше. Зная $y_2(i)$, можно найти частное решение неоднородного уравнения (1.36) методом вариации произвольных постоянных.

§ 1.8. Системы уравнений в конечных разностях

Пусть теперь имеется n функций $y_1(i)$, $y_2(i)$, ..., $y_n(i)$ дискретного аргумента i, $i=\overline{0,\infty}$. Каноническая форма системы уравнений в конечных разностях имеет вид

$$y_s(i+1)=f_s(i,\ y_1(i),\ y_2(i),\ ...,y_n(i)),\quad s=\overline{1,n}. \qquad (1.40)$$

Заметим, что если заданы $y_s(0)$, $s=\overline{0,n}$, то система однозначно определит все $y_s(1)$, $s=\overline{1,n}$. Зная $y_s(1)$, можно найти все $y_s(2)$, $s=\overline{1,n}$, затем все $y_s(3)$ и т.д. Таким образом, решение системы (1.40) зависит от n произвольных констант

$$y_s(i)=\varphi_s(i,\ C_1,\ C_2,\ ...,C_n),\quad s=\overline{0,n}.$$

Линейная неоднородная система уравнений в конечных разностях имеет вид

$$y_s(i+1)=\sum_{k=1}^{n}a_{sk}y_k(i)+q_s(i). \tag{1.41}$$

Вводя матричные обозначения

$$y(i)=\begin{bmatrix} y_1(i) \\ y_2(i) \\ \ldots \\ y_n(i) \end{bmatrix},\qquad q(i)=\begin{bmatrix} q_1(i) \\ q_2(i) \\ \ldots \\ q_n(i) \end{bmatrix},\qquad A=\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix},$$

систему (1.41) можно записать в виде

$$y(i+1)=Ay(i)+q(i). \tag{1.42}$$

Если $q(i)=0$, то система (1.42) называется однородной.

Рассмотрим сначала однородную систему. Будем искать решение в виде $y_s(i)=B_sz^i$. Подставляя это решение в однородную систему, получим

$$B_sz^{i+1}=\sum_{k=1}^{n}a_{sk}B_kz^i.$$

Сокращая на z^i, получим

$$\sum_{k=1}^{n}\left(a_{sk}-z\delta_{sk}\right)B_k=0. \tag{1.43}$$

Чтобы эта система имела ненулевое решение, необходимо, чтобы

$$\det(A-zE)=0,$$

откуда видно, что z есть собственные числа матрицы A.

Рассмотрим лишь случай, когда все n собственных чисел матрицы A различны. Тогда, как это видно из (1.43), B есть не что иное, как собственный вектор матрицы A, соответствующий собственному числу z.

Таким образом, общее решение однородной системы, соответствующей системе (1.42), можно записать в виде

$$y(i) = C_1 B_1 z_1^i + C_2 B_2 z_2^i + \ldots + C_n B_n z_n^i, \tag{1.44}$$

где z_s – собственные числа; B_s – собственные векторы матрицы A, а C_s – произвольные константы, значения которых находятся из начальных условий $y_s(0)$, $s = \overline{0,n}$.

Что касается решений неоднородной системы (1.42), то ее проще всего искать методом производящих функций.

Вводя векторные функции

$$Y(z) = \sum_{i=0}^{\infty} z^i y(i), \quad Q(z) = \sum_{i=0}^{\infty} z^i q(i), \tag{1.45}$$

из системы (1.42) получим

$$\frac{1}{z}\left(Y(z) - y(0)\right) = AY(z) + Q(z),$$

откуда

$$Y(z) = \left(\frac{1}{z}E - A\right)^{-1}\left(Q(z) + \frac{1}{z}y(0)\right). \tag{1.46}$$

§ 1.9. Преобразование Лапласа – Стилтьеса

Обычное преобразование Лапласа достаточно подробно изучалось в курсе математического анализа. В теории массового обслуживания чаще применяется вариант этого преобразования, известный как преобразование Лапласа – Стилтьеса.

Пусть функция $F(x)$ обладает следующими свойствами:

1. $F(x) \equiv 0$ при $x < 0$,

2. $\exists$ M и $s_0 > 0$ такие, что $|F(x)| \le Me^{s_0 x}$.

Тогда преобразованием Лапласа – Стилтьеса $F^*(s)$ от функции $F(x)$ называется функция

$$F^*(s) = \int_0^{\infty} e^{-sx} dF(x). \tag{1.47}$$

Заметим, что $F^*(s)$ есть аналитическая функция в полуплоскости $\operatorname{Re} s > s_0$ и однозначно определяет $F(x)$ во всех точках ее непрерывности:

$$F(x) = \lim_{s\to\infty} \sum_{k\le sx} \frac{(-s)^k}{k!} F^{*(k)}(s), \tag{1.48}$$

где $F^{*(k)}(s)$ есть производная k-го порядка от $F^*(s)$. Эта формула носит название формулы обращения

Таким образом, $F(x)$ и $F^*(s)$ связаны друг с другом взаимнооднозначно, исключая значения $F(x)$ в ее точках разрыва: $F(x) \Leftrightarrow F^*(s)$.

Приведем некоторые важнейшие свойства преобразования Лапласа – Стилтьеса.

1. Линейность. Если $F_1(x) \Leftrightarrow F_1^*(s)$, а $F_2(x) \Leftrightarrow F_2^*(s)$, то $C_1F_1(x) + C_2F_2(x) \Leftrightarrow C_1F_1^*(s) + C_2F_2^*(s)$.

2. $\int\limits_0^x F(y)dy \Leftrightarrow \frac{F^*(s)}{s} + \frac{1}{s}F(0)$.

$$\int\limits_0^\infty e^{-sx} d\int\limits_0^x F(y)dy = \int\limits_0^\infty e^{-sx}F(x)dx = -\frac{1}{s}\int\limits_0^\infty F(x)de^{-sx} =$$

$$= -\frac{1}{s}F(x)e^{-sx}\Big|_0^\infty + \frac{1}{s}\int\limits_0^\infty e^{-sx}dF(x) = \frac{1}{s}F(0) + \frac{1}{s}F^*(s).$$

3. $\int\limits_0^x Q(x-y)dF(y) \Leftrightarrow Q^*(s)F^*(s)$.

$$\int\limits_0^\infty e^{-sx} d\int\limits_0^x Q(x-y)dF(y) = \lim_{A\to\infty}\int\limits_0^A e^{-sx} d\int\limits_0^x Q(x-y)dF(y) =$$

$$= \lim_{A\to\infty}\left\{ e^{-sx}\int\limits_0^x Q(x-y)dF(y)\Big|_0^A - \int\limits_0^A\int\limits_0^x Q(x-y)dF(y)de^{-sx} \right\} =$$

$$= \lim_{A\to\infty}\left\{ e^{-sA}\int\limits_0^A Q(x-y)dF(y) + s\int\limits_0^A e^{-sx}\left(\int\limits_0^x Q(x-y)dF(y)\right)dx \right\} =$$

$$= \lim_{A\to\infty}\left\{s\int_0^A e^{-sx}\int_0^x Q(x-y)dF(y)dx\right\} = \lim_{A\to\infty}\left\{s\int_0^A\int_y^A e^{-sx}Q(x-y)dxdF(y)\right\} =$$

$$= \lim_{A\to\infty}\left\{s\int_0^A\int_0^{A-y} e^{-s(y+z)}Q(z)dzdF(y)\right\} = s\int_0^\infty e^{-sy}\int_0^\infty e^{-sz}Q(z)dzdF(y) =$$

$$= \int_0^\infty e^{-sy}dF(y)s\int_0^\infty e^{-sz}Q(z)dz = \int_0^\infty e^{-sy}dF(y)\int_0^\infty e^{-sz}dQ(z) \Leftrightarrow F^*(s)Q^*(s).$$

4. Обычно $F(x)$ имеет смысл функции распределения некоторой случайной величины ξ. Тогда

$$F^{*(k)}(s) = (-1)^k\int_0^\infty x^k e^{-sx}dF(x), \qquad F^{*(k)}(0) = (-1)^k\int_0^\infty x^k dF(x).$$

В частности,

$$M\{x\} = -F^{*\prime}(0), \qquad M\{x^2\} = F^{*\prime\prime}(0),$$

то есть начальные моменты величины ξ находятся простым дифференцированием функции $F^*(s)$.

5. Предельное поведение

$$\lim_{s\to 0} F^*(s) = F(\infty) - F(0),$$

$$\lim_{s\to\infty} F^*(s) = F^*(+0) - F(0).$$

6. При $F(0) = 0$ $\lim\limits_{s\to 0}\dfrac{F^*(s)}{s} = \int_0^\infty F(x)dx$.

Действительно, по свойству 2, $\int_0^x F(y)dy \Leftrightarrow \dfrac{F^*(s)}{s}$. Применяя предельные соотношения, получим

$$\lim_{s\to 0}\frac{F^*(s)}{s} = \int_0^\infty F(x)dx - \int_0^0 F(x)dx = \int_0^\infty F(x)dx.$$

7. $\lim\limits_{s\to 0} s\int_0^\infty e^{-sx}f(x)dx = \lim\limits_{x\to\infty} f(x)$.

Действительно,

$$s\int_0^\infty e^{-sx} f(x)dx = -\int_0^\infty f(x)de^{-sx} = -f(x)e^{-sx}\Big|_0^\infty + \int_0^\infty e^{-sx}df(x) = f(0) + \int_0^\infty e^{-sx}df(x).$$

Переходя к пределу, получим

$$\lim_{s\to 0} s\int_0^\infty e^{-sx} f(x)dx = \lim_{s\to 0}\left[f(0) + \int_0^\infty e^{-sx}df(x)\right] = f(0) + \int_0^\infty df(x) =$$

$$= f(0) + f(\infty) - f(0) = f(\infty) = \lim_{x\to\infty} f(x).$$

8. $\lim\limits_{s\to 0} sF^*(s) = \lim\limits_{x\to\infty} \dfrac{F(x)}{x}$.

$$\lim_{x\to\infty}\frac{F(x)}{x} = \lim_{s\to 0} s\int_0^\infty e^{-sx}\frac{F(x)}{x}dx = \lim_{s\to 0}\frac{\int_0^\infty e^{-sx}\frac{F(x)}{x}dx}{1/s} = \lim_{s\to 0}\frac{\int_0^\infty e^{-sx}F(x)dx}{1/s^2} =$$

$$= \lim_{s\to 0} s^2\int_0^\infty e^{-sx}F(x)dx = \lim_{s\to 0}\left[-s\int_0^\infty F(x)de^{-sx}\right] =$$

$$= \lim_{s\to 0}\left[-sF(x)e^{-sx}\Big|_0^\infty + s\int_0^\infty e^{-sx}dF(x)\right] =$$

$$\lim_{s\to 0}\left[sF(0) + s\int_0^\infty e^{-sx}dF(x)\right] = \lim_{s\to 0} sF^*(s).$$

Глава 2. ТЕОРИЯ ПОТОКОВ СОБЫТИЙ

§ 2.1. Определения и терминология

Представим себе ось времени, на которой в какие-то случайные моменты t_1, t_2, t_3 и т.д. наступают события. Такой случайный процесс называется случайным потоком однородных событий или случайным точечным процессом.

Описать такой процесс можно тремя способами:

1) задать статистические свойства последовательности моментов наступления событий t_1, t_2, t_3, ...;

2) задать статистические свойства длины интервалов τ_i между моментами наступления событий;

3) задать статистические свойства случайного процесса $i(t)$, где $i(t)$ – число событий, наступивших за время t.

Очевидно, что от одного описания можно перейти к другому, так как

$$\tau_1 = t_1,\ \tau_2 = t_2 - t_1,\ \dots,\ \tau_i = t_i - t_{i-1},$$

$$t_1 = \tau_1,\ t_2 = \tau_1 + \tau_2,\ t_i = \sum_{k=1}^{i} \tau_k .$$

Обычно предпочитают использовать описание потока через интервалы между событиями либо через $i(t)$ – число событий, наступивших за время t. Говоря о случайных потоках, обычно интересуются наличием или отсутствием следующих свойств:

А. Стационарность

Возьмем на временной оси n непересекающихся интервалов длиной Δ_i, $i = \overline{1,n}$, и пусть η_i – число событий, происшедших на интервале Δ_i. Поток называется стационарным, если вероятности

$$P\{\eta_i = k\},\quad k \ge 0,\quad i = \overline{1,n},$$

не зависят от местонахождения i -го интервала на оси времени, а определяются лишь его длиной. Смысл этого определения состоит в том, что вероятностные свойства потока не зависят от начала отсчета времени.

Б. Последействие

Если η_1, η_2, ..., η_n – независимые случайные величины, то говорят, что поток событий обладает свойством отсутствия последействия. Смысл этого определения заключается в том, что наступление (или не наступление) событий на каком-то интервале не зависит от того, сколько событий появилось на каких-то других интервалах.

В. Ординарность

Обозначим $P_{>1}(t_0, t)$ вероятность появления более одного события на интервале $[t_0, t]$. Если

$$\lim_{t \to t_0} \frac{P_{>1}(t_0, t)}{t - t_0} = 0,$$

то поток называется ординарным. Смысл этого определения заключается в том, что наступление более одного события на бесконечно малом интервале длины Δt есть бесконечно малая величина более высокого порядка, чем Δt. Более образно – на бесконечно малом интервале времени не может наступить более одного события.

Интенсивность и параметр потока

Пусть $P_1(t_0, t)$ есть вероятность того, что на интервале $[t_0, t]$ наступит ровно одно событие.

Величина

$$\lim_{t \to t_0} \frac{P_1(t_0, t)}{t - t_0} = \lambda(t_0)$$

называется **параметром** потока.

Обозначим через $m(t_0, t)$ **среднее** число событий, наступивших на интервале $[t_0, t]$. Величина

$$\lim_{t \to t_0} \frac{m(t_0, t)}{t - t_0} = \mu(t_0)$$

называется **интенсивностью** потока.

Легко сообразить, что для ординарного потока параметр потока и его интенсивность равны между собой. Не очень строго это может быть показано так:

$$m(t_0,\ t)=\sum_{k=0}^{\infty}kP_k(t_0,\ t)=P_1(t_0,\ t)+\sum_{k>2}^{\infty}kP_k(t_0,\ t).$$

В силу ординарности потока вторая сумма есть бесконечно малая величина по сравнению с $t-t_0$ (здесь, конечно, главная нестрогость: ряд бесконечный, при $k\to\infty$, все может быть и не так). Тогда, деля на $t-t_0$ и переходя к пределу $t\to t_0$, мы и получим, что $\lambda(t_0)=\mu(t_0)$.

В дальнейшем, для ординарных потоков мы не будем делать разницы между интенсивностью потока и его параметром. Заметим лишь, что для **неординарных** потоков всегда $\lambda(t)\le\mu(t)$, так как $m(t_0,t)\ge P_1(t_0,t)$.

§ 2.2. Пуассоновский поток событий

Пуассоновский поток событий является одним из самых «любимых» потоков в теории массового обслуживания. Он характеризуется выполнением двух свойств:

1) отсутствием последствия;

2) ординарностью.

Прежде чем выводить формулы, введем обозначения и сделаем некоторые следствия из ординарности. Пусть $P_к(t_0,t)$ есть вероятность того, что на интервале $[t_0,t]$ наступит ровно k событий. Тогда, в силу определения интенсивности, функция

$$\lambda(t)=\lim_{\Delta t\to 0}\frac{P_1(t,\ t+\Delta t)}{\Delta t} \tag{2.1}$$

будет интенсивностью пуассоновского потока. Это же соотношение можно записать как

$$P_1(t,\ t+\Delta t)=\lambda(t)\Delta t+o(\Delta t). \tag{2.2}$$

В силу ординарности потока

$$\lim_{\Delta t\to 0}\frac{P_k(t,\ t+\Delta t)}{\Delta t}=0 \quad \text{при} \quad k\ge 2,$$

и поэтому при $k\ge 2$

$$P_k(t,\ t+\Delta t)=o(\Delta t). \tag{2.3}$$

Так как выполнено условие нормировки

$$P_0(t,\ t+\Delta t)+P_1(t,\ t+\Delta t)+P_{>1}(t,\ t+\Delta t)=1,$$

то

$$P_0(t_1 t+\Delta t)=1-P_1(t,\ t+\Delta t)-P_{>1}(t,\ t+\Delta t)=1-\lambda(t)\Delta t+o(\Delta t). \quad (2.4)$$

Соотношения (2.2) – (2.4) являются основными для дальнейшего.

Выведем систему дифференциальных уравнений, определяющую $P_k(t_0,t)$. Учитывая отсутствие последствия, можно записать

$$P_0(t_0,\ t+\Delta t)=P_0(t_0,\ t)P_0(t,\ t+\Delta t)=P_0(t_0,\ t)\left[1-\lambda(t)\Delta t+o(\Delta t)\right],$$

отсюда

$$\frac{P_0(t_0,\ t+\Delta t)-P_0(t_0,\ t)}{\Delta t}=-P_0(t_0,\ t)\lambda(t)+\frac{o(\Delta t)}{\Delta t}.$$

Переходя к пределу $\Delta t\to 0$, получим

$$\frac{\partial P_0(t_0,t)}{\partial t}=-\lambda(t)P_0(t_0,\ t).$$

Аналогично, для $P_k(t_0,\ t)$ можно записать

$$P_k(t_0,t+\Delta t)=P_k(t_0,t)P_0(t,t+\Delta t)+P_{k-1}(t_0,t)P_1(t,t+\Delta t)+\sum_{s=2}^{k}P_{k-s}(t_0,t)P_s(t,t+\Delta t).$$

Используя соотношения (2.2) – (2.4), запишем

$$P_k(t_0,t+\Delta t)=P_k(t_0,t)\left[1-\lambda(t)\Delta t+o(\Delta t)\right]+P_{k-1}(t_0,t)\left[\lambda(t)\Delta t+o(\Delta t)\right]+o(\Delta t),$$

откуда

$$\frac{P_k(t_0,t+\Delta t)-P_k(t_0,t)}{\Delta t}=P_{k-1}(t_0,t)\lambda(t)-P_k(t_0,t)\lambda(t)+\frac{o(\Delta t)}{\Delta t},$$

и после предельного перехода $\Delta t\to 0$ получим

$$\frac{\partial P_k(t_0,t)}{\partial t}=\lambda(t)\left[P_{k-1}(t_0,t)-P_k(t_0,t)\right].$$

Итак,

$$\frac{\partial P_0(t_0,t)}{\partial t}=-\lambda(t)P_0(t_0,t),$$

$$\frac{\partial P_k(t_0,t)}{\partial t}=\lambda(t)\left[P_{k-1}(t_0,t)-P_k(t_0,t)\right],\quad k\ge 1. \quad (2.5)$$

Для однозначного решения этой системы надо добавить граничное условие, которое естественно брать в виде

$$P_k(t_0,t_0)=\begin{cases}1, & \text{если } k=0,\\ 0, & \text{если } k\geq 1,\end{cases} \tag{2.6}$$

так как в силу ординарности потока на интервале нулевой длины с вероятностью 1 не будет ни одного события.

Рассмотрим решение системы (2.5) методом производящих функций. Введем функцию

$$F(t_0,t,z)=\sum_{k=0}^{\infty}P_k(t_0,t)z^k .$$

Умножая первое уравнение системы (2.5) на z^0, а все остальные – на соответствующие z^k и суммируя все равенства, получим

$$\frac{\partial F(t_0,t,z)}{\partial t}=-\lambda(t)\sum_{k=0}^{\infty}P_k(t_0,t)z^k+$$

$$+\lambda(t)\sum_{k=1}^{\infty}P_{k-1}(t_0,t)z^k=-\lambda(t)F(t_0,t,z)+\lambda(t)zF(t_0,t,z).$$

Поэтому окончательно

$$\frac{\partial F(t_0,t,z)}{\partial t}=\lambda(t)(z-1)F(t_0,t,z) . \tag{2.7}$$

Это уравнение легко решается методом разделения переменных

$$\frac{dF(t_0,t,z)}{F(t_0,t,z)}=(z-1)\lambda(t)dt ,$$

откуда, интегрируя в пределах (t_0,t), получим

$$\ln F(t_0,t,z)-\ln F(t_0,t_0,z)=(z-1)\int_{t_0}^{t}\lambda(u)du .$$

Но, как следует из граничного условия (2.6),

$$F(t_0,t_0,z)=1 ,$$

и поэтому

$$F(t_0,t,z)=\exp\left[(z-1)\Lambda(t_0,t)\right]=e^{z\Lambda(t_0,\,t)}e^{-\Lambda(t_0,\,t)} ,$$

где $\Lambda(t_0,t)=\int_{t_0}^{t}\lambda(u)du$. Разлагая $e^{z\Lambda(t_0,\,t)}$ в ряд Тейлора, запишем

$$F(t_0,t,z)=\sum_{k=0}^{\infty}\frac{z^k\Lambda^k(t_0,t)}{k!}e^{-\Lambda(t_0,t)},$$

откуда получаем основную формулу, относящуюся к пуассоновскому потоку:

$$P_k(t_0,t)=\frac{\Lambda^k(t_0,t)}{k!}e^{-\Lambda(t_0,t)}. \qquad (2.8)$$

Эта формула вместе с формулами (2.2) – (2.4) и позволяет решать практически все задачи, относящееся к пуассоновскому потоку.

Приведем несколько примеров на использование этих формул.

1. Пусть мы рассматриваем интервал $[t_0,\infty)$ и нас интересует момент t_i наступления i-го события. Как найти плотность вероятностей $p_i(t_i)$ величин t_i?

Рассмотрим интервал $[t,\ t+dt]$. Вероятность того, что именно на этом интервале наступит i-е по счету событие, можно записать двумя путями:

$$P(t\le t_i\le t+dt)=p_i(t)dt=\frac{\Lambda^{i-1}(t_0,t)}{(i-1)!}e^{-\Lambda(t_0,t)}\lambda(t)dt,$$

так как на интервале $[t_0,t)$ наступило $i-1$ событие, а на интервале $[t,\ t+dt]$ ровно одно (уже i-е по счету). Сокращая на dt, получим

$$p_i(t)=\frac{\Lambda^{i-1}(t_0,t)}{(i-1)!}e^{-\Lambda(t_0,t)}\lambda(t). \qquad (2.9)$$

2. Рассмотрим интервал $[t_0,\ t_0+T]$ и найдем плотность вероятностей того, что на нем наступили n событий в моменты времени t_1, t_2, …, t_n. Выделяя бесконечно малые интервалы $[t_i,\ t_i+dt_i]$, будем иметь

$$p_n(t_1,t_2,\ldots t_n)dt_1dt_2\ldots dt_n=$$
$$=e^{-\Lambda(t_0,t_1)}\lambda(t_1)dt_1\cdot e^{-\Lambda(t_1,t_2)}\lambda(t_2)dt_2\cdot\ldots\cdot e^{-\Lambda(t_{n-1},t_n)}\lambda(t_n)dt_n\cdot e^{-\lambda(t_n,T)},$$

так как на интервалах $[t_0,t_1]$, $[t_1+dt_1,t_2]$, …, $[t_n+dt_n,T]$ не наступит ни одного события, а на интервалах $[t_i,\ t_i+dt_i]$ – ровно по одному. Сокращая на $dt_1dt_2\ldots dt_n$, получим

$$p_n(t_1,t_2,...,t_n) = e^{-\Lambda(t_0,t_0+T)} \cdot \prod_{i=1}^{n} \lambda(t_i) . \qquad (2.10)$$

3. Пусть по моментам t_i появления событий на интервале $[t_0,T]$ мы составляем статистику

$$\xi = \sum_{i\geq 1} f(t_i) ,$$

где $f(\cdot)$ – некоторая гладкая функция. Найдем $M\{\xi\}$ и $D\{\xi\}$.

А) Вспоминая выражение для $p_i(t)$, получим

$$M\{\xi\} = \sum_{i=1}^{\infty} M\{f(t_i)\} = \sum_{i=1}^{\infty} \int_{t_0}^{T} f(t) \frac{\Lambda^{i-1}(t_0,t)}{(i-1)!} e^{-\Lambda(t_0,t)} \lambda(t)dt =$$

$$= \int_{t_0}^{T} f(t)\lambda(t)e^{-\Lambda(t_0,t)} \sum_{i=1}^{\infty} \frac{\Lambda^{i-1}(t_0,t)}{(i-1)!} dt .$$

Но $\sum_{i=1}^{\infty} \frac{\Lambda^{i-1}(t_0,t)}{(i-1)!} = e^{\Lambda(t_0,t)}$, и поэтому

$$M\{\xi\} = \int_{t_0}^{T} f(t)\lambda(t)dt .$$

Б) Второй путь выглядит так:

$$M\{\xi\} = \sum_{N=1}^{\infty} P\{n=N\} \sum_{i=1}^{N} M\{f(t_i) \mid n=N\} .$$

Так как на интервале $[t_0,t_i)$ наступило $i-1$ событие, а на интервале (t_i,T) – $N-i$ событий, то

$$P\{n=N\}\, p(t_i \mid n=N)dt_i = \frac{\Lambda(t_0,t_i)^{i-1}}{(i-1)!} e^{-\Lambda(t_0,t_i)} \lambda(t_i)dt_i \frac{\Lambda(t_i,T)^{N-i}}{(N-i)!} e^{-\Lambda(t_i,T)} .$$

Поэтому

$$M\{\xi\} = \sum_{N=1}^{\infty} \sum_{i=1}^{N} \int_{t_0}^{T} f(t) \frac{\Lambda(t_0,t)^{i-1}}{(i-1)!} e^{-\Lambda(t_0,t)} \lambda(t) \frac{\Lambda(t,T)^{N-i}}{(N-i)!} e^{-\Lambda(t,T)} dt ,$$

где у переменной интегрирования t_i опущен индекс i. Вспоминая формулу бинома Ньютона, получим

$$\sum_{i=1}^{N} \frac{\Lambda(t_0,t)^{i-1}}{(i-1)!} \frac{\Lambda(t,T)^{(N-1)-(i-1)}}{[(N-1)-(i-1)]\ !} = \frac{[\Lambda(t_0,t)+\Lambda(t,T)]^{N-1}}{(N-1)!} = \frac{\Lambda(t,T)^{N-1}}{(N-1)!},$$

и следовательно,

$$M\{\xi\} = \int_{t_0}^{T} f(t)\lambda(t)e^{-\Lambda(t_0,t)} \sum_{N=1}^{\infty} \frac{\Lambda(t_0,t_0+T)^{N-1}}{(N-1)!} dt = \int_{t_0}^{T} f(t)\lambda(t)dt.$$

Для вычисления дисперсии вычислим $M\{\xi^2\}$. Имеем

$$M\{\xi^2\} = M\left\{\sum_{i\ge1} f(t_i) \sum_{j\ge1} f(t_j)\right\} =$$

$$= M\left\{\sum_{i<j} f(t_i)f(t_j)\right\} + M\left\{\sum_{j<i} f(t_i)f(t_j)\right\} + M\left\{\sum_{i\ge1} f^2(t_i)\right\}.$$

По аналогии с тем, что было при вычислении $M\left\{\sum_{i\ge1} f(t_i)\right\}$, имеем

$$M\left\{\sum_{i\ge1} f^2(t_i)\right\} = \int_{t_0}^{T} f^2(t)\lambda(t)dt.$$

Далее

$$M\left\{\sum_{i<j} f(t_i)f(t_j)\right\} = \sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} \int_{t_0}^{T} dt_i \int_{t_i}^{T} dt_j f(t_i)f(t_j)p(t_i,t_j).$$

Легко видеть, что

$$p(t_i,t_j)dt_i dt_j = \frac{\Lambda^{i-1}(t_0,t_i)}{(i-1)!} e^{-\Lambda(t_0,t_i)}\lambda(t_i)dt_i \frac{\Lambda^{j-i-1}(t_i,t_j)}{(j-i-1)!} e^{-\Lambda(t_i,t_j)}\lambda(t_j)dt_j,$$

так как на интервале $[t_0,t_i)$ произошло $i-1$ событий, а на интервале (t_i,t_j) – $(j-i-1)$ событие. Обозначая в интеграле t_i через u, а t_j через v, получим

$$M\left\{\sum_{i<j} f(t_i)f(t_j)\right\} =$$

$$= \sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} \int_{t_0}^{T} du \int_{u}^{T} dv f(u)f(v)\lambda(u)\lambda(v) \frac{\Lambda^{i-1}(t_0,u)}{(i-1)!} \frac{\Lambda^{j-i-1}(u,v)}{(j-i-1)!} e^{-\Lambda(t_0,v)}.$$

Но

$$\sum_{j=i+1}^{\infty}\frac{\Lambda^{j-i-1}(u,v)}{(j-i-1)!}=e^{\Lambda(u,v)},$$

$$\sum_{i=1}^{\infty}\frac{\Lambda^{i-1}(t_0,u)}{(i-1)!}=e^{\Lambda(t_0,u)},$$

и поэтому

$$M\left\{\sum_{i<j}f(t_i)f(t_j)\right\}=\int_{t_0}^{T}du\int_{u}^{T}dvf(u)f(v)\lambda(u)\lambda(v)=\frac{1}{2}\int_{t_0}^{T}f(u)\lambda(u)du\int_{t_0}^{T}f(v)\lambda(v)dv.$$

Легко получить, что $M\left\{\sum_{j<i}f(t_i)f(t_j)\right\}$ равно этому же выражению. Поэтому

$$M\left\{\xi^2\right\}=\int_{t_0}^{T}f(u)\lambda(u)du\int_{t_0}^{T}f(v)\lambda(v)dv+\int_{t_0}^{T}f^2(u)\lambda(u)du,$$

и окончательно

$$D\{\xi\}=M\left\{\xi^2\right\}-M^2\{\xi\}=\int_{t_0}^{T}f^2(u)\lambda(u)du.$$

§ 2.3. Варианты пуассоновского потока событий

А. Стационарный пуассоновский поток (простейший)

Если $\lambda(t)=\lambda=\text{const}$, то получившийся поток будет стационарным пуассоновским потоком. Для него $\Lambda(t_0,t)=\lambda\cdot(t-t_0)$ и

$$P_k(t_0,t)=\frac{[\lambda(t-t_0)]^k}{k!}e^{-\lambda(t-t_0)}. \tag{2.11}$$

Выведем еще несколько формул, касающихся этого потока.

А.1. Плотность вероятностей для длин интервалов между моментами наступления событий простейшего потока.

Пусть $\tau=t_k-t_{k-1}$ есть длина интервала между моментами наступления событий. Обозначая его плотность вероятностей через $p(\tau)$, можно записать

$$p(\tau)=\lambda e^{-\lambda\tau},\quad \tau\geq 0.$$

А.2. Отсутствие последействия.

Найдем $P\{\tau > x+t \mid \tau > t\}$. Смысл этой вероятности – это вероятность того, что новое событие потока не наступит еще в течение времени x, если известно, что оно уже не наступило в течение времени t.

Имеем

$$P\{\tau > x+t \mid \tau > t\} = \frac{P\{\tau > x+t, \tau > t\}}{P\{\tau > t\}} = \frac{P\{\tau > x+t\}}{P\{\tau > t\}} =$$

$$= \frac{\int_{x+t}^{\infty} \lambda e^{-\lambda\tau} d\tau}{\int_{t}^{\infty} \lambda e^{-\lambda\tau} d\tau} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x}.$$

Здесь использовано то, что событие $\{\tau > x+t, \tau > t\} = \{\tau > t+x\}$.

Таким образом, независимо от того, что событие не наступило в течение времени t, оставшееся до его наступления время распределено по тому же закону, что и исходное время между событиями, независимо от того, какое время t прошло. В этом и проявляется отсутствие последействия в пуассоновском потоке.

А.3. Условная равномерность наступления событий.

Пусть мы имеем временной интервал длительности T. Пусть t_1, t_2, …, t_n есть моменты наступления событий на этом интервале. Тогда формула (2.10) примет вид

$$p_n(t_1, t_2, \dots, t_n) = \lambda^n e^{-\lambda T}.$$

Так как вероятность наступления n событий на интервале $[0, T]$ равна $(\lambda T)^n e^{-\lambda T} / n!$, то условная плотность вероятностей моментов наступления событий t_1, t_2, …, t_n при условии, что их наступило ровно n, равна

$$p_n(t_1, t_2, \dots, t_n \mid n) = n!/T^n, \qquad 0 < t_1 < t_2 < \dots < t_n < T,$$

и она не зависит ни от λ, ни от t_1, t_2, …, t_n. Это говорит о том, что в данном случае t_1, t_2, …, t_n распределены **равномерно** по интервалу $[0, T]$ (с учетом их упорядочивания, которое проявляется в сомножителе $n!$).

Б. Эрланговский поток k-го порядка

Предположим, что в исходном пуассоновском потоке постоянной интенсивности λ после наступления какого-то события следующие $(k-1)$ событие пропадают и появляется только k-е по счету событие. Получающийся поток называется эрланговским потоком k-го порядка.

Пусть $p_k(\tau)$ есть плотность вероятностей длин интервалов между моментами наступления событий в таком потоке. Тогда

$$p_k(\tau)d\tau = \frac{(\lambda\tau)^{k-1}}{(k-1)!}e^{-\lambda\tau}\lambda d\tau,$$

так как на интервале $(0,\tau)$ наступило (хоть и пропали) $k-1$ событие, а на интервале $[\tau,\tau+d\tau]$ ровно одно (k-е по счету). Поэтому

$$p_k(\tau) = \frac{(\lambda\tau)^{k-1}\lambda}{(k-1)!}e^{-\lambda\tau}.$$

В получившейся формуле обычно заменяют λ на $k\cdot\lambda$, чтобы сохранить за λ смысл интенсивности потока, и пишут $p_k(\tau)$ в виде

$$p_k(\tau) = \frac{\lambda^k k^k \tau^{k-1}}{k!}e^{-\lambda k\tau} = \frac{(\lambda k)^k \tau^{k-1}}{k!}e^{-\lambda k\tau}. \quad (2.12)$$

Эрланговский поток можно интерпретировать еще и так: каждое событие, прежде чем появится на свет, должно пройти k фаз, каждая из которых имеет длительность, распределенную с плотностью вероятностей $p(\tau) = (\lambda k)e^{-\lambda k\tau}$, т.е. распределенной по экспоненциальному закону.

Поток называется гиперэрланговским, если $p(\tau)$ имеет вид

$$p(\tau) = \sum_{k=1}^{\infty} C_k \frac{(\lambda k)^k \tau^{k-1}}{k!}e^{-\lambda k\tau}, \quad (2.13)$$

где все $C_k \geq 0$ и $\sum_{k=1}^{\infty} C_k = 1$.

В. Пуассоновский неординарный поток

Рассмотрим поток событий, в котором выполняются свойства стационарности и отсутствия последействия, но не выполняется свойство ординарности.

Пусть события появляются в моменты t_1, t_2, t_3,... «пачками», так что в момент t_k появляются сразу η_k событий. Будем считать, что η_k есть последовательность независимых одинаково распределенных случайных величин с распределением

$$P\{\eta_k = n\} = f_n .$$

Такой поток называется простым (или пуассоновским) неординарным (или групповым) потоком однородных событий.

Найдем распределение вероятностей $P_k(t)$ числа событий потока, наступивших на интервале длиной t. Пусть ν_t – число событий, наступивших на этом интервале, и N_t – число моментов времени, в которые наступали группы событий, то есть

$$N_t = \max\{n : 0 < t_n < t\} .$$

Тогда $\nu_t = \sum_{k=1}^{N_t} \eta_k$. Введем производящую функцию

$$\varphi_t(z) = M\left\{z^{\nu_t}\right\} = \sum_{n=0}^{\infty} P\{N_t = n\} M\left\{z^{\nu_t} \mid N_t = n\right\} =$$

$$= \sum_{n=0}^{\infty} P\{N_t = n\} M\left\{z^{\eta_1+\eta_2+\ldots+\eta_n}\right\} = \sum_{n=0}^{\infty} P\{N_t = n\}\left[M\left\{z^{\eta_i}\right\}\right]^n ,$$

так как все η_i независимы и одинаково распределены. Введем функцию

$$f(z) = M\left\{z^{\eta_i}\right\} = \sum_{k=0}^{\infty} f_k z^k ,$$

имеющую смысл производящей функции числа событий в группе. Тогда

$$\varphi_t(z) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \left[f(z)\right]^n = \exp\{\lambda t[f(z) - 1]\} . \qquad (2.14)$$

Вероятности $P_k(t)$ в принципе можно определить из формул

$$P_k(t) = \frac{1}{k!} \left.\frac{\partial^k \varphi_t(z)}{\partial z^k}\right|_{z=0} , \quad k \ge 1 .$$

Найдем еще среднее число событий, наступивших на интервале $(0,t)$. Имеем

$$M\{\nu_t\} = \sum_{k=0}^{\infty} kP_k(t) = \left.\frac{\partial \varphi_t(z)}{\partial z}\right|_{z=1} =$$

$$= \left.\left\{\lambda t f'(z) \exp\left\{\lambda t \left[f(z)-1\right]\right\}\right\}\right|_{z=1} = \lambda t f'(1) = \lambda t \bar{k},$$

где $\bar{k} = \sum_{k=0}^{\infty} k f_k$. Отсюда интенсивность такого потока равна

$$\frac{M\{\nu_t\}}{t} = \lambda \bar{k},$$

где $\bar{k}$ – среднее число событий в группе. Смысл этой формулы очевиден.

Г. Простой поток разнотипных событий

Пусть $\{t_n\}$ – моменты, образующие пуассоновский поток событий интенсивности λ, а σ_n – независимые одинаково распределенные случайные величины, принимающие значения из множества целых чисел $\{1,\ 2,\ \ldots, k\}$.

Если $\sigma_n = \sigma(t_n) = i$, то в момент t_n происходит событие типа i. Обозначая $p_i = P\{\sigma_n = i\}$, мы получим простой поток разнотипных событий.

Обозначая через $\nu_i(t)$ число событий типа i на интервале $(0,t)$, определим производящую функцию

$$\varphi_t(z_1, z_2, \ldots, z_k) = M\left\{z_1^{\nu_1(t)} z_2^{\nu_2(t)} \ldots z_k^{\nu_k(t)}\right\}.$$

Для вычисления явного вида этой функции введем величины

$$\sigma_{n\,i} = \begin{cases} 1, \text{ если } & \sigma_n = i, \\ 0, \text{ если } & \sigma_n \neq i. \end{cases}$$

Тогда $\nu_i(t) = \sum_n \sigma_{n\,i}$, и мы имеем

$$\varphi_t(z_1, z_2, \ldots, z_k) = \sum_{m=0}^{\infty} P\{N_t = m\} M\left\{\prod_{i=1}^{k} z_i^{\sum_{i=1}^{m} \sigma_{n\,i}}\right\} =$$

$$= \sum_{m=0}^{\infty} P\{N_t = m\}\left\{M\left\{z_1^{\sigma_{1\,1}} z_2^{\sigma_{1\,2}} \ldots z_k^{\sigma_{1\,k}}\right\}\right\}^m,$$

где учтено, что в разные моменты времени величины σ_n независимы и одинаково распределены. Так как $\sigma_{n\,i}$ принимают значения 0 или 1 и значение 1 принимает только одна из них, то

$$M\left\{z_1^{\sigma_{1\,1}} z_2^{\sigma_{1\,2}} \ldots z_k^{\sigma_{1\,k}}\right\} = p_1 z_1 + p_2 z_2 + \ldots + p_k z_k .$$

Поэтому

$$\varphi_t(z_1, z_2, \ldots, z_k) = \sum_{m=0}^{\infty} \frac{(\lambda t)^m}{m!} e^{-\lambda t} \left[p_1 z_1 + p_2 z_2 + \ldots + p_k z_k\right]^m =$$

$$= \exp\left\{\lambda t\left[-1 + (p_1 z_1 + p_2 z_2 + \ldots + p_k z_k)\right]\right\} = \prod_{i=1}^{k} e^{-\lambda\, t\, p_i (1 - z_i)} . \qquad (2.15)$$

Отсюда следует, что $\nu_i(t)$ есть независимые пуассоновские потоки событий с параметрами $\lambda\, p_i$, и мы наблюдаем сумму (суперпозицию) этих потоков.

§ 2.4. Потоки восстановления

Пусть на временной оси некоторый момент времени выбран за 0 и пусть t_1, t_2, $t_3, \ldots, t_i > 0$ – моменты наступления событий (рис. 2.1).

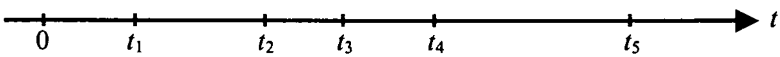


Рис. 2.1. Поток событий

Обозначим $\tau_1 = t_1$, $\tau_n = t_n - t_{n-1}$ – длины интервалов между моментами наступления событий. Отметим принципиальное отличие τ_1 от всех остальных τ_n: если τ_n, $n \geq 2$, отсчитываются от момента наступления предыдущего события потока, то τ_1 отсчитывается от нуля на оси времени, никакого отношения к потоку не имеющего.

Случайный поток называется потоком с **ограниченным последействием**, если $\{\tau_n\}$ – независимые случайные величины.

Смысл слов «ограниченное последействие» следующий: будущее поведение потока после момента времени t, $t_n < t < t_{n+1}$ зависит лишь от конечной величины $(t - t_n)$, а от остального прошлого не зависит.

Для описания потока с ограниченным последействием достаточно задать функции распределения $A_k(x) = P\{\tau_k < x\}$, $k \geq 1$.

Поток с ограниченным последействием, для которого $A_k(x) = A(x)$ для $k \geq 2$ (заметьте: $A_1(x)$ это не касается), называется потоком **восстановления.** В последнем случае вместо слов «наступило событие потока» часто говорят «наступило восстановление».

Пусть N_t – число событий, наступивших на интервале $(0,t)$. Функция $H(t) = M\{N_t\}$ называется **функцией восстановления**. Функция $h(t) = H'(t)$, если она существует, называется **плотностью восстановления**. По смыслу $h(t)dt$ (или $dH(t)$) есть вероятность того, что на интервале $[t,\ t+dt]$ наступит событие потока (интенсивность равна параметру потока).

Выведем теперь одно из основных уравнений – уравнение для $H(t)$. Имеем

$$P\{N_t = n\} = P\{t_n < t,\ t_{n+1} \geq t\}.$$

Но, с другой стороны,

$$P\{t_n < t\} = P\{t_{n+1} < t\} + P\{t_n < t,\ t_{n+1} > t\},$$

откуда

$$P\{t_n < t,\ t_{n+1} > t\} = P\{t_n < t\} \ - \ P\{t_{n+1} < t\}.$$

Поэтому для $H(t) = M\{N_t\}$ имеем

$$H(t) = M\{N_t\} = \sum_{n=1}^{\infty} nP\{N_t = n\} = 1[P\{t_1 < t\} - P\{t_2 < t\}] +$$

$$+2[P\{t_2 < t\} - P\{t_3 < t\}] + 3[P\{t_3 < t\} - P\{t_4 < t\}] + \ldots = \sum_{n=1}^{\infty} P\{t_n < t\}.$$

Так как $t_n = \tau_1 + \tau_2 + \tau_3 \ldots + \tau_n$, то

$$P\{t_1 < t\} = F_1(t) = A_1(t),$$

$$P\{t_2 < t\} = F_2(t) = P\{t_1 + \tau_2 < t\} = \int_0^t F_1(t-x)\ dA(x),$$

$$P\{t_3 < t\} = F_3(t) = P\{t_2 + \tau_3 < t\} = \int_0^t F_2(t-x)\ dA(x),$$

....

$$P\{t_n < t\} = F_n(t) = P\{t_{n-1} + \tau_n < t\} = \int_0^t F_{n-1}(t-x)\ dA(x).$$

Поэтому

$$H(t) = \sum_{n=1}^{\infty} F_n(t) = A_1(t) + \sum_{n=2}^{\infty} \int_0^t F_{n-1}(t-x)\ dA(x) =$$

$$= A_1(t) + \int_0^t \left(\sum_{n=2}^{\infty} F_{n-1}(t-x) \right) dA(x) = A_1(t) + \int_0^t H(t-x)\ dA(x).$$

Итак, функция восстановления удовлетворяет интегральному уравнению

$$H(t) = A_1(t) + \int_0^t H(t-x)\ dA(x). \tag{2.16}$$

Решить его можно через преобразования Лапласа – Стилтьеса. Действительно, применяя это преобразование к обеим частям уравнения, получим

$$H^*(s) = A_1^*(s) + H^*(s) A^*(s),$$

откуда

$$H^*(s) = \frac{A_1^*(s)}{1 - A^*(s)}. \tag{2.17}$$

Зная $A^*(s)$ и $A_1^*(s)$, отсюда находится $H^*(s)$, а по таблицам обратного преобразования Лапласа – Стилтьеса и $H(t)$.

Пример

Пусть $A(x) = 1 - e^{-\lambda x}$, $A_1(x) = 1 - e^{-\mu x}$. Тогда

$$A^*(s) = \int_0^{\infty} e^{-sx} d(1 - e^{-\lambda x}) = \lambda \int_0^{\infty} e^{-sx-\lambda x} dx = \frac{\lambda}{\lambda + s},$$

$$A_1^*(s) = \frac{\mu}{\mu + s}.$$

Поэтому

$$H^*(s) = \frac{\mu/(\mu+s)}{1 - \lambda/(\lambda+s)} = \frac{\mu(\lambda+s)}{(\mu+s)s} = \frac{\lambda}{s} + \frac{\mu - \lambda}{\mu + s},$$

и отсюда

$$H(t)=\lambda\ \ t+\frac{\mu-\lambda}{\mu}(1-e^{-\mu t})\,.$$

Рассмотрим теперь некоторые следствия из уравнения (2.16) и его решения (2.17).

1. Пусть $a=\int\limits_0^\infty x dA(x)$ и $\lambda=1/a$. Заметим, что

$$A^*(0)=A(\infty)-A(0)=1\,.$$

Тогда

$$\lim_{t\to\infty}\frac{H(t)}{t}=\lim_{s\to 0}sH^*(s)=\lim_{s\to 0}\frac{A_1^*(s)\ \ s}{1-A^*(s)}=\frac{\lim\limits_{s\to 0}A_1^*(s)}{\lim\limits_{s\to 0}\dfrac{A^*(0)-A^*(s)}{s}}=\frac{1}{-A^{*\prime}(0)}=\frac{1}{a}=\lambda\,,$$

то есть при $t\to\infty$ $H(t)$ ведет себя как λt.

2. Пусть $Q(t)$ – неотрицательная интегрируемая на $(0,\ \infty)$ функция. Тогда по свойствам преобразования Лапласа – Стилтьеса

$$\int\limits_0^t Q(t-x)dH(x)\Leftrightarrow Q^*(s)H^*(s)=\frac{Q^*(s)A_1^*(s)}{1-A^*(s)}\,.$$

По свойствам преобразования Лапласа – Стилтьеса

$$\lim_{t\to\infty}\int\limits_0^t Q(t-x)dH(x)=\lim_{s\to 0}Q^*(s)H^*(s)=$$

$$=\lim_{s\to 0}\frac{Q^*(s)}{s}\lim_{s\to\infty}A_1^*(s)\lim_{s\to 0}\frac{s}{1-A^*(s)}=\frac{1}{a}\int\limits_0^\infty Q(x)dx.$$

Окончательно

$$\lim_{t\to\infty}\int\limits_0^t Q(t-x)dH(x)=\lambda\int\limits_0^\infty Q(x)dx\,,\quad \lambda=\frac{1}{a}\,. \tag{2.18}$$

Эта формула называется основной теоремой теории восстановления.

§ 2.5. Распределение величины перескока и недоскока для потоков восстановления

Рассмотрим некоторый произвольный момент времени t в потоке восстановления, который окружают моменты t_n и t_{n+1} наступления событий потока. Величина $t_{n+1} - t = \gamma(t)$ называется временем (или величиной) **перескока**, а величина $\gamma^* = t - t_n$ – величиной **недоскока** (рис. 2.2).

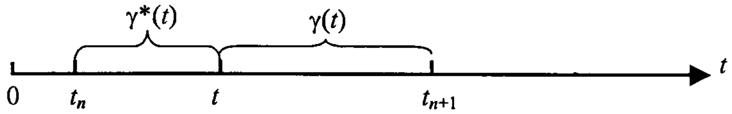


Рис. 2.2. Перескок и недоскок

Пусть $F(t,\ x) = P\{\gamma(t) < x\}$. Выведем уравнение для $F(t,\ x)$.

Возможны следующие варианты:

1. До момента t вообще не наступило ни одного события (рис. 2.3).

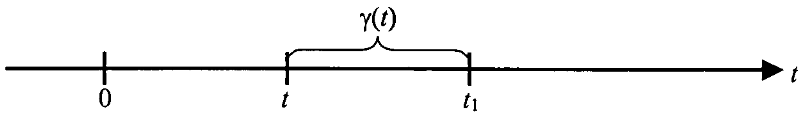


Рис. 2.3

Это будет тогда, когда $t_1 > t$. Так как при этом $\gamma(t) = t_1 - t$, то событие $\gamma(t) < x$ эквивалентно $t_1 < t + x$. Итак, в этом случае

$$P\{\gamma(t) < x\} = P\{\ t < t_1 < t + x\} = A_1(t + x) - A_1(t).$$

2. Пусть $t_n < t < t_{n+1}$ (рис. 2.4).

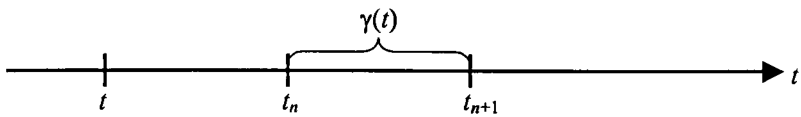


Рис.2.4

В этом случае должно быть, с одной стороны, $\tau > t - t_n$, где $\tau = t_{n+1} - t_n$ – длина интервала между моментами наступления событий в потоке восстановления. Далее, так как $\tau = (t - t_n) + \gamma(t)$, то событие $\gamma(t) < x$ эквивалентно $\tau < (t - t_n) + x$. Итак, окончательно должно выполняться условие

$$t - t_n < \tau < t - t_n + x.$$

Так как n может быть любым в интервале значений $\overline{1,\infty}$, то вероятность этой ситуации равна ($t_n \equiv s$)

$$\sum_{n=1}^{\infty}\int_0^t [A(t-s+x)-A(t-s)]dF_n(s) = \int_0^t [A(t-s+x)-A(t-s)]dH(s).$$

Складывая эти два выражения, получим

$$F(t,x) = A_1(t+x) - A_1(t) + \int_0^t [A(t-s+x)-A(t-s)]dH(s). \quad (2.19)$$

Найдем $F(x) = \lim\limits_{t\to\infty} F(t,x)$, предполагая, что этот предел существует. Тогда

$$\lim_{t\to\infty}[A_1(t+x) - A_1(t)] = A_1(\infty) - A_1(\infty) = 1 - 1 = 0.$$

Далее, представляя второе слагаемое в (2.19) в виде

$$\int_0^t [A(t-s+x)-A(t-s)]dH(s) =$$

$$= \int_0^t [1-A(t-s)]dH(s) - \int_0^t [1-A(t-s+x)]dH(s),$$

(введение 1 необходимо для того, чтобы подынтегральная функция была интегрируемой на $(0, \infty)$) и используя основную теорему теории восстановления, получим

$$F(x) = \lambda\int_0^{\infty}[1-A(z)]dz - \lambda\int_0^{\infty}[1-A(z+x)]dz \underset{z+x=u}{=} \lambda\int_0^{\infty}[1-A(z)]dz -$$

$$-\lambda\int_x^{\infty}[1-A(u)]du.$$

Окончательно имеем следующую важную формулу:

$$F(x) = \lambda\int_0^x [1-A(z)]dz, \quad \lambda = 1/a, \quad (2.20)$$

дающую функцию распределения величины перескока в асимптотическом случае $t\to\infty$.

Рассмотрим теперь величину недоскока $\gamma^*(t)$ и введем функцию $F^*(t,x) = P\{\gamma^*(t) < x\}$. Выведем уравнение для этой функции. Опять-таки возможны два варианта.

1. До момента t не наступило ни одного события. Тогда $\gamma^*(t) = t$.

А) Если взять $x < t$ (рис. 2.5), то мы получим невозможную комбинацию, так как $\gamma^*(t) = t > x$, а нам надо найти $P\{\gamma^*(t) = t < x\}$. Поэтому вероятность такой ситуации равна нулю.

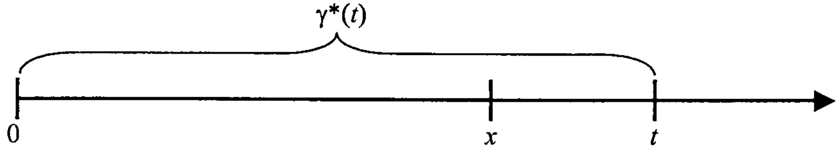


Рис. 2.5

Б) Если взять $x > t$ (рис. 2.6), то рассматриваемая ситуация наступит в том случае, когда $\tau_1 > t$. Вероятность этого равна $1 - A_1(t)$.

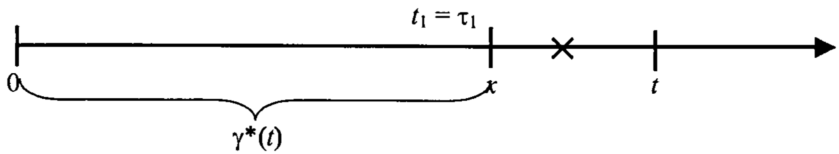


Рис. 2.6

Итак, искомая вероятность данной ситуации равна

$$F_1(t,x) = \begin{cases} 1 - A_1(t)\ , \text{ если } & t < x, \\ 0\ , \text{ если } & t > x. \end{cases}$$

2. Пусть $t_n < t < t_{n+1}$ (рис. 2.7).

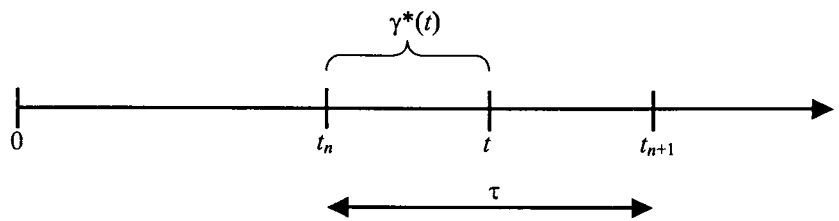


Рис. 2.7

Тогда должно быть 1) $\tau > t - t_n$; 2) $t - t_n = \gamma^*(t) < x$. Отсюда следует, что t_n лежит в интервале $t - x < t_n < t$. Так как вероятность события $\tau > t - t_n$ равна $1 - A(t - t_n)$, то вероятность этой ситуации равна ($t_n \equiv s$)

$$\sum_{n=1}^{\infty} \int_{t-x}^{t} [1 - A(t-s)] \; dF_n(s) = \int_{t-x}^{t} [1 - A(t-s)] \; dH(s).$$

Окончательно

$$F^*(t,x) = F_1(t,x) + \int_0^t [1 - A(t-s)] \; dH(s) - \int_0^{t-x} [1 - A(t-s)] \; dH(s). \quad (2.21)$$

Найдем $F^*(x) = \lim\limits_{t\to\infty} F^*(t,x)$, предполагая, что предел существует. Легко видеть, что $\lim\limits_{t\to\infty} F_1(t,x) = 0$. Далее, по основной теореме теории восстановления

$$\lim_{t\to\infty} \int_0^t [1 - A(t-s)] dH(s) = \lambda \int_0^{\infty} [1 - A(z)] dz,$$

$$\lim_{t\to\infty} \int_0^{t-x} [1 - A(t-s)] dH(s) = \lim_{(t-x)\to\infty} \int_0^{(t-x)} \big[1 - A[(t-x) + x - s]\big] dH(s) =$$

$$= \lim_{T\to\infty} \int_0^T \big[1 - A[T + x - s]\big] dH(s) = \lambda \int_0^{\infty} [1 - A(u + x)] du = \lambda \int_x^{\infty} [1 - A(z)] dz.$$

Окончательно

$$F^*(x) = \lambda \int_0^{\infty} [1 - A(z)] dz - \lambda \int_x^{\infty} [1 - A(z)] dz = \lambda \int_0^{x} [1 - A(z)] dz,$$

то есть $F^*(x) = F(x)$, даваемое формулой (2.20).

Заметим, что **распределение величины перескока и недоскока можно получить и другим способом**, используя тот факт, что процессы $\gamma(t)$ и $\gamma^*(t)$ являются марковскими.

Очевидны следующие выкладки.

Для процесса $\gamma(t)$, характеризующего величину перескока, обозначим

$$P(\gamma(t) < x) = F(t,x),$$

тогда

$$F(t+\Delta t, x) = F(t, x+\Delta t) - F(t, \Delta t) + F(t, \Delta t)A(x) + o(\Delta t).$$

Разложим функцию F по приращению аргументов в ряд Тейлора с точностью до $o(\Delta t)$:

$$F(t,x) + \Delta t \frac{\partial F(t,x)}{\partial t} = F(t,x) + \Delta t \frac{\partial F(t,x)}{\partial x} - \Delta t \frac{\partial F(t,0)}{\partial x} +$$

$$+\Delta t \frac{\partial F(t,0)}{\partial x} A(x) + o(\Delta t).$$

Уничтожая $F(t,x)$ в обеих частях и сокращая на Δt, получим

$$\frac{\partial F(t,x)}{\partial t} = \frac{\partial F(t,x)}{\partial x} - \frac{\partial F(t,0)}{\partial x} + \frac{\partial F(t,0)}{\partial x} A(x) = \frac{\partial F(t,x)}{\partial x} - \frac{\partial F(t,0)}{\partial x}\left(1 - A(x)\right).$$

В стационарном режиме $F(t,x) \equiv F(x)$, тогда $F'(x) = F'(0)\left(1 - A(x)\right)$, следовательно,

$$F(x) = F'(0)\int_0^x \left(1 - A(s)\right) ds .$$

Полагая $x \to \infty$, получим

$$1 = F'(0)\int_0^\infty \left(1 - A(s)\right) ds = F'(0)a ,$$

откуда

$$F'(0) = \frac{1}{a} = \lambda .$$

Таким образом, функция распределения $F(x)$ величины перескока имеет вид

$$F(x) = \lambda\int_0^x \left(1 - A(s)\right) ds .$$

Для процесса $\gamma^*(t)$, характеризующего величину недоскока, обозначим

$$P(\gamma^*(t) < x) = G(t,x) ,$$

$$P(x \le \gamma^*(t) < x + dx) = g(t,x)dx ,$$

тогда

$$g(t+\Delta t, x+\Delta t) = g(x,t)P(\tau > x+\Delta t \mid \tau > x) = g(x,t)\frac{P(\tau > x+\Delta t)}{P(\tau > x)} =$$

$$= g(x,t)\frac{1-A(x+\Delta t)}{1-A(x)} = g(x,t)\frac{1-A(x)+\Delta t A'(x)}{1-A(x)} = g(x,t)\left(1-\Delta t\frac{A'(x)}{1-A(x)}\right),$$

и, следовательно,

$$\frac{\partial g(t,x)}{\partial t} + \frac{\partial g(t,x)}{\partial x} = -\frac{A'(x)}{1-A(x)}g(t,x).$$

В стационарном режиме $g(t,x) \equiv g(x)$, поэтому $g(x)$ удовлетворяет уравнению

$$g'(x) = -\frac{A'(x)}{1-A(x)}g(x),$$

решение которого имеет вид

$$g(x) = C\left(1-A(x)\right).$$

Интегрируя это равенство по x в интервале $0 \le x < \infty$, получим

$$1 = C\int_0^\infty \left(1-A(s)\right)ds = Ca,$$

откуда следует, что

$$C = \frac{1}{a} = \lambda.$$

Таким образом,

$$g(x) = \lambda\left(1-A(x)\right),$$

а для функции распределения $G(x)$ величины недоскока можно записать как

$$G(x) = \lambda \int_0^x \left(1-A(s)\right)ds.$$

Парадокс остаточного времени

Пусть поток трамваев есть поток событий восстановления с функцией распределения интервалов между трамваями $A(x)$. В случайный момент времени t человек приходит на остановку. Сколько в среднем времени ему ждать трамвая?

1. **Первое решение.** Так как в среднем длина интервала времени между трамваями равна $a=\int\limits_0^\infty x dA(x)$ и человек приходит совершенно произвольно, то ему ждать в среднем $a/2$.

2. **Второе решение.** Время ожидания трамвая – это время перескока, распределенное с функцией распределения $\frac{1}{a}\int\limits_0^x [1-A(z)]dz$, т.е. с плотностью вероятностей $\frac{1}{a}[1-A(x)]$. Поэтому ждать ему в среднем

$$\frac{1}{a}\int\limits_0^\infty x[1-A(x)]dx=\frac{1}{2a}\int\limits_0^\infty [1-A(x)]d(x^2)=\frac{1}{2a}x^2[1-A(x)]\Big|_0^\infty+$$

$$+\frac{1}{2a}\int\limits_0^\infty x^2 dA(x)=\frac{M\{x^2\}}{2a}=\frac{a^2+\sigma_a^2}{2a}=\frac{a}{2}+\frac{\sigma_a^2}{2a}>\frac{a}{2}. \qquad (2.22)$$

Как видно, два решения дали два разных результата, причем второе решение дает большее среднее время ожидания, чем первое. Какой же результат верен?

Верен второй результат. Разгадка в том, что промежутки времени между приходами трамваев разные. И, приходя в **произвольный** момент времени, человек с большей вероятностью попадает на более длинный промежуток времени между трамваями, чем на более короткий.

§ 2.6. Основное свойство рекуррентных потоков

Определение. Стационарный поток восстановления называется рекуррентным.

Теорема. Если поток восстановления является стационарным, то есть является рекуррентным и $a=\int\limits_0^\infty x dA(x)<+\infty$, то

$$A_1(x)=\frac{1}{a}\int\limits_0^x [1-A(z)]\ dz. \qquad (2.23)$$

Доказательство

1. Докажем, что если поток стационарен, то $H(t)$ – линейная по t функция.

Так как поток стационарен, то величины $N_{t+s}-N_s$ и $N_t-N_0=N_t$ распределены одинаково, тогда

$$H(t+s)=M\{N_{t+s}\}=M\{N_{t+s}-N_s+N_s-N_0\}=$$

$$=M\{N_{t+s}-N_s\}+M\{N_s-N_0\}=M\{N_t\}+M\{N_s\}=H(t)+H(s).$$

Следовательно, $H(t)$ является линейной функцией вида

$$H(t)=kt,$$

а так как $\lim\limits_{t\to\infty}\frac{H(t)}{t}=\lambda$, то $k=\lambda=\frac{1}{a}$.

2. Возьмем уравнение для функции восстановления

$$H(t)=A_1(t)+\int_0^t A(t-s)dH(s)$$

и подставим туда $H(t)=kt$, где $k=\frac{1}{a}$. Тогда

$$\frac{1}{a}t=A_1(t)+\frac{1}{a}\int_0^t A(t-s)\,ds=A_1(t)+\frac{1}{a}\int_0^t A(z)\,dz,$$

откуда

$$A_1(t)=\frac{1}{a}t-\frac{1}{a}\int_0^t A(t-s)\,ds=\frac{1}{a}\int_0^t[1-A(z)]dz,$$

или, если заменить $t=x$,то

$$A_1(x)=\frac{1}{a}\int_0^x[1-A(z)]dz.$$

Таким образом, рекуррентный поток определяется единственной функцией распределения $A(x)$.

Следствие. Если $A_1(x)\neq\frac{1}{a}\int_0^x[1-A(z)]dz$, то процесс восстановления не может быть стационарным, то есть не является рекуррентным.

Частный случай. Найдем, при каких условиях выполняется соотношение $A_1(x)=A(x)$. Тогда

$$A(x)=\frac{1}{a}\int_0^x[1-A(z)]dz=\lambda\int_0^x[1-A(z)]dz.$$

Дифференцируя по x, получим

$$A'(x) = \lambda[1 - A(x)] = -[1 - A(x)]' = -\frac{d[1 - A(x)]}{dx}.$$

Разделяя переменные

$$\frac{d(1 - A(x))}{1 - A(x)} = -\lambda dx$$

и интегрируя, получим

$$\ln(1 - A(x)) = \lambda x,$$

откуда $A(x) = 1 - e^{-\lambda x}$ и поток является пуассоновским потоком с постоянной интенсивностью λ.

Сумма независимых рекуррентных потоков

Рассмотрим сумму двух независимых рекуррентных потоков, определяемых функциями $A_1(x)$ и $A_2(x)$ (рис. 2.8).

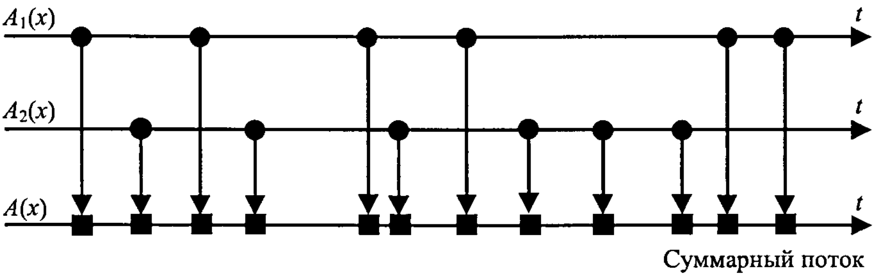


Рис. 2.8. Схема образования суммарного потока

Пусть τ – величина перескока суммарного потока. Тогда очевидно, что $\tau = \min(\tau_1, \tau_2)$, где τ_1, τ_2 – независимые величины перескоков для первого и второго суммируемых рекуррентных потоков.

Поэтому

$$P(\tau > x) = P(\min(\tau_1, \tau_2) > x) = P(\tau_1 > x)P(\tau_2 > x) =$$

$$= \left(1 - \lambda_1 \int_0^x (1 - A_1(s))ds\right)\left(1 - \lambda_2 \int_0^x (1 - A_2(s))ds\right).$$

Следовательно,

$$1-\lambda\int_0^x(1-A(s))ds=\left(1-\lambda_1\int_0^x(1-A_1(s))ds\right)\left(1-\lambda_2\int_0^x(1-A_2(s))ds\right),$$

тогда

$$\lambda\int_0^x(1-A(s))ds=\lambda_1\int_0^x(1-A_1(s))ds+\lambda_2\int_0^x(1-A_2(s))ds-$$

$$-\lambda_1\lambda_2\int_0^x(1-A_1(s))ds\int_0^x(1-A_2(s))ds\,.$$

Таким образом,

$$\lambda(1-A(x))=\lambda_1(1-A_1(x))+\lambda_2(1-A_2(x))-$$

$$-\lambda_1\lambda_2\left[(1-A_1(x))\int_0^x(1-A_2(s))ds+(1-A_2(x))\int_0^x(1-A_1(s))ds\right].$$

Так как $\lambda=\lambda_1+\lambda_2$, то функция $A(x)$ для суммы рекуррентных потоков будет иметь вид

$$A(x)=1-\frac{\lambda_1}{\lambda_1+\lambda_2}(1-A_1(x))-\frac{\lambda_2}{\lambda_1+\lambda_2}(1-A_2(x))+$$

$$+\frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}\left[(1-A_1(x))\int_0^x(1-A_2(s))ds+(1-A_2(x))\int_0^x(1-A_1(s))ds\right].$$

В частности, если $A_k(x)=1-e^{-\lambda_k x}$, $k=1,\ 2$, то есть суммируются простейшие потоки, то

$$A(x)=1-\frac{\lambda_1}{\lambda_1+\lambda_2}e^{-\lambda_1 x}-\frac{\lambda_2}{\lambda_1+\lambda_2}e^{-\lambda_2 x}+$$

$$+\frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}\left[e^{-\lambda_1 x}\frac{1}{\lambda_2}(1-e^{-\lambda_2 x})+e^{-\lambda_2 x}\frac{1}{\lambda_1}(1-e^{-\lambda_1 x})\right]=1-\frac{1}{\lambda_1+\lambda_2}\times$$

$$\times\left[\lambda_1 e^{-\lambda_1 x}+\lambda_2 e^{-\lambda_2 x}-\lambda_1 e^{-\lambda_1 x}(1-e^{-\lambda_2 x})-\lambda_2 e^{-\lambda_2 x}(1-e^{-\lambda_1 x})\right]=1-e^{-(\lambda_1+\lambda_2)x}\,.$$

Следовательно, суммой независимых простейших потоков является простейший поток с параметром, равным сумме параметров исходных потоков.

Биноминальная схема деления рекуррентного потока

Биноминальная схема деления рекуррентного потока реализуется следующим образом. Каждое событие исходного потока с заданной и постоянной вероятностью p назначается в первый, а с вероятностью $1-p$ во второй разделенные потоки соответственно (рис. 2.9).

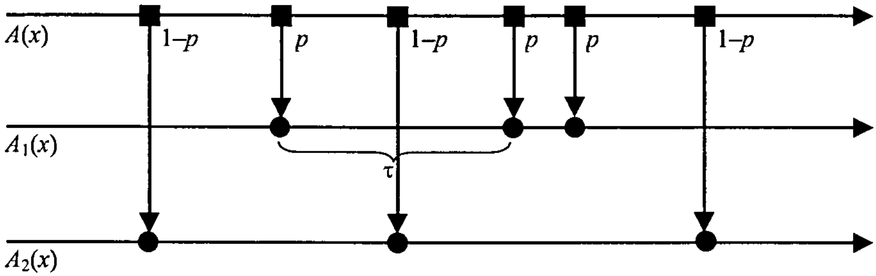


Рис. 2.9. Схема деления потока

Исходный поток определяется функцией распределения $A(x)$ длин интервалов между моментами наступления событий. Обозначим через $A_1(x)$ и $A_2(x)$ функции распределения длин интервалов между моментами наступления событий в разделенных потоках. Найдем характеристическую функцию $M\left\{e^{-\alpha\tau}\right\}$ длин интервалов для первого из разделенных потоков.

Пусть τ – длина интервала в первом из разделенных потоков, а ν – случайное число интервалов исходного потока, реализующий интервал длины τ. Очевидно,

ν	Вероятность
1	p
2	$(1-p)p$
k	$(1-p)^{k-1}p$

Тогда

$$M\left\{e^{-\alpha\tau}\right\}=\sum_{k=1}^{\infty}M\left\{e^{-\alpha\tau}\mid\nu=k\right\}P\{\nu=k\}=\sum_{k=1}^{\infty}\varphi^k(\alpha)p(1-p)^{k-1}=$$

$$=p\varphi(\alpha)\frac{1}{1-(1-p)\varphi(\alpha)}.$$

Обращая эту функцию, можно найти функцию распределения $A_1(x)$ для первого потока. Аналогично определяется и $A_2(x)$.

Найдем среднее значение длины интервалов в первом из разделенных потоков. Используя свойства характеристических функций и дифференцируя по α в нуле это равенство, запишем

$$-M\tau = -a_1 = p\left.\frac{\varphi'(\alpha)(1-(1-p)\varphi(\alpha)) + \varphi(\alpha)(1-p)\varphi'(\alpha)}{\left[1-(1-p)\varphi(\alpha)\right]^2}\right|_{\alpha=0} =$$

$$= p\frac{-ap-(1-p)a}{p^2} = -\frac{a}{p},$$

то есть средняя длина интервала для выделенного потока имеет вид $a_1 = a/p$.

Тогда $\lambda_1 = 1/a_1 = p/a = \lambda p \Rightarrow \lambda_1 = p\lambda$.

Для простейшего потока с параметром λ имеем

$$\varphi(\alpha) = \int_0^\infty e^{-\alpha x}\lambda \; e^{-\lambda x}dx = \frac{\lambda}{\lambda+\alpha},$$

следовательно,

$$M\left\{e^{-\alpha\tau}\right\} = p\frac{\lambda/(\lambda+\alpha)}{1-(1-p)\lambda/(\lambda+\alpha)} = p\frac{\lambda}{\lambda+\alpha-(1-p)\lambda} = \frac{\lambda p}{\alpha+\lambda p}.$$

При биномиальной схеме деления простейшего потока каждый поток – простейший с параметром λp или $\lambda(1-p)$.

Глава 3. МАРКОВСКИЕ МОДЕЛИ МАССОВОГО ОБСЛУЖИВАНИЯ

§ 3.1. Модели и обозначения

Всякая система массового обслуживания (СМО) состоит из:

1) входящих потоков заявок (требований);
2) накопителей (бункеров) для ожидающих обслуживания заявок или источника повторных вызовов (ИПВ);
3) обслуживающих приборов (линий);
4) дисциплины обслуживания.

Все это кратко записывается символикой Кендалла в виде

$$A \mid B \mid \mathrm{N}_1 \mid \mathrm{N}_2 \mid \mathrm{f}$$

Здесь первый символ означает входящий поток заявок. Наиболее часто встречаются следующие символы: M – простейший поток заявок с постоянной интенсивностью; $M(t)$ – пуассоновский с $\lambda = \lambda(t)$; E_k – эрланговский поток k-го порядка; GI – рекуррентный поток; G – поток с зависимыми интервалами.

Если потоков n, то индексом указывается число потоков и вверху ставится стрелка. Например, $\vec{M}_n$ означает, что в систему поступает n простейших потоков заявок постоянной интенсивности.

Второй символ – распределение времени обслуживания: M – экспоненциальное распределение; E_k – эрланговское порядка k; G – распределение времени обслуживания произвольное.

Если время обслуживания разное для разных типов заявок, то над соответствующим символом ставится стрелка; N_1 – число обслуживающих приборов; N_2 – число мест для ожидания.

Что касается дисциплины обслуживания, то в этой главе мы рассмотрим лишь те дисциплины, которые встречаются в СМО с одним входящим потоком. Наиболее часто используются:

FIFO – «first input – first output» – дисциплина «первым пришел – первым обслужен»;

LIFO – «last input – first output» – дисциплина «последний пришел – первый обслужен».

§ 3.2. Нестационарный режим в системе $M \mid M \mid \infty$

Рассмотрим (рис. 3.1) бесконечно линейную СМО, на вход которой поступает простейший с параметром λ поток заявок, обслуживание экспоненциальное с параметром μ. Пусть $i(t)$ – число заявок в системе, то есть число приборов, занятых в момент времени t.

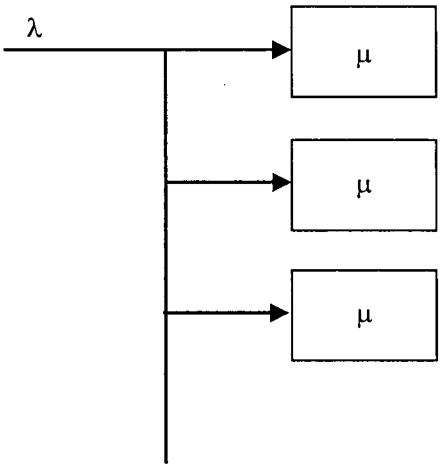


Рис. 3.1. Блок-схема бесконечно линейной СМО

Рассмотрим этот случайный процесс. В силу того, что входящий поток простейший (обладает свойством отсутствия последействия), а время обслуживания имеет экспоненциальное распределение (также обладает этим свойством), то очевидно, что процесс $i(t)$ является цепью Маркова с непрерывным временем.

Систему массового обслуживания будем называть марковской, если процесс изменения ее состояний является цепью Маркова с непрерывным временем. Следовательно, рассматриваемая система $M \mid M \mid \infty$ относится к классу марковских моделей массового обслуживания и поэтому для ее исследования можно применить теорию цепей Маркова с непрерывным временем и прежде всего составить прямую систему дифференциальных уравнений Колмогорова для следующего распределения вероятностей:

$$P_i(t) = P(i(t) = i).$$

Из курса «Теории случайных процессов» известно, что такая система имеет вид

$$P_i'(t) = \sum_j P_j(t) q_{ji},$$

где

$$q_{ji} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(i(t+\Delta t) = j \mid i(t) = i) \quad \text{при} \quad i \neq j,$$

$$q_{ii} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[P(i(t+\Delta t) = i \mid i(t) = i) - 1 \right].$$

Так как рассматриваемый случайный процесс $i(t)$ является процессом гибели и размножения, то ненулевыми могут быть лишь $q_{i,i-1}$, q_{ii} и $q_{i,i+1}$ для $i > 0$, а также q_{00} и q_{01}. При этом, как обычно, имеет место равенство

$$q_{ii} = -(q_{i,i-1} + q_{i,i+1}).$$

Из свойств простейшего потока для рассматриваемой СМО очевидно

$$P(i(t+\Delta t) = i+1 \mid i(t) = i) = \lambda \Delta t + o(\Delta t),$$

и поэтому $q_{i,i+1} = \lambda$.

Для $i > 0$ рассмотрим переходные вероятности $P(i(t+\Delta t) = i-1 \mid i(t) = i)$. Это – условная вероятность того, что к моменту времени $t + \Delta t$ в системе останется $i-1$ заявка при условии, что в момент времени t было занято i приборов, то есть за время Δt освободится один из приборов. Так как остаточное время обслуживания каждым прибором экспоненциальное с параметром μ, а занятых приборов i и освободиться может любой из них, то

$$P(i(t+\Delta t) = i-1 \mid i(t) = i) = i\mu\Delta t + o(\Delta t),$$

откуда $q_{i,i-1} = i\mu$ и $q_{ii} = -(\lambda + i\mu)$.

Таким образом, прямая система дифференциальных уравнений Колмогорова для рассматриваемого распределения вероятностей $P_i(t)$ имеет вид

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t),$$

$$P_i'(t) = \lambda P_{i-1}(t) - (\lambda + i\mu) P_i(t) + (i+1)\mu P_{i+1}(t). \tag{3.1}$$

В теории массового обслуживания вывод этих уравнений осуществляется обычно по-другому. Для этого записывают допредельные равенства

$$P_0(t+\Delta t) = (1 - \lambda\Delta t) P_0(t) + \mu \Delta t P_1(t) + o(\Delta t),$$

$$P_i(t+\Delta t) = \lambda \Delta t P_{i-1}(t) + [1 - (\lambda + i\mu)\Delta t] P_i(t) + i\mu\Delta t P_{i+1}(t) + o(\Delta t),$$

которые получаются с использованием формулы полной вероятности. По сути, вероятности $P_i(t+\Delta t)$ в момент времени $t + \Delta t$ выражаются через вероятности $P_i(t)$ в момент времени t и вероятности переходов в состояние i за бесконечно малый промежуток времени Δt.

Затем от полученных допредельных равенств переходим к пределу при $\Delta t \to 0$ и получаем систему (3.1).

Такой способ вывода уравнений Колмогорова в теории массового обслуживания называют иногда Δt-методом, хотя достаточно очевидно, что Δt-метод есть просто процедура вывода прямой системы дифференциальных уравнений Колмогорова для распределения вероятностей цепи Маркова с непрерывным временем, описывающих процесс изменения состояний СМО.

Для однозначного решения системы (3.1) надо задать начальные условия. Рассмотрим случай, когда в момент времени $t = 0$ система пуста, то есть все приборы свободны. Тогда начальные условия имеют вид

$$P_i(0) = \begin{cases} 1, & \text{если } i = 0, \\ 0, & \text{если } i > 0. \end{cases}$$

Чтобы решить систему (3.1), определим производящую функцию $F(z,\ t) = \sum_{i=0}^{\infty} z^i P_i(t)$, для которой выполнены следующие свойства:

$$F(z,\ 0) = 1,$$

$$\sum_{i=1}^{\infty} z^i P_{i-1}(t) = zF(z,\ t),$$

$$\sum_{i=1}^{\infty} iz^i P_i(t) = z\frac{\partial F(z,\ t)}{\partial z},$$

$$\sum_{i=0}^{\infty} (i+1) z^i P_{i+1}(t) = \frac{\partial F(z,\ t)}{\partial z}.$$

Следовательно, из системы (3.1) можно получить уравнение для производящей функции в виде

$$\frac{\partial F(z,\ t)}{\partial t} = \lambda z F(z,\ t) - \lambda F(z,\ t) - \mu z \frac{\partial F(z,\ t)}{\partial z} + \mu \frac{\partial F(z,\ t)}{\partial z}.$$

Делая несложные преобразования и опуская аргументы у $F(z,\ t)$, получим

$$\frac{\partial F}{\partial t} + \mu(z-1)\frac{\partial F}{\partial z} = \lambda(z-1)F. \tag{3.2}$$

Для полученного дифференциального уравнения в частных производных первого порядка система обыкновенных дифференциальных уравнений для характеристик имеет вид

$$\frac{dt}{1}=\frac{dz}{\mu(z-1)}=\frac{dF}{\lambda(z-1)F}.$$

Для этой системы найдем два первых интеграла. Сначала рассмотрим первое равенство. Для него

$$t=\frac{1}{\mu}\ln(z-1)+\frac{1}{\mu}\ln C_1,$$

следовательно, $e^{\mu t}=C_1(z-1)$, тогда $C_1=\frac{1}{z-1}e^{\mu t}$.

Из второго равенства получим $\frac{dz}{\mu}=\frac{dF}{\lambda F}$, откуда следует $\ln F=\frac{\lambda}{\mu}z+\ln C_2$, тогда $F=e^{\frac{\lambda}{\mu}z}C_2$, а $C_2=e^{-\frac{\lambda}{\mu}z}F$.

Следовательно, общее решение уравнения (3.2) имеет вид

$$e^{-\frac{\lambda}{\mu}z}F=\varphi\left[\frac{1}{z-1}e^{\mu t}\right],$$

$$F(z,\ t)=e^{\frac{\lambda}{\mu}z}\varphi\left[\frac{1}{z-1}e^{\mu t}\right],$$

где φ – произвольная дифференцируемая функция. Найдем ее вид, используя начальное условие $F(z,0)=1$:

$$F(z,\ 0)=e^{\frac{\lambda}{\mu}z}\varphi\left[\frac{1}{z-1}\right]=1.$$

Отсюда

$$\varphi\left[\frac{1}{z-1}\right]=e^{-\frac{\lambda}{\mu}z}.$$

Выполнив замену $\frac{1}{z-1}=y$, получим

$$\varphi(y)=e^{-\frac{\lambda}{\mu}\left(1+\frac{1}{y}\right)}. \qquad (3.3)$$

Следовательно, $F(z,t)$ имеет вид

$$F(z,t)=e^{\frac{\lambda}{\mu}z}\varphi\left[\frac{1}{z-1}e^{\mu t}\right],$$

где φ определяется равенством (3.3), используя которое, запишем

$$F(z,\ t)=e^{\frac{\lambda}{\mu}z}e^{-\frac{\lambda}{\mu}\left(1+\frac{1}{\frac{1}{z-1}e^{\mu t}}\right)}=e^{\frac{\lambda}{\mu}z}e^{-\frac{\lambda}{\mu}\left(1+(z-1)e^{-\mu t}\right)}=e^{\frac{\lambda}{\mu}(z-1)\left[1-e^{-\mu t}\right]}=e^{-\frac{\lambda}{\mu}(1-z)\left[1-e^{-\mu t}\right]}.$$

Таким образом, производящая функция $F(z,t)$ имеет вид

$$F(z,\ t)=e^{-\frac{\lambda}{\mu}(1-z)\left[1-e^{-\mu t}\right]}.$$

Разлагая эту функцию в ряд по степеням z^i, найдем вероятности $P_i(t)$ следующим образом:

$$e^{-\frac{\lambda}{\mu}(1-z)\left[1-e^{-\mu t}\right]}=e^{-\frac{\lambda}{\mu}\left[1-e^{-\mu t}\right]}e^{\frac{\lambda}{\mu}z\left[1-e^{-\mu t}\right]}=e^{-\frac{\lambda}{\mu}\left[1-e^{-\mu t}\right]}\sum_{i=0}^{\infty}\frac{\left[\frac{\lambda}{\mu}z\left[1-e^{-\mu t}\right]\right]^i}{i!}=$$

$$=\sum_{i=0}^{\infty}z^i\frac{\left[\frac{\lambda}{\mu}\left[1-e^{-\mu t}\right]\right]^i}{i!}e^{-\frac{\lambda}{\mu}\left[1-e^{-\mu t}\right]}=\sum_{i=0}^{\infty}z^iP_i(t).$$

Следовательно, распределение является пуассоновским:

$$P_i(t)=\frac{\left[\frac{\lambda}{\mu}\left[1-e^{-\mu t}\right]\right]^i}{i!}e^{-\frac{\lambda}{\mu}\left[1-e^{-\mu t}\right]}.$$

Отсюда определим финальное распределение

$$\lim_{t\to\infty}P_i(t)=\pi(i)=\frac{(\lambda/\mu)^i}{i!}e^{-\lambda/\mu}.$$

Известно, что финальное распределение совпадает со стационарным. Следовательно, стационарное распределение $\pi(i)$ числа приборов занятых в системе $M\mid M\mid\infty$, является пуассоновским с параметром λ/μ.

§ 3.3. Нестационарный режим в системе $M(t) \mid M \mid \infty$

Аналогично рассмотрим бесконечно линейную СМО с нестационарным пуассоновским потоком (рис. 3.1), когда интенсивность входящего потока $\lambda(t)$ зависит от времени.

Пусть $i(t)$ – число приборов, занятых в момент t. Обозначим $P(i(t)=i)=P(i,t)$. Начальные условия в момент t_0 зададим в общем виде $P(i,t_0)=q(i)$, где $q(i)$ – заданное распределение вероятностей.

Так как

$$P(0,t+\Delta t)=(1-\lambda(t)\Delta t)P(0,t)+\mu\Delta tP(1,t),$$

$$P(i,t+\Delta t)=\left[1-(\lambda(t)+i\mu)\Delta t\right]P(i,t)+\lambda(t)P(i-1,t)+(i+1)\mu\Delta tP(i+1,t),$$

то распределение $P(i,t)$ удовлетворяет системе уравнений, аналогичной (3.1):

$$\frac{\partial P(0,t)}{\partial t}+\lambda(t)P(0,t)=\mu P(1,t),$$

$$\frac{\partial P(i,t)}{\partial t}+(\lambda(t)+i\mu)P(i,t)=\lambda(t)P(i-1,t)+(i+1)\mu P(i+1,t). \quad (3.4)$$

Применяя метод производящих функций, запишем

$$\sum_{i=0}^{\infty}z^iP(i,t)=G(z,t), \quad \sum_{i=0}^{\infty}z^iq(i)=Q(z),$$

$$\sum_{i=1}^{\infty}iz^iP(i,t)=z\left(\sum_{i=0}^{\infty}z^iP(i,t)\right)'_z=z\frac{\partial G(z,t)}{\partial z},$$

$$\sum_{i=0}^{\infty}z^i(i+1)P(i+1,t)=\frac{\partial}{\partial z}\left\{\sum_{i=0}^{\infty}z^{i+1}P(i+1,t)\right\}=\frac{\partial}{\partial z}\left\{\sum_{j=1}^{\infty}z^jP(j,t)\right\}=\frac{\partial G(z,t)}{\partial z}.$$

Из (3.4) получим равенство

$$\frac{\partial G(z,t)}{\partial t}+\lambda(t)G(z,t)+\mu z\frac{\partial G(z,t)}{\partial z}=\lambda(t)zG(z,t)+\mu\frac{\partial G(z,t)}{\partial z},$$

из которого вытекает уравнение для производящей функции

$$\frac{\partial G(z,t)}{\partial t}+\mu(z-1)\frac{\partial G(z,t)}{\partial z}=\lambda(t)(z-1)G(z,t) \quad (3.5)$$

с начальным условием

$$G(z,\ t_0)=Q(z)\,. \tag{3.6}$$

Аналогично уравнению (3.2), решим уравнение (3.5), используя систему обыкновенных дифференциальных уравнений для характеристик

$$\frac{dt}{1}=\frac{dz}{\mu(z-1)}=\frac{dG}{\lambda(t)(z-1)G}\,.$$

Из уравнения $\mu dt=\dfrac{dz}{z-1}$ найдем первый интеграл:

$$\mu t+\ln c_1=\ln(z-1)\,,\qquad c_1=(z-1)e^{-\mu t}$$

$$c_1e^{\mu t}=z-1\,. \tag{3.7}$$

Из второго уравнения найдем другой первый интеграл:

$$\lambda(t)(z-1)dt=\frac{dG}{G}\,.$$

Переменную z заменим, используя равенство (3.7):

$$\lambda(t)c_1e^{\mu t}dt=\frac{dG}{G}\,,$$

откуда получим

$$\ln G=\ln c_2+\int\limits_{t_0}^{t}\lambda(s)c_1e^{\mu s}ds\,.$$

Выполняя несложные преобразования, запишем

$$G=c_2\exp\left\{\int\limits_{t_0}^{t}\lambda(s)c_1e^{\mu s}ds\right\}=c_2\exp\left\{\int\limits_{t_0}^{t}\lambda(s)(z-1)e^{-\mu(t-s)}ds\right\}=\left[t-s=\tau\right]=$$

$$=c_2\exp\left\{-\int\limits_{t-t_0}^{0}\lambda(t-\tau)(z-1)e^{-\mu\tau}d\tau\right\}=c_2\exp\left\{(z-1)\int\limits_{0}^{t-t_0}\lambda(t-\tau)e^{-\mu\tau}d\tau\right\}.$$

Таким образом, общее решение уравнения (3.5) имеет вид

$$G=\Phi\left((z-1)e^{-\mu t}\right)\ \exp\left\{(z-1)\int\limits_{0}^{t-t_0}\lambda(t-\tau)e^{-\mu\tau}d\tau\right\},$$

где $\Phi(\cdot)$ – произвольная функция.

Используя начальное условие (3.6), запишем

$$G(z,t_0)=Q(z)=\Phi\left((z-1)e^{-\mu t_0}\right).$$

Выполнив замены $(z-1)e^{-\mu t_0}=y$, $z=1+ye^{\mu t_0}$, получим

$$\Phi(y)=Q(1+ye^{\mu t_0}),$$

следовательно, производящая функция $G(z,t)$ имеет вид

$$G(z,t)=Q(1+(z-1)e^{-\mu(t-t_0)})\exp\left\{(z-1)\int\limits_0^{t-t_0}\lambda(t-\tau)e^{-\mu\tau}d\tau\right\}. \qquad (3.8)$$

Разложив ее в ряд по степеням z^i, нетрудно найти нестационарное распределение вероятностей $P(i,t)$, числа приборов, занятых в системе $M(t)|\,M|\,\infty$.

§ 3.4. Стационарный режим в $M\,|\,M\,|\,\infty$

Пусть $\lambda(t)=\lambda$, $\sum\limits_i P(i,t)z^i=G(z,t)$. Используя равенство (3.8)

$$G(z,t)=Q(1+(z-1)e^{-\mu(t-t_0)})\exp\left\{(z-1)\lambda\frac{1-e^{-\mu(t-t_0)}}{\mu}\right\},$$

найдем $\lim\limits_{t\to\infty}G(z,t)=H(z)$ – производящую функцию финального распределения. Выполняя предельный переход, получим

$$H(z)=\exp\{(z-1)\rho\}=\sum_{i=0}^{\infty}z^i\frac{\rho^i}{i!}e^{-\rho}=\sum_{i=0}^{\infty}\pi(i)z^i.$$

Естественно, что полученный результат совпадает с распределением, полученным ранее:

$$\pi(i)=\frac{\rho^i}{i!}e^{-\rho}.$$

Покажем, что если в однородной СМО в качестве начального распределения $Q(z)$ взять финальное $H(z)$, то система функционирует в стационарном режиме:

$$Q(z)=H(z),$$

$$G(z,t)=H(1+(z-1)e^{-\mu(t-t_0)})\exp\{(z-1)\rho(1-e^{-\mu(t-t_0)})\}=$$

$$=\exp\{\rho(z-1)e^{-\mu(t-t_0)}\}\exp\{\rho(z-1)(1-e^{-\mu(t-t_0)})\}=\exp\{\rho(z-1)\},$$

$$\mathrm{P}(i,t)=\pi(i)=\frac{\rho^i}{i!}e^{-\rho}.$$

§ 3.5. Система $M \mid M \mid 1 \mid$ ИПВ

Рассмотрим однолинейную марковскую СМО с источником повторных вызовов (рис. 3.2)

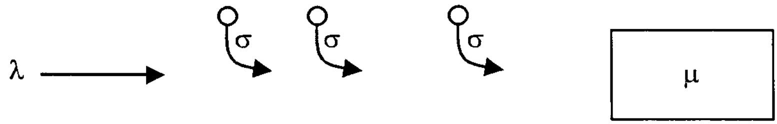


Рис. 3.2. Блок-схема СМО с источником повторных вызовов

На вход системы поступает простейший поток заявок с интенсивностью λ. Требование, заставшее прибор свободным, занимает его для обслуживания в течение случайного времени, распределенного по экспоненциальному закону с параметром μ. Если прибор занят, то поступившая заявка переходит в источник повторных вызовов (ИПВ), в котором осуществляет случайную задержку, продолжительность которой имеет экспоненциальное распределение с параметром σ. Из ИПВ после случайной задержки заявка вновь обращается к прибору с повторной попыткой его захвата. Если прибор свободен, то заявка из ИПВ занимает его на случайное время обслуживания, если же он занят, то заявка мгновенно возвращается в источник повторных вызовов для реализации следующей задержки случайной продолжительности.

Примером подобной СМО является обычный телефонный аппарат, стоящий у абонента, которому звонят клиенты. Она является наиболее простой моделью возникающей ситуации.

Пусть $i(t)$ – число заявок в ИПВ, а $k(t)$ определяет состояние прибора следующим образом:

$$k(t)=\begin{cases}0, & \text{если прибор свободен,}\\ 1, & \text{если прибор занят.}\end{cases}$$

Обозначим $P(k(t)=k,\ \ i(t)=i)=P_k(i,t)$.

Нетрудно показать, что распределение $P_k(i,t)$ удовлетворяет равенствам

$$P_0(0,t+\Delta t)=P_1(0,t)\mu\Delta t+P_0(0,t)(1-\lambda\Delta t)+o(\Delta t),$$

$$P_1(0,t+\Delta t)=\lambda\Delta tP_0(0,t)+\sigma\Delta tP_0(1,t)+\left[1-(\lambda+\mu)\Delta t\right]P_1(0,t)+o(\Delta t),$$

$$P_0(i,t+\Delta t)=\mu\Delta tP_1(i,t)+\left[1-(\lambda+i\sigma)\Delta t\right]P_0(i,t)+o(\Delta t),$$

$$P_1(i,t+\Delta t)=\lambda\Delta t P_0(i,t)+(i+1)\sigma\Delta t P_0(i+1,t)+\lambda\Delta t P_1(i-1,t)+$$
$$+\left[1-(\lambda+\mu)\Delta t\right]P_1(i,t)+o(\Delta t).$$

Отсюда следует, что распределение $P_k(i,t)$ определяется следующей системой дифференциальных уравнений:

$$P_0'(0,t)=-\lambda P_0(0,t)+\mu P_1(0,t),$$
$$P_1'(0,t)=\lambda P_0(0,t)+\sigma P_0(1,t)-(\lambda+\mu)P_1(0,t),$$
$$P_0'(i,t)=\mu P_1(i,t)-(\lambda+i\sigma)P_0(i,t),$$
$$P_1'(i,t)=\lambda P_0(i,t)+(i+1)\sigma P_0(i+1,t)+\lambda P_1(i-1,t)-(\lambda+\mu)P_1(i,t). \quad (3.9)$$

Составим систему уравнений, определяющих производящие функции:

$$F_k(z,t)=\sum_{i=0}^{\infty}z^i P_k(i,t).$$

Из (3.9) получим

$$\frac{\partial F_0(z,t)}{\partial t}=\mu F_1(z,t)-\lambda F_0(z,t)-\sigma\sum_{i=1}^{\infty}z^i i P_0(i,t),$$
$$\frac{\partial F_1(z,t)}{\partial t}=\lambda F_0(z,t)-\lambda z F_1(z,t)+$$
$$+\sigma\sum_{i=0}^{\infty}(i+1)z^i P_0(i+1,t)-(\lambda+\mu)F_1(z,t),$$
$$\frac{\partial F_0(z,t)}{\partial t}=\mu F_1(z,t)-\lambda F_0(z,t)-\sigma z\frac{\partial F_0(i,z)}{\partial z},$$
$$\frac{\partial F_1(z,t)}{\partial t}=\lambda F_0(z,t)-\lambda z F_1(z,t)+\sigma\frac{\partial F_0(z,t)}{\partial z}-(\lambda+\mu)F_1(z,t)-\sigma z\frac{\partial F_1(z,t)}{\partial z}.$$

Следовательно, производящие функции $F_k(z,t)$ распределения вероятностей $P_k(i,t)$ состояний (k,i) СМО с источником повторных вызовов определяются системой двух уравнений с частными производными первого порядка следующего вида:

$$\frac{\partial F_0(z,t)}{\partial t}+\sigma z\frac{\partial F_0(z,t)}{\partial z}=\mu F_1(z,t)-\lambda F_0(z,t),$$

$$\frac{\partial F_1(z,t)}{\partial t} - \sigma\frac{\partial F_0(z,t)}{\partial z} = \lambda F_0(z,t) + (\lambda z - \lambda - \mu)F_1(z,t).$$

Решение этой системы в стационарном случае не представляет особого труда.

§ 3.6. Графы вероятностей переходов цепей Маркова

Удобным способом задания марковских СМО являются графы вероятностей переходов состояний таких систем.

Для марковских СМО графы вероятностей переходов строятся совершенно аналогично построению таких графов для цепей Маркова с непрерывным временем.

Вершина графа обозначает состояние системы. Ребра графа ориентированы и показывают возможные переходы из одного состояния в другое. В графе рисуют лишь те ребра, которые показывают переходы с ненулевыми инфинитезимальными характеристиками. Эти характеристики обычно пишут рядом с ребрами и называют весами ребер.

Удобство такого способа описания СМО заключается в его наглядности и возможности реализации простого правила построения системы дифференциальных уравнений Колмогорова для вероятностей состояний СМО.

Это правило имеет следующий вид: производная по времени от вероятности состояния в момент времени t равна сумме произведений вероятностей состояний на веса ребер, входящих в данное состояние (как будто вероятности **втекают** в данное состояние), минус произведение вероятности рассматриваемого состояния на сумму весов всех ребер, выходящих из него (как будто вероятность **вытекает** из рассматриваемого состояния).

Приведем несколько примеров таких графов для простейших марковских СМО и соответствующих систем дифференциальных уравнений, определяющих распределение вероятностей состояний этих систем.

1) $M\,|\,M\,|\,1\,|\,\infty$

Рассматриваемая СМО состоит из одного обслуживающего прибора (одной линии), на который поступает простейший поток заявок (требований) с интенсивностью λ. Если в системе нет заявок, то поступившая заявка занимает прибор для своего обслуживания, продолжительность которого случайная, распределенная по экспоненциальному закону с параметром μ. Если в момент поступления заявки прибор занят, то поступившая заявка становится в очередь (поступает в бункер). Считается, что длина очереди (объем бункера) может быть сколь угодно велика. В момент освобождения прибора из очереди по какому-то правилу выбирается следующая заявка для обслуживания.

Пусть $i(t)$ есть число заявок, находящихся в системе в момент времени t. В силу свойств простейшего потока и экспоненциального обслуживания процесс $i(t)$ является цепью Маркова с непрерывным временем. Граф вероятностей переходов для процесса $i(t)$ изображен на рис. 3.3.

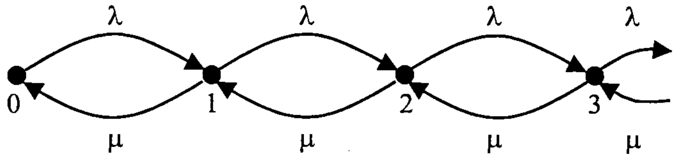


Рис. 3.3. Граф переходов в СМО $M \mid M \mid 1 \mid \infty$

$$P'(i,t) = \lambda P(i-1,t) + \mu P(i+1,t) - (\lambda + \mu) P(i,t).$$

2) $M \mid M \mid \infty$

Описание этой системы приведено в § 3.2. Граф вероятностей переходов для процесса $i(t)$ изображен на рис. 3.4.

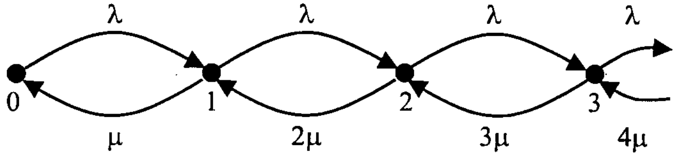


Рис. 3.4. Граф переходов в СМО $M \mid M \mid \infty$

$$P'(i,t) = \lambda P(i-1,t) + (i+1)\mu P(i+1) - (\lambda + i\mu) P(i,t).$$

3) ***СМО с источником повторных вызовов***

Описание этой системы приведено в § 3.5. Граф вероятностей переходов для процесса $i(t)$ изображен на рис. 3.5.

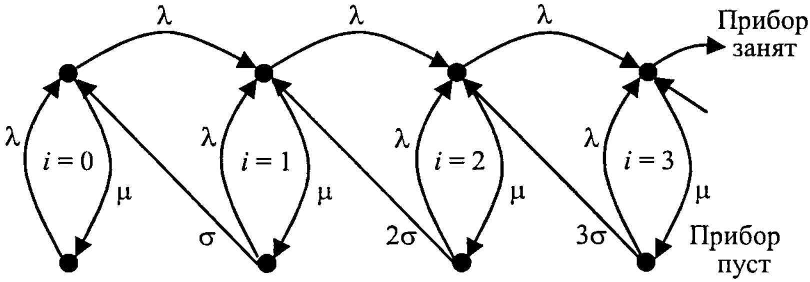


Рис. 3.5. Граф переходов в СМО с источником повторных вызовов

$$P_0'(i,t) = \mu P_1(i,t) - (\lambda + i\sigma)P_0(i,t)$$

$$P_1'(i,t) = \lambda P_1(i-1,t) + (i+1)\sigma P_0(i+1,t) + \lambda P_0(i,t) - (\lambda + \mu)P_1(i,t).$$

§ 3.7. Эргодичность цепей Маркова

Существование стационарного распределения вероятностей состояний СМО определяется эргодическими свойствами соответствующей цепи Маркова. Приведем основные результаты из эргодической теории цепей Маркова.

Эргодические свойства цепей Маркова с непрерывным временем полностью определяются эргодическими свойствами вложенных цепей Маркова с дискретным временем.

Определение 1. Состояние i называется **несущественным**, если существует такое состояние j, в которое система может перейти за конечное число шагов n, но не может вернуться в i-е ни за какое число шагов, то есть

$$p_{ij}(n) > 0 \quad p_{ji}(m) = 0 \quad \forall m.$$

Все остальные состояния существенные.

Определение 2. Состояния i и j называются **сообщающимися** ($i \leftrightarrow j$), если существуют n и k такие, что

$$p_{ij}(n) > 0 \quad p_{ji}(k) > 0.$$

Определение 3. Все существенные состояния можно разбить на классы, которые состоят из сообщающихся состояний, и ни из одного состояния данного класса нельзя перейти в состояние другого класса. Такие классы называют **неразложимыми** или **замкнутыми**.

Рассмотрим неразложимый класс S.

Тогда для $\forall i \in S \quad \exists n$, что $p_{ii}(n) > 0$. Пусть M_i – множества числа шагов n, для которых $P_{ii}(n) > 0$.

Определение 4. Наибольший общий делитель d_i этих чисел называется периодом состояния i.

Теорема солидарности. Все состояния одного неразложимого класса имеют одинаковый период d.

Определение 5. Если $d = 1$, то класс называется **непериодическим,** или **эргодическим**.

Введем вероятность

$$f_i(n) = P(\xi(n) = i \mid \xi(0) = i, \xi(1) \neq i, \xi(2) \neq i, .., \xi(n-1) \neq i)$$

того, что система впервые возвращается в i-е состояние на n-м шаге.

Тогда вероятность

$$f_i^* = \sum_{n=1}^{\infty} f_i(n)$$

можно рассматривать как вероятность того, что система, выйдя из i-го состояния, хотя бы один раз вернется в него.

Определение 6. Состояние i называется **возвратным**, если $f_i^* = 1$, и **невозвратным**, если $f_i^* < 1$.

Теорема солидарности. Если имеется два сообщающихся состояния и одно из них возвратно, то второе возвратно также.

Для возвратного состояния i вероятности $f_i(n)$ образуют вероятностное распределение времени возвращения в i-е состояние.

Определение 7. Если i – возвратное состояние и $\sum_{n=1}^{\infty} n f_i(n) = \infty$, то состояние i называется **нулевым**, если $\sum_{n=1}^{\infty} n f_i(n) < \infty$, то возвратное состояние называется ненулевым или **положительным**.

Основная эргодическая теорема для марковских цепей

Теорема. Рассмотрим возвратную, неприводимую, непериодическую марковскую цепь, тогда имеет место равенство

$$\lim_{n\to\infty} P_{ji}(n) = \frac{1}{\sum_{n=1}^{\infty} n f_i(n)} \begin{cases} = 0, & \text{если } i \quad \text{нулевое состояние}, \\ > 0, & \text{если } i \text{ положительное состояние}. \end{cases}$$

При этих же условиях

$$\lim_{n\to\infty} P_{ji}(n) = \lim_{n\to\infty} P_{ii}(n).$$

Теорема альтернативы. Пусть для марковской цепи со счетным числом состояний и матрицей переходных вероятностей P_{ij} существуют пределы

$$\lim_{n\to\infty} P_{ij}(n) = \pi_j \quad \forall j.$$

Тогда

$$\sum_i \pi_i P_{ij} = \pi_j, \text{ и либо все } \pi_i = 0, \text{ либо } \sum_i \pi_i = 1.$$

Если $\pi_i = 0$, то стационарное распределение не существует; если $\sum_i \pi_i = 1$, то π_i называются финальными вероятностями, а их набор образует эргодическое распределение, которое совпадает с единственным стационарным распределением.

Таким образом, для того, чтобы существовало стационарное распределение вероятностей, необходимо и достаточно, чтобы цепь Маркова была

1) неразложима;
2) непериодична;
3) возвратна;
4) положительна.

Для проверки условий возвратности и положительности сформулируем две конструктивные теоремы.

Теорема Фостера (эргодическая теорема Фостера)

Для того, чтобы неприводимая, непериодическая цепь Маркова была строго эргодической, необходимо и достаточно, чтобы система уравнений

$$x(j) = \sum_{i\in X} x(i) P_{ij}, \quad j \in X,$$

имела нетривиальное решение $x(i)$ $i \in X$ такое, что $\sum_{i\in X} |x(i)| < \infty$ $i \in X$.

При этом существует единственное стационарное распределение, которое совпадает с эргодическим.

Теорема Мустафы (эргодическая теорема Мустафы)

Для того, чтобы неприводимая, непериодическая цепь Маркова была эргодической, достаточно существование $\varepsilon > 0$, натурального числа i_0 и набора неотрицательных чисел x_0, x_1, x_2, ..., таких, что

$$\sum_{j \in X} P_{ij} x_j \le x_i - \varepsilon \quad \text{для} \quad i > i_0,$$

$$\sum_{j \in X} P_{ij} x_j < \infty \quad \text{для} \quad i \le i_0.$$

При этом существует единственное эргодическое распределение, которое совпадает со стационарным.

Теорему Мустафы можно сформулировать и по-другому.

Теорема. Для того, чтобы неприводимая непериодическая цепь Маркова была эргодической, достаточно существование $\varepsilon > 0$, натурального числа i_0 и неотрицательной функции $f(i)$ такой, что

$$M\{f(\xi_{n+1}) \mid \xi_n = i\} \le f(i) - \varepsilon \quad \text{для} \quad i > i_0,$$

$$M\{f(\xi_{n+1}) \mid \xi_n = i\} \le \infty \quad \text{для} \quad i \le i_0.$$

При этом существует единственное эргодическое распределение, которое совпадает со стационарным.

§ 3.8. Стационарный режим в системе $M \mid M \mid 1 \mid \infty$

В СМО при определенных условиях существуют так называемые финальные вероятности, то есть $\lim\limits_{t \to \infty} P_i(t) = \pi_i$. Эти распределения π_i для однородной СМО обладают свойством стационарности, то есть они не меняются со временем. Стационарный режим представляет наибольший интерес при изучении СМО, и обычно, анализируя какую-то СМО, как раз и стремятся найти эти финальные вероятности. Как же их вычислять?

Наиболее строгим путем является получение сначала $P_i(t)$, а затем нахождение $\lim\limits_{t \to \infty} P_i(t) = \pi(i)$, если этот предел существует.

Рассмотрим систему $M \mid M \mid 1 \mid \infty$, то есть однолинейную систему, на вход которой поступает простейший с интенсивностью λ поток заявок, время обслуживания экспоненциальное с параметром μ. В системе имеется бункер с неограниченным числом мест для ожидания (рис. 3.6).

Пусть $i(t)$ – число заявок в системе в момент времени t. Обозначим $P_i(t) = P(i(t) = i)$. Так как

$$P_0(t+\Delta t) = (1-\lambda\Delta t)P_0(t) + \mu\Delta t P_1(t) + o(\Delta t),$$

$$P_i(t+\Delta t) = \lambda\Delta t P_{i-1}(t) + (1-(\lambda+\mu)\Delta t)P_i(t) + \mu\Delta t P_{i+1}(t) + o(\Delta t),$$

то, выполнив несложные преобразования, получим систему дифференциальных уравнений

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t),$$

$$P_i'(t) = \lambda P_{i-1}(t) - (\lambda+\mu)P_i(t) + \mu P_{i+1}(t), \quad i \ge 1, \qquad (3.10)$$

определяющую распределение вероятностей $P_i(t)$ числа заявок в рассматриваемой СМО.

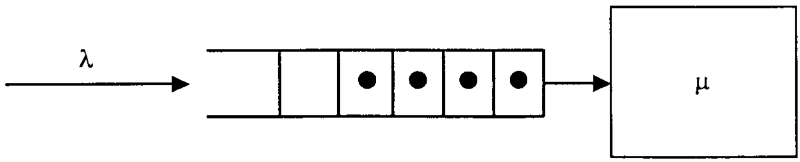


Рис. 3.6. Блок-схема однолинейной СМО с бесконечным бункером

Заметим, что решение системы (3.10) не выражается в элементарных функциях и найти его достаточно сложно.

Однако указанный выше способ нахождения финального распределения хоть и является строгим, но очень громоздок, так как требует предварительного нахождения $P_i(t)$ или их преобразований Лапласа. Поэтому поступают обычно следующим образом: переходят к пределу $t \to \infty$ не в решении, а в самих уравнениях, определяющих $P_i(t)$, считая $\lim\limits_{t\to\infty} P_i'(t) = 0$. По-хорошему, прежде чем делать такой предельный переход, следовало бы доказать существование $\lim\limits_{t\to\infty} P_i(t) = \pi_i$ и тем более то, что $\lim\limits_{t\to\infty} P_i'(t) = 0$. Однако это очень сложно, и обычно это не делают, а о существовании $\lim\limits_{t\to\infty} P_i(t)$ судят исходя из решения.

Продемонстрируем это на системе $M \mid M \mid 1 \mid \infty$. Делая в уравнениях (3.10) предельный переход $t \to \infty$, получим систему

$$\begin{cases} -\lambda\pi_0 + \mu\pi_1 = 0, \\ \lambda\pi_{i-1} - (\lambda+\mu)\pi_i + \mu\pi_{i+1} = 0, \quad i \ge 1. \end{cases} \tag{3.11}$$

Решим систему (3.11). Второе уравнение этой системы является однородным уравнением в конечных разностях второго порядка. Его характеристическое уравнение

$$\mu z^2 - (\lambda+\mu)z + \lambda = 0$$

имеет корни $z_1 = 1$ и $z_2 = \dfrac{\lambda}{\mu} = \rho$, так что общее решение имеет вид

$$\pi_i = C_1 + C_2\rho^i .$$

Первое уравнение (3.11) имеет смысл «граничного условия». Подставляя это решение, получим

$$-\lambda(C_1 + C_2) + \mu(C_1 + C_2\rho) = 0,$$

откуда получится, что $C_1 = 0$ и $\pi_i = C_2\rho^i$.

При решении систем уравнений, определяющих финальные вероятности, следует иметь в виду, что одна из констант всегда остается неопределенной; она находится из условия нормировки $\sum_{i=0}^{\infty}\pi_i = 1$. Это условие нормировки следует всегда добавлять к системе уравнений, определяющих финальные вероятности.

В нашем случае условие нормировки имеет вид

$$C_2\sum_{i=0}^{\infty}\rho^i = 1 .$$

Так как входящий сюда ряд сходится при $\rho < 1$, то и стационарный режим существует лишь при $\rho < 1$. В этом случае $C_2 = 1-\rho$ и

$$\pi_i = (1-\rho)\rho^i . \tag{3.12}$$

Финальные вероятности позволяют найти ряд важных характеристик СМО.

А. Среднее число заявок в системе

$$M\{\ i\ \} = \sum_{i=0}^{\infty} i\pi_i = (1-\rho)\sum_{i=0}^{\infty} i\rho^i = (1-\rho)\rho\sum_{i=0}^{\infty} i\rho^{i-1} = (1-\rho)\rho\left(\sum_{i=0}^{\infty}\rho^i\right)' = \frac{\rho}{1-\rho} .$$

Б. Средняя длина очереди

Так как длина очереди равна i–1 и очередь есть лишь при $i > 1$, то

$$M\{n\}=\bar{n}=\sum_{i=2}^{\infty}(i-1)\pi_i=(1-\rho)\sum_{i=1}^{\infty}(i-1)\rho^i=(1-\rho)\rho^2\sum_{s=0}^{\infty}s\rho^{s-1}=\frac{\rho^2}{1-\rho}.$$

В. Дисперсия числа заявок

$$M\{i(i-1)\}=\sum_{i=2}^{\infty}i(i-1)\pi_i=(1-\rho)\rho^2\sum_{i=2}^{\infty}i(i-1)\rho^{i-2}=(1-\rho)\rho^2\left(\sum_{i=0}^{\infty}\rho^i\right)''=\frac{2\rho^2}{(1-\rho)^2}.$$

Отсюда

$$D\{i\}=M\{i^2\}-M\{i\}^2=M\{i(i-1)\}+M\{i\}-M\{i^2\}=$$

$$=\frac{2\rho^2}{(1-\rho)^2}+\frac{\rho}{1-\rho}-\frac{\rho^2}{(1-\rho)^2}=\frac{\rho}{(1-\rho)^2}.$$

Г. Дисперсия длины очереди (Найдите сами).

§ 3.9. Виртуальное время ожидания $M\,|\,M\,|\,1\,|\,\infty$ (процедура FIFO)

Представим себе, что в некоторый произвольный момент времени t в СМО поступает заявка, которую будем называть виртуальной, так как реальная заявка может и не поступить в систему в момент t. Время, прошедшее между поступлением виртуальной заявки в систему и моментом начала ее обслуживания, называется виртуальным временем ожидания. Мы будем обозначать его через ξ.

Разумеется, ξ – случайная величина. Обычно интересуются такими параметрами, как $M\{\xi\}=\tau_{\text{ож}}$, $D\{\xi\}$, $p_\xi(\tau)$. Несмотря на то, что в системе $M|M|1|\infty$ все эти величины можно вычислить очень просто, мы изложим ниже более общий, хотя и более сложный способ вычисления этих величин. Этот метод также основан на графах и свойствах марковских процессов.

Пусть в момент времени t_1, когда заявка поступила в систему, там уже находилось i заявок. Так как те заявки, которые поступят после момента t_1, уже не влияют на интересующее нас время ожидания, то их можно не учитывать. Надо учитывать лишь окончание обслуживания заявок, стоящих в очереди впереди находящихся в системе. Это происходит с интенсивностью μ.

Определим соответствующий марковский процесс $j(t)$ как число оставшихся (не обслуженных) к моменту t заявок из тех i, которые были в системе в момент времени t_1 поступления виртуальной заявки.

Процесс $j(t)$ является процессом чистой гибели. Его граф переходов имеет вид, изображенный на рис. 3.7.

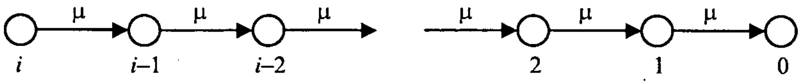


Рис. 3.7

Найдем время ξ_j перехода построенного процесса $j(t)$ из состояния j в состояние 0. Обозначим $M\{\xi_j \mid j\}$ через m_j, здесь $0 \le j \le i$. Нас интересует величина m_i.

Далее рассуждаем, как в Δt-методе. Пусть прошло время Δt. Если за это время не произошло освобождение прибора (это будет с вероятностью $1 - \mu\Delta t + o(\Delta t)$), то наше время ожидания составит $\xi_j = \Delta t + \xi'_j$. Основным является то, что так как спустя Δt мы снова окажемся в состоянии j, то ξ'_j распределено так же, как и ξ_j. Если же за время Δt произойдет освобождение прибора (это будет с вероятностью $\mu\Delta t + o(\Delta t)$), то время ожидания составит $\Delta t + \xi'_{j-1}$, так как мы перейдем в состояние $j-1$. Поэтому, усредняя по ξ, мы получим

$$M\{\xi_j \mid j\} = m_j = (1 - \mu\Delta t)\left(\Delta t + M\{\xi'_j\}\right) + \mu\Delta t\left(\Delta t + M\{\xi'_{j-1}\}\right) + o(\Delta t) =$$

$$= (1 - \mu\Delta t)(\Delta t + m_j) + \mu\Delta t(\Delta t + m_{j-1}) + o(\Delta t).$$

Как обычно, раскрывая скобки, получим

$$m_j = m_j + \Delta t - \mu m_j \Delta t + \mu m_{j-1}\Delta t + o(\Delta t).$$

Уничтожая m_j, деля на Δt и переходя к пределу $\Delta t \to 0$, получим

$$m_i - m_{i-1} = \frac{1}{\mu}. \qquad (3.13)$$

Сюда надо добавить естественное граничное условие $m_0 = 0$.

Уравнение (3.13) есть линейное неоднородное уравнение первого порядка; корень его характеристического уравнения равен 1. Поэтому его частное решение следует искать в виде $m_j = A \cdot j$. Подставляя это решение в (3.13), получим

$$A \cdot j - A(j-1) = \frac{1}{\mu},$$

откуда следует, что $A = \frac{1}{\mu}$, и общее решение (3.13) имеет вид

$$m_j = \frac{j}{\mu} + C.$$

Граничное условие $m_0 = 0$ дает $C = 0$, так что окончательно $m_j = j/\mu$, $0 \le j \le i$.

Следовательно, $m_i = i/\mu$.

Осталось усреднить по i. В стационарном режиме приходящая заявка застает в СМО i заявок с вероятностью π_i. Поэтому

$$M\{\xi\} = \tau_{\text{ож}} = \sum_{i=0}^{\infty} m_i \pi_i = \frac{1}{\mu} M\{i\} = \frac{1}{\mu}\frac{\rho}{1-\rho}. \tag{3.14}$$

Для вычисления дисперсии найдем $M\{\xi_i^2 \mid i\} = m_2(i)$. Аналогично тому, как выводилось уравнение (3.13), теперь для всех $0 \le j \le i$ можно записать

$$m_2(j) = M\{\xi_j^2 \mid j\} = (1-\mu\Delta t)M\{(\Delta t + \xi'_j)^2\} + \mu\Delta t M\{(\Delta t + \xi'_{j-1})^2\} + o(\Delta t).$$

Раскрывая скобки и усредняя по ξ, получим

$$m_2(j) = (1-\mu\Delta t)\left[m_2(j) + 2\Delta t m_j + \Delta t^2\right] +$$
$$+\mu\Delta t\left[m_2(j-1) + 2\Delta t m_{j-1} + \Delta t^2\right] + o(\Delta t).$$

Отсюда

$$m_2(j) = m_2(j) - m_2(j)\mu\Delta t + 2m_j\Delta t + \mu m_2(j-1)\Delta t + o(\Delta t),$$

и мы обычным способом получаем уравнение относительно $m_2(j)$:

$$m_2(j) - m_2(j-1) = \frac{2}{\mu} m_j.$$

Но m_j нам уже известно – оно равно $m_j = \frac{j}{\mu}$, так что

$$m_2(j)-m_2(j-1)=\frac{2j}{\mu^2} \qquad (3.15)$$

с очевидным граничным условием $m_2(0)=0$.

Ищем решение этого уравнения в виде $m_2(j)=A\cdot j+B\cdot j^2$ (при этом условие $m_2(0)=0$ выполняется автоматически). Подставляя это решение в (3.15), получим

$$A\cdot j+B\cdot j^2-A(j-1)-B(j-1)^2=\frac{2j}{\mu^2}=2B\cdot j+A-B.$$

Отсюда, приравнивая коэффициенты при одинаковых степенях j, получим

$$2B=\frac{2}{\mu^2}, \qquad A-B=0,$$

откуда для всех $0\le j\le i$ получим

$$m_2(j)=\frac{j^2+j}{\mu^2}.$$

Следовательно, при $j=i$ имеем

$$m_2(i)=\frac{i^2+i}{\mu^2}. \qquad (3.16)$$

Усредняя по i, получим окончательно

$$M\left\{\ \xi^2\right\}=\sum_{i=0}^{\infty}m_2(i)\pi_i=\frac{1}{\mu^2}\left[M\left\{\ i^2\right\}+M\left\{\ i\right\}\right]=$$

$$=\frac{1}{\mu^2}\left[M\left\{\ i(i-1)\right\}+2M\left\{\ i\right\}\right]=\frac{2\rho}{\mu^2(1-\rho)^2},$$

$$D\{\xi\}=M\left\{\xi^2\right\}-M\left\{\xi\right\}^2=\frac{\rho(2-\rho)}{(1-\rho)^2}. \qquad (3.17)$$

Вычислим, наконец, плотность вероятностей времени ожидания ξ, точнее, преобразования Лапласа от $p_\xi(x)$:

$$g_\xi(s)=M\left\{\ e^{-s\xi}\right\}=\int_0^{\infty}e^{-sx}p_\xi(x)dx.$$

Для вычисления $g_\xi(s)$ приходится вводить вспомогательные функции $g_i(s)=M\left\{e^{-s\xi_i}\right\}$, где ξ_i означает время ожидания, если в начальный момент поступления заявки в системе уже находилось i заявок. Тогда рассуждения, аналогичные тем, что применялись при выводе уравнения (3.13) для всех $0\le j\le i$, дают

$$g_j(s)=M\left\{e^{-s\xi_j}\right\}=(1-\mu\Delta t)M\left\{e^{-s(\Delta t+\xi'_j)}\right\}+\mu\Delta t M\left\{e^{-s(\Delta t+\xi'_{j-1})}\right\}=$$

$$=(1-\mu\Delta t)e^{-s\Delta t}\cdot g_j(s)+\mu\Delta t\cdot e^{-s\Delta t}\cdot g_{j-1}(s)+o(\Delta t).$$

Так как $e^{-s\Delta t}=1-s\Delta t+o(\Delta t)$, то

$$g_j(s)=(1-\mu\Delta t)(1-s\Delta t)g_j(s)+\mu\Delta t(1-s\Delta t)g_{j-1}(s)+o(\Delta t)=$$

$$=g_j(s)-(\mu+s)\Delta t g_j(s)+\mu\Delta t g_{j-1}(s)+o(\Delta t).$$

Отсюда обычным путем получаем

$$(\mu+s)g_j(s)-\mu g_{j-1}(s)=0.$$

Так как $\xi_0=0$ с вероятностью 1, то $g_0(s)=1$. Отсюда легко получить, что для всех $0\le j\le i$

$$g_j(s)=\left(\frac{\mu}{\mu+s}\right)^j.$$

Следовательно $g_i(s)=\left(\dfrac{\mu}{\mu+s}\right)^i$.

Для получения $g_\xi(s)$ надо $g_i(s)$ усреднить по i

$$g_\xi(s)=\sum_{i=0}^{\infty}\left(\frac{\mu}{\mu+s}\right)^i\pi_i=(1-\rho)\sum_{i=0}^{\infty}\left(\frac{\mu}{\mu+s}\right)^i\rho^i=\frac{1-\rho}{1-\dfrac{\mu\rho}{\mu+s}}=\frac{(1-\rho)(\mu+s)}{s+\mu(1-\rho)}.$$

Разлагая это дробно рациональное выражение на простейшие дроби, получим

$$g_\xi(s)=1-\rho+\rho\frac{(1-\rho)\mu}{s+\mu(1-\rho)},$$

и, используя обратное преобразование Лапласа – Стилтьеса, будем иметь функцию распределения виртуального времени ожидания в виде

$$F_\xi(x) = 1 - \rho e^{-\mu(1-\rho)x}. \tag{3.18}$$

Отсюда можно еще раз найти $M\{\xi\}$ и $D\{\xi\}$

$$M\{\xi\} = \rho\int_0^\infty \mu(1-\rho)xe^{-\mu(1-\rho)x}dx = \frac{\rho}{\mu(1-\rho)},$$

$$M\{\xi^2\} = \rho\int_0^\infty \mu(1-\rho)x^2e^{-\mu(1-\rho)x}dx = \frac{2\rho}{\mu^2(1-\rho)^2},$$

что совпадает с тем, что было ранее.

§ 3.10. Виртуальное время ожидания $M \mid M \mid 1 \mid \infty$ (процедура LIFO)

Рассмотрим таким же образом статистические характеристики времени ожидания при процедуре обслуживания LIFO (инверсионный порядок обслуживания – «последним пришел – первым обслужен»). В этом случае значение имеют только те заявки, которые пришли **после** интересующей нас заявки. Обозначая через $j(t)$ число оставшихся к моменту t тех заявок, которые пришли в систему **после** интересующей нас заявки, поступившей в момент времени t_1, за время ее ожидания, мы получим следующий граф переходов для процесса $j(t)$, где состояние (–1) означает, что интересующая нас заявка берется на обслуживание (рис. 3.8).

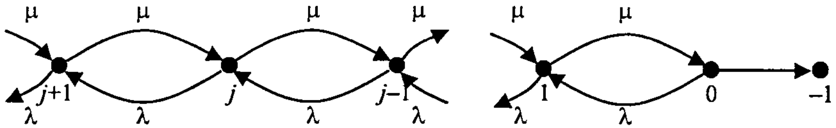


Рис. 3.8

Пусть ξ_j по-прежнему означает время перехода из состояния j в состояние (–1). Нас интересует ξ_0. В этом случае Δt-метод дает

$$\xi_j = \begin{cases} \Delta t + \xi_{j+1} & \text{с вероятностью } \lambda\Delta t, \\ \Delta t + \xi'_j & \text{с вероятностью } (1-\lambda\Delta t - \mu\Delta t), \\ \Delta t + \xi_{j-1} & \text{с вероятностью } \mu\Delta t. \end{cases}$$

Поэтому для $m_j = M\{\xi_j\}$ получаем

$$m_j = (1-\lambda\Delta t - \mu\Delta t)M\left\{\Delta t + \xi'_j\right\} + \lambda\Delta t M\left\{\Delta t + \xi'_{j+1}\right\} + \mu\Delta t M\left\{\Delta t + \xi'_{j-1}\right\} +$$

$$+o(\Delta t) = (1-\lambda\Delta t - \mu\Delta t)(\Delta t + m_j) + \lambda\Delta t(\Delta t + m_{j+1}) + \mu\Delta t(\Delta t + m_{j-1}) + o(\Delta t)\,.$$

Отсюда

$$\lambda m_{j+1} - (\lambda+\mu)m_j + \mu m_{j-1} = -1\,.$$

Для величины m_0 получается особое уравнение, так как переход в состояние (–1) означает начало обслуживания

$$m_0 = (1-\lambda\Delta t - \mu\Delta t)(\Delta t + m_0) + \lambda\Delta t(\Delta t + m_1) + \mu\Delta t\Delta t\,,$$

откуда

$$\lambda m_1 - (\lambda+\mu)m_0 = -1\,.$$

Формально это совпадает с предыдущим уравнением, если положить $m_{-1} = 0$. Таким образом, окончательно имеем

$$\begin{cases} \lambda m_{i+1} - (\lambda+\mu)m_i + \mu m_{i-1} = -1, \\ m_{-1} = 0. \end{cases} \qquad (3.19)$$

Так как характеристическое уравнение для (3.19) имеет вид

$$\lambda z^2 - (\lambda+\mu)z + \mu = 0\,,$$

то его корни $z_1 = 1$ и $z_2 = \dfrac{\mu}{\lambda} = \dfrac{1}{\rho}$. Следовательно, частное решение неоднородного уравнения (3.19) ищем в виде $m_i = A\cdot i$. Подстановка в уравнение дает

$$\lambda A(i+1) - (\lambda+\mu)A\cdot i + \mu A(i-1) = -1\,,$$

откуда $A = \dfrac{1}{\mu-\lambda}$, и поэтому

$$m_i = C_1 + \frac{C_2}{\rho^i} + \frac{i}{\mu-\lambda}\,.$$

Но слагаемое $\dfrac{C_2}{\rho^i}$ растет до бесконечности в геометрической прогрессии при $i\to\infty$, чего быть не может, поэтому следует взять $C_2 = 0$. Условие $m_{-1} = 0$ дает окончательно

$$m_i = \frac{i+1}{\mu-\lambda}.$$

При вычислении $M\{\xi\}$ надо принять во внимание то, что с вероятностью $\pi_0 = 1-\rho$ поступающая заявка сразу попадает в состояние (–1) (система пуста) и с вероятностью $1-\pi_0 = \rho$ она попадает в состояние 0. Поэтому

$$M\{\xi\} = \tau_{0m} = (1-\rho)m_{-1} + \rho m_0 = \frac{\rho}{\mu-\lambda} = \frac{\rho}{\mu(1-\rho)}. \qquad (3.20)$$

Интересно отметить, что среднее время ожидания в процедуре LIFO такое же, как и в процедуре FIFO. Аналогично тому, как это проделано выше, можно было бы вычислить $D\{\xi\}$. Но мы продемонстрируем вычисление другим путем.

Введем снова $g_j(s) = M\left\{ e^{-s\xi_j} \right\}$. Тогда Δt-метод дает

$$g_j(s) = \lambda\Delta t M\left\{e^{-s(\Delta t+\xi_{j+1})}\right\} + (1-\lambda\Delta t-\mu\Delta t)M\left\{e^{-s(\Delta t+\xi_j)}\right\} + \mu\Delta t M\left\{e^{-s(\Delta t+\xi_{j-1})}\right\} +$$

$$+ o(\Delta t) = \lambda\Delta t e^{-s\Delta t} g_{j+1}(s) + (1-\lambda\Delta t-\mu\Delta t)e^{-s\Delta t} g_j(s) + \mu\Delta t e^{-s\Delta t} g_{j-1}(s) + o(\Delta t).$$

Представляя $e^{-s\Delta t} = 1 - s\Delta t + o(\Delta t)$, получим после стандартных преобразований

$$\lambda g_{j+1}(s) - (\lambda+\mu+s)g_j(s) + \mu g_{j-1}(s) = 0, \qquad (3.21)$$

с граничным условием $g_{-1}(s) = 1$ (так как $\xi_{-1} = 0$ с вероятностью 1).

Характеристическое уравнение для (3.21)

$$\lambda z^2 - (\lambda+\mu+s)z + \mu = 0.$$

Его корни равны $\tilde{z}_1 = \frac{1}{z_2}$ и $\tilde{z}_2 = \frac{1}{z_1}$, где z_1 и z_2 определяются формулами

$$z_{1,2} = \frac{\lambda+\mu+s \pm \sqrt{(\lambda+\mu+s)^2 - 4\lambda\mu}}{2\mu},$$

поэтому

$$g_i(s) = \frac{C_1}{z_1^i} + \frac{C_2}{z_2^i}.$$

Но $|z_1|<1$, и поэтому надо брать $C_1=0$, чтобы выполнялось условие $g_i(s)\le 1$. Условие $g_i(0)=1$ дает окончательно $g_i(s)=1/z_2^{\,i+1}$.

Учитывая снова, что заявка попадает в состояние (–1) с вероятностью $1-\rho$ и в состояние 0 с вероятностью ρ, получаем

$$g_\xi(s)=1-\rho+\frac{\rho}{z_2}=1-\rho+\rho\frac{2\mu}{\lambda+\mu+s+\sqrt{(\lambda+\mu+s)^2-4\lambda\mu}}=$$

$$=1-\rho+\rho\frac{\lambda+\mu+s-\sqrt{(\lambda+\mu+s)^2-4\lambda\mu}}{2\lambda}. \qquad (3.22)$$

Используя таблицы обратного преобразования Лапласа, можно получить, что

$$p_\xi(x)=(1-\rho)\delta(x)+\rho\sqrt{\frac{\mu}{\lambda}}\frac{1}{x}I_1(2\sqrt{\lambda\mu}x)e^{-(\lambda+\mu)x}, \qquad (3.23)$$

где $I_1(\cdot)$ есть модифицированная функция Бесселя.

Но знание $p_\xi(x)$ совсем не обязательно для нахождения $M\{\xi\}$ и $D\{\xi\}$, их можно определить и непосредственно из формулы (3.22) для $g_\xi(s)$. Имеем

$$g'_\xi(s)=\frac{\rho}{2\lambda}\left[1-\frac{\lambda+\mu+s}{\sqrt{(\lambda+\mu+s)^2-4\lambda\mu}}\right],$$

откуда

$$M\{\xi\}=-g'_\xi(0)=-\frac{\rho}{2\lambda}\left[1-\frac{\lambda+\mu}{\mu-\lambda}\right]=\frac{\rho}{\mu-\lambda}=\frac{\rho}{\mu(1-\rho)}, \qquad (3.24)$$

что совпадает с (3.20). Далее

$$g''_\xi(s)=\frac{\rho}{2\lambda}\left[-\frac{1}{\sqrt{(\lambda+\mu+s)^2-4\lambda\mu}}+\frac{(\lambda+\mu+s)^2}{\left[\sqrt{(\lambda+\mu+s)^2-4\lambda\mu}\right]^{3/2}}\right],$$

откуда

$$M\{\xi^2\}=g''_\xi(0)=\frac{\rho}{2\lambda}\left[-\frac{1}{\mu-\lambda}+\frac{(\lambda+\mu)^2}{(\mu-\lambda)^3}\right]=\frac{2\rho}{\mu^2(1-\rho)^3},$$

и поэтому

$$D\{\xi\} = M\{\xi^2\} - M\{\xi\}^2 = \frac{\rho(2-\rho+\rho^2)}{\mu^2(1-\rho)^3}. \quad (3.25)$$

Легко видеть, что $D\{\xi\}$ по процедуре LIFO всегда больше $D\{\xi\}$ по процедуре FIFO.

§ 3.11. Задача Эрланга для системы $M \mid M \mid N \mid 0$

В заключение данной главы приведем решение классической задачи Эрланга. Рассмотрим N-линейную СМО, на вход которой поступает простейший с параметром λ поток заявок. Обслуживание каждым прибором экспоненциальное с параметром μ. Необходимо найти стационарное распределение $\pi(i)$ числа занятых приборов, если заявки, поступившие в систему, когда заняты все приборы, теряются.

Очевидно, финальные вероятности $\pi(i)$ удовлетворяют системе уравнений

$$\lambda\pi(0) = \mu\pi(1),$$

$$(\lambda + i\mu)\pi(i) = \lambda\pi(i-1) + (i+1)\mu\pi(i+1),$$

$$N\mu\pi(N) = \lambda\pi(N-1), \quad (3.26)$$

из которой нетрудно получить рекуррентные соотношения

$$\pi(1) = \rho\pi(0),$$

$$\pi(i+1) = \frac{\rho}{i+1}\pi(i),$$

$$\pi(N) = \frac{\rho}{N}\pi(N-1),$$

откуда следует, что $\pi(i)$ имеют вид

$$\pi(0) = 1 \Big/ \sum_{j=0}^{N} \frac{\rho^j}{j!},$$

$$\pi(i) = \frac{\rho^i}{i!} \Big/ \sum_{j=0}^{N} \frac{\rho^j}{j!},$$

$$\pi(N) = \frac{\rho^N}{N!} \Big/ \sum_{j=0}^{N} \frac{\rho^j}{j!}. \quad (3.27)$$

Заметим, что при $N = \infty$ (бесконечно линейная СМО) эти формулы принимают очень простой вид

$$\pi(i) = \frac{\rho^i}{i!} e^{-\rho}. \tag{3.28}$$

Полученные формулы (3.27) называются формулами Эрланга и получены в 1921 г. В то же время Эрланг ставил задачу и пытался найти аналогичное распределение вероятностей, если обслуживание неэкспоненциальное. Эту задачу решил Б.А. Севастьянов, опубликовав ее в работе «Эргодическая теорема для марковских процессов и ее применение к телефонным линиям» в журнале «Теория вероятностей и ее применение» еще в 1957 г.

В теории телетрафика этими формулами пользуются до сих пор, хотя в связи с повышенной компьютеризацией и передачей по телефонным каналам разнородной информации требуется дальнейшее обобщение этих классических результатов.

Такие обобщения мы рассмотрим в главе по немарковским системам массового обслуживания.

Глава 4. ПОЛУМАРКОВСКИЕ МОДЕЛИ МАССОВОГО ОБСЛУЖИВАНИЯ

§ 4.1. Основные идеи

Перейдем теперь к рассмотрению полумарковских однолинейных систем, в которых одно из двух – либо длины интервалов между моментами поступления заявок потока, либо время обслуживания имеет произвольное распределение, то есть к рассмотрению систем типа $M \mid G$ или $GI \mid M$. Это СМО с рекуррентным потоком или рекуррентным обслуживанием. Основной трудностью, возникающей здесь, является то, что процесс $i(t)$, где i – число требований в системе в **произвольный** момент времени t, перестает быть марковским процессом, так как либо поток заявок, либо время обслуживания теперь обладает последействием. Однако такой случайный процесс можно свести к марковскому процессу, и поэтому его называют полумарковским.

Для преодоления этой трудности используют в основном два приема: метод дополнительной переменной или метод вложенных цепей Маркова. Наиболее популярным является так называемый метод вложенных цепей Маркова, основанный на двух идеях:

1. СМО рассматривается не в произвольный момент времени t, а в некоторые **специальным образом подобранные** моменты времени t_1, t_2, t_3,...

2. Эти моменты времени t_n подбираются так, чтобы процесс $i(t_n)$, рассматриваемый только в эти моменты времени, был марковским процессом – цепью Маркова с дискретным временем.

Таким образом, происходит возврат к марковским процессам, но – ценой отказа от рассмотрения системы в произвольный момент времени.

§ 4.2. Система $M \mid G \mid 1 \mid \infty$. Метод вложенных цепей Маркова

Рассмотрим СМО с одним обслуживающим прибором, на вход которой поступает простейший поток заявок интенсивности λ. Обслуживание произвольное с функцией распределения длительности обслуживания $B(x)$.

При применении метода вложенных цепей Маркова прежде всего надо выбрать особые моменты времени t_n. Так как обслуживание произвольное, то эти моменты должны быть связаны именно с обслуживанием. Возьмем поэтому в качестве рассматриваемых моментов времени **моменты** окончания обслуживания (можно взять и моменты начала обслуживания, результаты будут аналогичными), и пусть ν_n есть число заявок в СМО сразу после ухода n-й по счету обслуженной заявки (то есть сама ушедшая заявка не считается). Очевидно, что $\nu_n = 0,1,2,3,...$

Обозначим момент окончания обслуживания n-й заявки через t_n и через Δ_n – число заявок, поступивших в систему на интервале времени, равном длительности обслуживания n-й заявки. Тогда, если $\nu_{n-1} > 0$, то $\nu_n = \nu_{n-1} + \Delta_n - 1$, так как после окончания обслуживания немедленно (так как $\nu_{n-1} > 0$) началось обслуживание n-й заявки. За время ее обслуживания в систему пришло Δ_n заявок, и в момент окончания обслуживания n-й заявки она покинула систему. Но если $\nu_{n-1} = 0$, то $\nu_n = \Delta_n$ потому, что произошло следующее: в момент окончания обслуживания $n-1$-й заявки система оказалась пустой. Пришедшая в систему n-я заявка встала сразу же на обслуживание, то есть в системе оказалась одна заявка. Затем в течение времени ее обслуживание пришло еще Δ_n заявок; в момент окончания эта самая n-я заявка покинула систему и в ней осталось ровно Δ_n заявок. Итак,

$$\nu_n = \begin{cases} \nu_{n-1} + \Delta_n - 1, & \text{если} \quad \nu_{n-1} > 0, \\ \Delta_n \ , & \text{если} \quad \nu_{n-1} = 0. \end{cases}$$

Обозначим $P\{\Delta_n = k\} = f_k$. Если время обслуживания n-й заявки равно x, то

$$P\{\Delta_n = k \mid x\} = \frac{(\lambda x)^k}{k!} e^{-\lambda x} .$$

Но так как x есть значение случайной величины с функцией распределения $B(x)$, то, усредняя по x, получим

$$f_n = \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x} dB(x) .$$

Обратим внимание на то, что вероятности f_k вообще от ν_{n-1} не зависят. Поэтому распределение вероятностей величины ν_n зависит лишь от величины ν_{n-1} и ни от чего более. Это и говорит о том, что ν_n образуют марковский случайный процесс.

Найдем теперь переходные вероятности $p_{ij}\{\ \nu_n = j \mid \nu_{n-1} = i\}$. Если $\nu_{n-1} = 0$, то $\nu_n = \Delta_n$, и поэтому

$$p_{0j} = f_j, \quad j = 0,1,2,...$$

Если же $\nu_{n-1} = i > 0$, то при минимальном значении $\Delta_n = 0$ окажется, что $\nu_n = i-1$. Поэтому $p_{ij} = 0$, если $j < i-1$. В силу соотношения $\nu_n = \nu_{n-1} + \Delta_n - 1$ при $j \geq i-1$

$$p_{ij} = f_{j-i+1},$$

так как $\Delta_n = \nu_n - \nu_{n-1} + 1$. Окончательно получаем, что матрица переходных вероятностей имеет вид

$$P = \begin{bmatrix} f_0 & f_1 & f_2 & f_3 & \dots \\ f_0 & f_1 & f_2 & f_3 & \dots \\ 0 & f_0 & f_1 & f_2 & \dots \\ 0 & 0 & f_0 & f_1 & \dots \\ 0 & 0 & 0 & f_0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Отсюда видно, что:

1. Эта цепь неприводима, так как из любого состояния i можно попасть в состояние $i-1$, а из состояния 0 – в любое другое состояние.

2. Цепь непериодична, так как из состояния 0 в состояние 0 можно попасть за один шаг.

3. Используя эргодическую теорему Мустафы с функцией $f(i)$ вида $f(i) = i$, получим, что при $i > 0$

$$M\left\{\nu_{n+1} \mid \nu_n = i\right\} = M\left\{\nu_n + \Delta_{n+1} - 1 \mid \nu_n = i\right\} = i + M\left\{\Delta_{n+1}\right\} - 1,$$

но

$$M\left\{\Delta_{n+1}\right\} = \sum_{k=0}^{\infty} k \int_0^{\infty} \frac{(\lambda x)^k}{k!} e^{-\lambda x} dB(x) = \lambda \int_0^{\infty} x dB(x) = \lambda b,$$

где $b = \int_0^{\infty} x dB(x)$. Величина $\rho = \lambda b$ имеет смысл загрузки системы. Тогда

$$M\left\{\nu_{n+1} \mid \nu_n = i\right\} = i - (1 - \lambda b) \quad \text{при } i > 0,$$

$$M\left\{\nu_{n+1} \mid \nu_n = 0\right\} = M\left\{\Delta_n\right\} = \lambda b < +\infty,$$

откуда следует, что при $\lambda b < 1$ в системе существует стационарный режим.

Найдем теперь финальные вероятности π_k из системы

$$\begin{cases} \pi_k = \sum_{j=0}^{\infty} \pi_j p_{jk}, \ k = 0,1,2,\dots, \\ \sum_{k=0}^{\infty} \pi_k = 1. \end{cases}$$

Основное – первое уравнение этой системы имеет вид

$$\pi_k = \pi_0 f_k + \pi_1 f_k + \pi_2 f_{k-1} + \pi_3 f_{k-2} + \dots + \pi_{k+1} f_0 = \pi_0 f_k + \sum_{j=0}^{k} \pi_{k-j+1} f_j, \quad (4.1)$$

т.к. в матрице P суммирование идет по элементам k-го столбца, а они равны $f_k, f_k, f_{k-1}, \dots, f_0$.

Введем производящие функции

$$f(z) = \sum_{k=0}^{\infty} f_k z^k, \qquad \pi(z) = \sum_{k=0}^{\infty} \pi_k z^k.$$

Тогда, умножая (4.1) на z^k и складывая все уравнения системы, получим

$$\pi(z) = \sum_{k=0}^{\infty} \pi_k z^k = \pi_0 f(z) + \sum_{k=0}^{\infty} z^k \sum_{j=0}^{k} \pi_{k-j+1} f_j, \quad (4.2)$$

но

$$\sum_{k=0}^{\infty} z^k \sum_{j=0}^{k} \pi_{k-j+1} f_j = \sum_{j=0}^{\infty} f_j \sum_{k=j}^{\infty} \pi_{k-j+1} z^k = \sum_{j=0}^{\infty} f_j z^j \sum_{k=j}^{\infty} \pi_{k-j+1} z^{k-j} =$$

$$= \sum_{j=0}^{\infty} f_j z^j \left(\sum_{r=0}^{\infty} \pi_{r+1} z^r \right) = f(z) \frac{1}{z} \left[\pi(z) - \pi_0 \right].$$

Поэтому уравнение (4.2) примет вид

$$\pi(z) = f(z) \frac{\pi(z) - \pi_0}{z} + \pi_0 f(z),$$

откуда

$$\pi(z) = \pi_0 \frac{f(z)(1-z)}{f(z) - z}. \tag{4.3}$$

Осталось найти π_0. Для этого воспользуемся условием нормировки

$$\lim_{z \to 1} \pi(z) = \sum_{k=0}^{\infty} \pi_k = 1.$$

Выполняя в (4.3) предельный переход и пользуясь правилом Лопиталя, получим

$$\lim_{z \to 1} \pi(z) = 1 = \pi_0 \frac{-1}{f'(z)\big|_{z=1} - 1},$$

так как $f(1)$ в силу условия нормировки также равно 1. Отсюда $\pi_0 = 1 - f'(1)$ и

$$\pi(z) = \left[1 - f'(1)\right] \frac{f(z)(1-z)}{f(z) - z}.$$

Осталось получить явное выражение для $f(z)$. Обозначая через $B^*(s)$ преобразование Лапласа – Стилтьеса от $B(x)$

$$B^*(s) = \int_0^{\infty} e^{-sx} dB(x),$$

получим

$$f(z) = \sum_{k=0}^{\infty} z^k \int_0^{\infty} \frac{(\lambda x)^k}{k!} e^{-\lambda x} dB(x) = \int_0^{\infty} e^{-\lambda x} e^{\lambda x z} dB(x) = \int_0^{\infty} e^{-\lambda(1-z)x} dB(x) = B^*\left(\lambda(1-z)\right).$$

Тогда

$$f'(1)=\int_0^\infty \lambda x e^{-\lambda(1-z)x}dB(x)\Big|_{z=1}=\lambda\int_0^\infty x dB(x)=\lambda b .$$

Поэтому окончательно

$$\pi(z)=\frac{(1-\lambda b)B^*\left(\lambda(1-z)\right)(1-z)}{B^*\left(\lambda(1-z)\right)-z}. \tag{4.4}$$

Эта формула называется формулой Полачека – Хинчина. Она, по крайней мере в принципе, позволяет вычислить стационарное распределение вероятностей числа заявок в системе в моменты t_n+0.

§ 4.3. Распределение числа заявок в системе $M\,|\,G\,|\,1\,|\,\infty$ в произвольный момент времени. Метод дополнительной переменной

Формула Полачека – Хинчина определяет распределение числа заявок в СМО в моменты окончания обслуживания заявок. А каково распределение вероятностей для произвольных моментов времени?

Обозначим

$$\pi(k)=\lim_{t\to\infty}P\left(i(t)=k\right).$$

Для того чтобы найти $\pi(k)$, введем дополнительную переменную, рассмотрев **двумерный** процесс $\{i(t),z(t)\}$, где $z(t)$ – длина интервала от момента t до момента окончания обслуживания заявки, стоящей на приборе в момент t, то есть компонента $z(t)$ определяется только в те моменты времени, когда $i(t)>0$. Если $i(t)=0$, компонента $z(t)$ не определяется.

Пусть

$$P_0(t)=P\left(i(t)=0\right),\quad P_i(z,t)=P\left(i(t)=i,z(t)<z\right),$$

тогда в стационарном режиме

$$P_0(t)=\pi(0),\quad P_i(z,t)=P_i(z),\quad P_i(\infty)=\pi(i).$$

Так как двумерный процесс $\{i(t),z(t)\}$ марковский, то для распределения вероятностей $P_0(t)$, $P_i(z,t)$ можно составить систему уравнений

$$P_0(t+\Delta t)=(1-\lambda\Delta t)P_0(t)+P_1(\Delta t,t)+o(\Delta t),$$

$$P_1(z,t+\Delta t) = \lambda\Delta t P_0(t)B(z) + (1-\lambda\Delta t)\left[P_1(z+\Delta t,t) - P_1(\Delta t,t)\right] +$$
$$+P_2(\Delta t,t)B(z) + o(\Delta t),$$

$$P_i(z,t+\Delta t) = \lambda\Delta t P_{i-1}(z,t) + (1-\lambda\Delta t)\left[P_i(z+\Delta t,t) - P_i(\Delta t,t)\right] +$$
$$+P_{i+1}(\Delta t,t)B(z) + o(\Delta t).$$

Откуда в стационарном режиме для $P_0 = \pi(0)$ и $P_i(z)$ получим следующую систему уравнений:

$$\lambda P_0 = P_1'(0),$$

$$\lambda P_1(z) = P_1'(z) - P_1'(0) + \lambda P_0 B(z) + P_2'(0)B(z),$$

$$\lambda P_i(z) = P_i'(z) - P_i'(0) + \lambda P_{i-1}(z) + P_{i+1}'(0)B(z), \tag{4.5}$$

с краевым условием $P_i(0) = 0$, $\forall\ i \geq 1$.

При $z \to \infty$ из системы (4.5) получим

$$\lambda\pi(0) = P_1'(0),$$

$$\lambda\pi(1) = -P_1'(0) + \lambda\pi(0) + P_2'(0),$$

$$\lambda\pi(i) = -P_i'(0) + \lambda\pi(i-1) + P_{i+1}'(0),$$

откуда нетрудно получить, что

$$\lambda\pi(0) = P_1'(0),$$

$$\lambda\pi(1) = P_2'(0),$$

$$\lambda\pi(i) = P_{i+1}'(0).$$

Следовательно, систему (4.5) можно записать в виде

$$\lambda P_1(z) = P_1'(z) - \lambda\pi(0) + \lambda\pi(0)B(z) + \lambda\pi(1)B(z),$$

$$\lambda P_i(z) = P_i'(z) - \lambda\pi(i-1) + \lambda P_{i-1}(z) + \lambda\pi(i)B(z).$$

Пусть $\sum\limits_{i=1}^{\infty} x^i P_i(z) = F(x,z)$, $\sum\limits_{i=0}^{\infty} x^i \pi(i) = \Phi(x)$, тогда

$$\lambda F(x,z) = \frac{\partial F(x,z)}{\partial z} - \lambda x\Phi(x) + \lambda x\pi(0)B(z) + \lambda x F(x,z) + \lambda B(z)\left[\Phi(x) - \pi(0)\right],$$

и, следовательно,

$$\lambda(1-x)F(x,z) = \frac{\partial F(x,z)}{\partial z} - \lambda\left(x - B(z)\right)\Phi(x) - \lambda\pi(0)B(z)\left[1-x\right].$$

Решая полученное линейное неоднородное дифференциальное уравнение первого порядка относительно $F(x,z)$, получим

$$F(x,z)=e^{\lambda(1-x)z}\int\limits_0^z e^{-\lambda(1-x)s}\left\{\lambda\left(x-B(s)\right)\Phi(x)+\lambda\pi(0)B(s)\left[1-x\right]\right\}ds\,.$$

Так как существует $\lim\limits_{z\to\infty}F(x,z)=\Phi(x)-\pi(0)$ и первый сомножитель в правой части последнего равенства неограниченно возрастает, то при $z\to\infty$ имеет место равенство

$$\int\limits_0^\infty e^{-\lambda(1-x)s}\left\{\lambda\left(x-B(s)\right)\Phi(x)+\lambda\pi(0)B(s)\left[1-x\right]\right\}ds=0\,,$$

откуда можно записать

$$\Phi(x)\int\limits_0^\infty e^{-\lambda(1-x)s}\left(x-B(s)\right)ds=-\pi(0)\left[1-x\right]\int\limits_0^\infty e^{-\lambda(1-x)s}B(s)ds\,,$$

и, следовательно,

$$\Phi(x)=\frac{\left[1-x\right]\int\limits_0^\infty e^{-\lambda(1-x)s}B(s)ds}{\int\limits_0^\infty e^{-\lambda(1-x)s}\left(B(s)-x\right)ds}\pi(0)\,.$$

Обозначим $B^*(s)=\int\limits_0^\infty e^{-sx}dB(x)$, тогда

$$\int\limits_0^\infty e^{-sx}B(x)dx=-\frac{1}{s}\int\limits_0^\infty B(x)de^{-sx}=-\frac{1}{s}\left[B(x)e^{-sx}\Big|_0^\infty-\int\limits_0^\infty e^{-sx}dB(x)\right]=$$

$$=-\frac{1}{s}\int\limits_0^\infty e^{-sx}dB(x)=\frac{1}{s}B^*(s),$$

и, следовательно,

$$\Phi(x)=\frac{\frac{1-x}{\lambda(1-x)}B^*\left(\lambda(1-x)\right)}{\frac{1}{\lambda(1-x)}B^*\left(\lambda(1-x)\right)-x\frac{1}{\lambda(1-x)}}\pi(0)=\frac{(1-x)B^*\left(\lambda(1-x)\right)}{B^*\left(\lambda(1-x)\right)-x}\pi(0)\,.$$

Таким образом, производящая функция $\Phi(x)$ числа заявок $i(t)$ для произвольного момента времени t в стационарном режиме определяется равенством

$$\Phi(x)=\frac{(1-x)B^*\left(\lambda(1-x)\right)}{B^*\left(\lambda(1-x)\right)-x}\pi(0), \tag{4.6}$$

совпадающим с формулой (4.4) Полачека – Хинчина.

Следовательно, стационарное распределение числа заявок в системе в произвольный момент времени совпадает с распределением, даваемым формулой Полачека – Хинчина для числа заявок в системе в моменты времени t_n+0, непосредственно следующие за моментами окончания обслуживания.

§ 4.4. Период занятости в системе $M \mid G \mid 1 \mid\infty$

Периодом занятости СМО будем называть интервал от момента прихода заявки в пустую систему до момента ее следующего опустошения.

Пусть ξ есть длительность периода занятости и $F(x)=P\{\ \xi<x\ \}$.

Пусть СМО пуста. Представим себе, что в некоторый момент времени $t=0$ в систему поступила заявка, которая обслуживалась время y. Возможны следующие варианты.

А) За время y в систему других заявок не поступит. Вероятность этого события равна $e^{-\lambda y}$, и в этом случае $\xi=y$.

Б) За время y в систему поступит ровно одна заявка. Вероятность такого события равна $(\lambda y)e^{-\lambda y}$. Самое главное заключается в том, что после окончания обслуживания исходной заявки мы вновь оказываемся в той же ситуации, с которой начиналось исходное рассмотрение – то есть в СМО одна заявка и ее обслуживание только начинается. Поэтому в этой ситуации $\xi=y+\xi_1$ и (и это самое главное) ξ_1 распределено также, как и ξ, то есть $P\{\xi_1<x\}=F(x)$.

В) Пусть вообще за время обслуживания исходной заявки в систему поступило n заявок. Вероятность этого равна $(\lambda y)^n e^{-\lambda y}/n!$. Пусть они обслуживаются по процедуре LIFO (какая нам разница, прибор же все равно занят). Обозначим через $t_n, t_{n-1}, t_{n-2}, \ldots, t_1$ моменты начала обслуживания n-й, $n-1$-й,..., 1-й из пришедших заявок и через t_0 – момент освобождения системы.

Обозначим $t_{i-1}-t_i$ через ξ_i. Самое главное – это уяснить себе, что эти интервалы (то есть $t_{i-1}-t_i$) имеют точно такую же структуру, как и весь период занятости, то есть в этом случае

$$\xi = y+\xi_1+\xi_2+\ldots+\xi_n, \tag{4.7}$$

где ξ_i независимы и распределены по закону $P\{\xi_i<x\}=F(x)$.

Если это понять, то все дальнейшее очень просто. Из (4.7) с учетом вероятностей поступления n заявок записываем

$$F(x)=\sum_{n=0}^{\infty}\int_0^x \frac{(\lambda y)^n}{n!}e^{-\lambda y}F^{(n)}(x-y)dB(y),$$

где $F^{(n)}(\cdot)$ есть n-кратная свертка $F(x)$. Переходя к преобразованию Лапласа – Стилтьеса, получим

$$F^*(\alpha)=\int_0^\infty e^{-\alpha x}dF(x)=\int_0^\infty e^{-\alpha x}d_x\left\{\sum_{n=0}^{\infty}\int_0^x \frac{(\lambda y)^n}{n!}e^{-\lambda y}F^{(n)}(x-y)dB(y)\right\}=$$

$$=\sum_{n=0}^{\infty}\int_0^\infty e^{-\alpha x}d_x\int_0^x F^{(n)}(x-y)d_y\left\{\int_0^y \frac{(\lambda t)^n}{n!}e^{-\lambda t}dB(t)\right\}=$$

$$=\sum_{n=0}^{\infty}F^{*n}(\alpha)\int_0^\infty e^{-\alpha y}d_y\left\{\int_0^y \frac{(\lambda t)^n}{n!}e^{-\lambda t}dB(t)\right\}=$$

$$=\sum_{n=0}^{\infty}F^{*n}(\alpha)\int_0^\infty e^{-\alpha y}\frac{(\lambda y)^n}{n!}e^{-\lambda y}dB(y)=\int_0^\infty e^{-\alpha y}e^{\lambda y F^*(\alpha)}e^{-\lambda y}dB(y)=$$

$$=\int_0^\infty e^{-(\lambda+\alpha-\lambda F^*(\alpha))y}dB(y)=B^*(\lambda+\alpha-\lambda F^*(\alpha)),$$

где $B^*(s)$ есть преобразование Лапласа – Стилтьеса от $B(x)$. Итак, $F^*(\alpha)$ определяется как корень уравнения

$$F^*(\alpha)=B^*(\lambda+\alpha-\lambda F^*(\alpha)), \tag{4.8}$$

лежащий в интервале $0<F^*(\alpha)<1$.

§ 4.5. Виртуальное время ожидания в системе $M \mid G \mid 1 \mid \infty$

Обозначим через $\gamma(t)$ виртуальное время ожидания начала обслуживания виртуальной заявки, поступившей в систему в некоторый произвольный момент времени t, и пусть $F(t,x)=P\{\gamma(t)<x\}$. Выведем уравнение для $F(t,x)$, учитывая, что $\gamma(t)$ – марковский процесс.

Рассмотрим для этого два бесконечно близких момента времени t и $t+\Delta t$. Что могло произойти за этот период?

А) В систему за период Δt с вероятностью $1-\lambda\Delta t$ не пришло ни одной заявки. Тогда довольно очевидно, что $\gamma(t+\Delta t)=\gamma(t)-\Delta t$, и поэтому

$$F(t+\Delta t,x)=P\{\gamma(t+\Delta t)<x\}=P\{\gamma(t)<x+\Delta t\}=F(t,x+\Delta t).$$

Б) В систему за период Δt с вероятностью $\lambda\Delta t$ пришла заявка. Тогда виртуальной заявке, пришедшей в момент времени $t+\Delta t$, придется ждать время $\gamma(t)-\Delta t+y$, где y – время обслуживания заявки, пришедшей на интервале Δt. Так как y имеет функцию распределения $B(y)$, то

$$F(t+\Delta t,x-\Delta t)=(1-\lambda\Delta t)F(t,x)+\lambda\Delta t\int_0^x B(x-y)d_yF(t,y)+o(\Delta t).$$

Сократив на Δt и устремляя Δt к нулю, получим уравнение, определяющее $F(t,x)$:

$$\frac{\partial F(t,x)}{\partial t}=-\lambda F(t,x)+\frac{\partial F(t,x)}{\partial x}+\lambda\int_0^x B(x-y)d_yF(t,y). \qquad (4.9)$$

Предположим, что существует $\lim\limits_{t\to\infty}F(t,x)=F(x)$ и что при этом $\lim\limits_{t\to\infty}\frac{\partial F(t,x)}{\partial t}=0$. Тогда, переходя к этому пределу в (4.9), получим уравнение

$$\lambda F(x)=F'(x)+\lambda\int_0^x B(x-y)dF(y). \qquad (4.10)$$

Найдем преобразование Лапласа– Стилтьеса от производной $F'(x)$, учитывая, что $F(0)=P\big(i(t)=0\big)=1-\lambda b$:

$$\int_0^\infty e^{-\alpha x}dF'(x) = F'(+0) + \int_{+0}^\infty e^{-\alpha x}dF'(x) = F'(+0) + e^{-\alpha x}F'(x)\Big|_{+0}^\infty - \int_{+0}^\infty F'(x)de^{-\alpha x} =$$

$$= F'(+0) - F'(+0) + \alpha\int_{+0}^\infty e^{-\alpha x}F'(x)dx = \alpha\int_{+0}^\infty e^{-\alpha x}F'(x)dx = \alpha\int_{+0}^\infty e^{-\alpha x}dF(x) =$$

$$= \alpha\int_0^\infty e^{-\alpha x}dF(x) - \alpha(1-\lambda b) = \alpha F^*(\alpha) - \alpha(1-\lambda b).$$

Теперь найдем преобразование Лапласа – Стилтьеса от правой и левой части (4.10):

$$\lambda F^*(\alpha) = \alpha F^*(\alpha) - \alpha(1-\lambda b) + \lambda B^*(\alpha)F^*(\alpha), \tag{4.11}$$

$$\left(\lambda - \alpha - \lambda B^*(\alpha)\right)F^*(x) = -\alpha(1-\lambda b),$$

откуда получаем

$$F^*(x) = \frac{\alpha(1-\lambda b)}{\lambda B^*(\alpha) + \alpha - \lambda} = \frac{1-\lambda b}{1 - \dfrac{\lambda}{\alpha}\left(1 - B^*(\alpha)\right)}. \tag{4.12}$$

Зная явный вид $B^*(\alpha)$, можно найти $F^*(\alpha)$ и, следовательно, явный вид $F(x)$.

Найдем математическое ожидание $M\{\gamma(t)\}$, то есть среднее виртуальное время ожидания в очереди. Уравнение (4.11) перепишем в виде

$$\alpha F^*(\alpha) = \alpha(1-\lambda b) + \lambda\left(1 - B^*(\alpha)\right)F^*(\alpha).$$

Продифференцируем это соотношение дважды по α:

$$F^*(\alpha) + \alpha F^{*\prime}(\alpha) = 1 - \lambda b + \lambda F^{*\prime}(\alpha) - \lambda F^{*\prime}(\alpha)B^*(\alpha) - \lambda F^*(\alpha)B^{*\prime}(\alpha),$$

$$2F^{*\prime}(\alpha) + \alpha F^{*\prime\prime}(\alpha) = \lambda F^{*\prime\prime}(\alpha) - \lambda F^{*\prime\prime}(\alpha)B^*(\alpha) - 2\lambda F^{*\prime}(\alpha)B^{*\prime}(\alpha) -$$

$$-\lambda F^*(\alpha)B^{*\prime\prime}(\alpha),$$

и подставим $\alpha = 0$, помня, что

$$B^{*\prime}(0) = -b = -\int_0^\infty x dB(x), \quad B^{*\prime\prime}(0) = b_2 = \int_0^\infty x^2 dB(x),$$

$$F^{*\prime}(0) = -T_{\text{ож}}, \quad F^{*\prime\prime}(0) = D = \int_0^\infty x^2 dF(x),$$

а также то, что $F^*(0) = B^*(0) = 1$. В результате получим

$$-2T_{\text{ож}} = \lambda D - \lambda D - 2\lambda T_{\text{ож}} b - \lambda b_2,$$

откуда

$$T_{\text{ож}} = \frac{\lambda b_2}{2(1-\lambda b)} = \frac{\lambda(b^2 + \sigma_b^2)}{2(1-\lambda b)}, \tag{4.13}$$

где σ_b^2 есть дисперсия времени обслуживания заявки.

Обратим внимание на один очень интересный момент: среднее время ожидания растет с увеличением дисперсии времени обслуживания σ_b^2. Поэтому, чем более стандартизовано время обслуживания заявок, тем меньше среднее время ожидания начала обслуживания, и $T_{\text{ож}}$ минимально при $\sigma_b^2 = 0$.

§ 4.6. Система $GI \mid M \mid 1 \mid \infty$. Метод вложенных цепей Маркова

Пусть на обслуживающее устройство поступает рекуррентный поток заявок с функцией распределения длин интервалов между моментами поступления заявок $A(x)$. Обслуживание будем предполагать экспоненциальным с интенсивностью μ, то есть $B(x) = 1 - e^{-\mu x}$.

Изучим данную систему также методом вложенных цепей Маркова. Так как произвольным является поток, то и вложенные моменты времени t_n должны быть связаны с потоком. Выберем в качестве моментов времени, в которые рассматривается система, моменты прихода в систему n-й по счету заявки. Если $i(t)$ есть число заявок в момент времени t_n, то в качестве величин, характеризующих систему, выберем величины $\nu_n = i(t_n - 0)$, то есть число заявок в системе перед приходом заявки. Очевидно, что $\nu_n = 0, 1, 2, \ldots$

Рассчитаем переходные вероятности в данной системе, то есть величины $p_{ij} = P\{\nu_{n+1} = j \mid \nu_n = i\}$. Возможны следующие варианты.

1. $j > i+1$. Такого быть не может, так как это означало бы, что на интервале $[t_n - 0,\ t_{n+1} - 0]$ в систему пришло более одной заявки. Поэтому в этом случае $p_{ij} = 0$.

2. $i \geq 1$, $1 \leq j \leq i+1$. Это означает, что на интервале $[t_n, t_{n+1})$ обслужено $i+1-j$ заявок. Вероятность этого

$$p_{ij} = \int_0^\infty \frac{(\mu x)^{i+1-j}}{(i+1-j)!} e^{-\mu x} dA(x).$$

3. $i \geq 1$, $j = 0$. Это означает, что момент окончания обслуживания последней, $i+1$, заявки, лежит левее точки $t_n - 0$.

Соответствующие p_{ij} проще всего найти из условия нормировки $\sum_{j=0}^{i+1} p_{ij} = 1$, откуда, при $i+1-j = n$, получим

$$p_{i0} = 1 - \sum_{j=1}^{i+1} p_{ij} = 1 - \sum_{j=1}^{i+1} \int_0^\infty \frac{(\mu x)^{i+1-j}}{(i+1-j)!} e^{-\mu x} dA(x) = \int_0^\infty \left[1 - \sum_{n=0}^{i} \frac{(\mu x)^n}{n!} e^{-\mu x}\right] dA(x).$$

4. Аналогично получаем

$$p_{01} = \int_0^\infty e^{-\mu x} dA(x), \quad p_{00} = \int_0^\infty \left[1 - e^{-\mu x}\right] dA(x).$$

Выписав матрицу переходных вероятностей, легко получить, что цепь Маркова неприводима и непериодична, так как эта матрица имеет вид

i	j					
	0	1	2	3	4	…
0	*	*				…
1	*	*	*			…
2	*	*	*	*		…
3	*	*	*	*	*	…
…	…	…	…	…	…	…

где * означает ненулевые величины.

Система уравнений, определяющая финальные вероятности, примет вид

$$\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij} = \sum_{i=j-1}^{\infty} \pi_i p_{ij}$$

для $j \geq 1$. Не рассматривая пока уравнения для π_0, попробуем найти решение этого уравнения.

Подставляя сюда явное выражение для p_{ij}, получим

$$\pi_j = \sum_{i=j-1}^{\infty} \pi_i \int_0^{\infty} \frac{(\mu x)^{i+1-j}}{(i+1-j)!} e^{-\mu x} dA(x), \quad j \geq 1. \tag{4.14}$$

Будем искать решение этих уравнений в виде $\pi_j = Cz^j$. Подставляя это решение в (4.14) и выполнив замену $i+1-j=n$, получим

$$Cz^j = \sum_{i=j-1}^{\infty} Cz^i \int_0^{\infty} \frac{(\mu x)^{i+1-j}}{(i+1-j)!} e^{-\mu x} dA(x) = Cz^{j-1} \sum_{n=0}^{\infty} z^n \int_0^{\infty} \frac{(\mu x)^n}{n!} e^{-\mu x} dA(x) =$$

$$= Cz^{j-1} \int_0^{\infty} e^{-\mu x + \mu x z} dA(x) = Cz^{j-1} A^*\left(\mu(1-z)\right),$$

где $A^*(\alpha)$ есть преобразование Лапласа – Стилтьеса от $A(x)$. Сокращая на Cz^{j-1}, получим

$$z = A^*\left(\mu(1-z)\right). \tag{4.15}$$

Теорема. Если $a = \int_0^{\infty} x dA(x) > \frac{1}{\mu}$, то уравнение $z = A^*\left(\mu(1-z)\right)$ имеет единственный корень, лежащий в интервале $(0,\ 1)$.

Доказательство

Обозначая $\mu(1-z) = x$, приведем уравнение к виду

$$A^*(x) = 1 - \frac{x}{\mu}.$$

Рассмотрим вид графика функции

$$u(x) = A^*(x) - \left(1 - \frac{x}{\mu}\right).$$

Имеем:

1) $u(0) = A^*(0) - 1 = 1 - 1 = 0$;

2) $u'(x) = A^{*\prime}(x) + \frac{1}{\mu}$,

$$u'(0) = A^{*\prime}(0) + \frac{1}{\mu} = \frac{1}{\mu} - a < 0,$$

то есть в окрестности точки $x = 0$ $u(x)$ монотонно убывает;

3) $u''(x) = A^{*\prime\prime}(x) = \int\limits_0^\infty t^2 e^{-xt} dA(t) \ge 0$,

то есть $u(x)$ выпуклая вниз функция;

4) при $x = \mu$ $u(\mu) = A^*(\mu) \ge 0$.

Все это говорит о том, что график функции $u(x)$ имеет вид, изображенный на рис. 4.1. Из него следует, что уравнение $A^*(x) = 1 - \frac{x}{\mu}$ имеет единственный корень x_1, лежащий в интервале $(0, \mu)$. Возвращаясь к z, получим, что уравнение $z = A^*(\mu(1-z))$ имеет единственный корень z_1, лежащий в интервале $(0, 1)$.

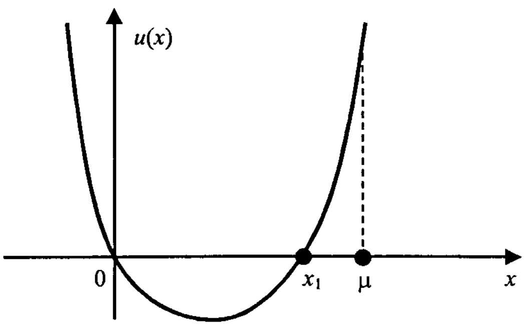


Рис. 4.1. Примерный вид графика функции $u(x)$

Таким образом, уравнениям (4.14) удовлетворят решения вида $\pi_j = Cz_1^j$.

Но у нас еще не было записано уравнение для π_0, так как оно особое, из-за особого вида величин p_{i0}. Покажем теперь, что найденное решение удовлетворяет и этому уравнению.

Используя выражение для p_{i0}, получим

$$\pi_0 = \sum_{i=0}^{\infty}\pi_i p_{i0} = \sum_{i=0}^{\infty}\pi_i \int_0^{\infty}\left[1-\sum_{n=0}^{i}\frac{(\mu x)^n}{n!}e^{-\mu x}\right]dA(x). \qquad (4.16)$$

Подставим сюда предполагаемое решение. Левая часть (4.16) будет тогда равна C. Посмотрим, чему равна правая часть:

$$\sum_{i=0}^{\infty}C\ z_1^i\left[1-\int_0^{\infty}\sum_{n=0}^{i}\frac{(\mu x)^n}{n!}e^{-\mu x}dA(x)\right]=\frac{C}{1-z_1}-C\sum_{n=0}^{\infty}\sum_{i=n}^{\infty}z_1^i\int_0^{\infty}\frac{(\mu x)^n}{n!}e^{-\mu x}dA(x)=$$

$$=\frac{C}{1-z_1}-C\sum_{n=0}^{\infty}\frac{z_1^n}{1-z_1}\int_0^{\infty}\frac{(\mu x)^n}{n!}e^{-\mu x}dA(x)=\frac{C}{1-z_1}-\frac{C}{1-z_1}\int_0^{\infty}e^{-\mu(1-z_1)x}dA(x)=$$

$$=\frac{C\left[1-A^*\left(\mu(1-z_1)\right)\right]}{1-z_1}=C.$$

Таким образом, предполагаемое решение обращает (4.16) в тождество.

Что касается C, то она находится из условия нормировки. Окончательно

$$\pi_i=(1-z_1)z_1^i, \qquad (4.17)$$

что по своей форме совпадает с видом финальных вероятностей в системе $M\,|\,M\,|\,1\,|\,\infty$. Только вместо ρ стоит z_1 – корень уравнения (4.15).

Найдем **время ожидания** в системе $GI\,|\,M\,|\,1\,|\,\infty$.

Приходящая в систему заявка с вероятностью π_n застает там n заявок. Все время ожидания будет равно

$$\tau_{\text{ож}}=\tau_1+\tau_2+\ldots+\tau_n,$$

где τ_i – время обслуживания n-й заявки. Так как все τ_i независимы и распределены экспоненциально, то

$$M\left\{e^{-s\tau_{\text{ож}}}\,|\,n\right\}=\left(\frac{\mu}{\mu+s}\right)^n.$$

Усредняя по n, получим

$$M\left\{\ e^{-s\tau_{\text{ож}}}\ \right\}=\sum_{n=0}^{\infty}(1-z_1)z_1^n\left(\frac{\mu}{\mu+s}\right)^n=\frac{1-z_1}{1-\dfrac{z_1\mu}{\mu+s}}=\frac{(1-z_1)(\mu+s)}{s+\mu(1-z_1)}=$$

$$= 1 - z_1 + z_1 \frac{\mu(1-z_1)}{s+\mu(1-z_1)}.$$

Обратное преобразование Лапласа – Стилтьеса определяет функцию распределения для времени ожидания

$$F_{\tau_{ож}}(x) = P\{\tau_{ож} < x\} = 1 - z_1 e^{-\mu(1-z_1)x},$$

что по своей форме также совпадает с результатом для системы $M \mid M \mid 1 \mid \infty$ с заменой ρ на z_1. В частности,

$$M\{\tau_{ож}\} = \frac{z_1}{\mu(1-z_1)}.$$

Отметим, что аналогично марковской модели функция распределения $F_{\tau_{ож}}(x)$ в нуле имеет разрыв и $F_{\tau_{ож}}(+0) = 1 - z_1$.

§ 4.7. Распределение числа заявок в системе $GI \mid M \mid 1 \mid \infty$ в произвольный момент времени. Метод дополнительной переменной

Так как для рассматриваемой СМО случайный процесс $i(t)$ немарковский, то марковизируем его, введя дополнительную переменную $z(t)$, равную длине интервала от момента t до момента поступления следующей заявки. Тогда для произвольной функции распределения $A(x)$ – определяющей входной рекуррентный поток заявок – двумерный процесс $\{i(t), z(t)\}$ будет марковским.

Пусть $P_i(z,t) = P(i(t) = i, z(t) < z)$, тогда

$$P_0(z, t+\Delta t) = P_0(z+\Delta t, t) - P_0(\Delta t, t) + \mu\Delta t P_1(z,t) + o(\Delta t),$$

$$P_1(z, t+\Delta t) = P_0(\Delta t, t)A(z) + \mu\Delta t P_2(z,t) +$$
$$+(1-\mu\Delta t)\left(P_1(z+\Delta t, t) - P_1(\Delta t, t)\right) + o(\Delta t),$$

$$P_i(z, t+\Delta t) = P_{i-1}(\Delta t, t)A(z) + \mu\Delta t P_{i+1}(z,t) +$$
$$+(1-\mu\Delta t)\left(P_i(z+\Delta t, t) - P_i(\Delta t, t)\right) + o(\Delta t).$$

В стационарном режиме $P_i(z,t) = P_i(z)$ удовлетворяет системе уравнений

$$0 = P_0'(z) - P_0'(0) + \mu P_1(z),$$

$$\mu P_1(z) = P_1'(z) - P_1'(0) + P_0'(0)A(z) + \mu P_2(z),$$

$$\mu P_i(z) = P_i'(z) - P_i'(0) + P_{i-1}'(0)A(z) + \mu P_{i+1}(z). \qquad (4.18)$$

Обозначим $\lim\limits_{z\to\infty} P_i(z) = \pi(i)$, тогда из системы (4.18) следует, что

$$0 = -P_0'(0) + \mu\pi(1),$$

$$\mu\pi(1) = -P_1'(0) + P_0'(0) + \mu\pi(2),$$

$$\mu\pi(i) = -P_i'(0) + P_{i-1}'(0) + \mu\pi(i+1),$$

откуда получим

$$P_0'(0) = \mu\pi(1),$$

$$P_1'(0) = \mu\pi(2),$$

$$P_i'(0) = \mu\pi(i+1),$$

а систему (4.18) перепишем в виде

$$0 = P_0'(z) - \mu\pi(1) + \mu P_1(z),$$

$$\mu P_1(z) = P_1'(z) - \mu\pi(2) + \mu\pi(1)A(z) + \mu P_2(z),$$

$$\mu P_i(z) = P_i'(z) - \mu\pi(i+1) + \mu\pi(i)A(z) + \mu P_{i+1}(z).$$

Решать эту систему будем в виде

$$P_0(z) = \pi(0)F_0(z),\ P_i(z) = \pi(i)F(z) \quad \text{для} \quad i = 1,2,...$$

Тогда получим систему

$$0 = \pi(0)F_0'(z) - \mu\pi(1) + \mu\pi(1)F(z); \qquad (4.19)$$

$$\begin{aligned} \mu\pi(1)F(z) &= \pi(1)F'(z) - \mu\pi(2) + \mu\pi(1)A(z) + \mu\pi(2)F(z), \\ \mu\pi(i)F(z) &= \pi(i)F'(z) - \mu\pi(i+1) + \mu\pi(i)A(z) + \mu\pi(i+1)F(z), \end{aligned} \qquad (4.20)$$

в которой две неизвестные функции распределения $F_0(z)$ и $F(z)$, а также неизвестное распределение $\pi(i)$ дискретной случайной величины.

Сначала рассмотрим систему (4.20), которую представим в виде

$$\begin{aligned} \mu\left(1 - \frac{\pi(2)}{\pi(1)}\right)F(z) &= F'(z) - \mu\frac{\pi(2)}{\pi(1)} + \mu A(z), \\ \mu\left(1 - \frac{\pi(i+1)}{\pi(i)}\right)F(z) &= F'(z) - \mu\frac{\pi(i+1)}{\pi(i)} + \mu A(z). \end{aligned} \qquad (4.21)$$

Пусть для всех $i=1,2,...$ $\frac{\pi(i+1)}{\pi(i)}=r$. Тогда в системе (4.21) все уравнения совпадают и имеют вид обыкновенного дифференциального уравнения первого порядка

$$\mu(1-r)F(z)=F'(z)-\mu(r-A(z)), \quad F(0)=0,$$

решение $F(z)$ которого представим как

$$F(z)=e^{\mu(1-r)z}\int_0^z e^{-\mu(1-r)s}\mu\left(r-A(s)\right)ds.$$

Так как существует $\lim\limits_{z\to\infty}F(z)=1$, то

$$\int_0^\infty e^{-\mu(1-r)s}\mu\left(r-A(s)\right)ds=0,$$

следовательно,

$$r=\mu(1-r)\int_0^\infty e^{-\mu(1-r)s}A(s)ds=\int_0^\infty e^{-\mu(1-r)s}dA(s)=A^*\left(\mu(1-r)\right)$$

и уравнение

$$r=A^*\left(\mu(1-r)\right), \tag{4.22}$$

определяющее величину r, совпадает с уравнением (4.15), определяющим корень z_1 в методе вложенных цепей Маркова для этой же СМО.

Таким образом, если в качестве r взять корень уравнения (4.22), то

$$\pi(i)=\pi(1)r^{i-1},$$

а функция распределения $F(z)$ имеет вид

$$F(z)=e^{\mu(1-r)z}\int_0^z e^{-\mu(1-r)s}\mu\left(r-A(s)\right)ds.$$

Чтобы найти $\pi(1)$ и $F_0(z)$, рассмотрим уравнение (4.19), которое представим в виде

$$F_0'(z)=\mu\frac{\pi(1)}{\pi(0)}\left(1-F(z)\right),$$

откуда

$$F_0(z)=\mu\frac{\pi(1)}{\pi(0)}\int_0^z\left(1-F(s)\right)ds. \tag{4.23}$$

Но так как $F_0(\infty)=1$, то $\mu\dfrac{\pi(1)}{\pi(0)}\int\limits_0^\infty\left(1-F(s)\right)ds=1$, следовательно,

$$\pi_1=\frac{\pi(0)}{\mu\int\limits_0^\infty\left(1-F(s)\right)ds}. \tag{4.24}$$

Полученные формулы (4.23) и (4.24) определяют $F_0(z)$ и $\pi(1)$. Вероятность $\pi(0)$ определим условием нормировки $\sum\limits_{i=0}^{\infty}\pi(i)=1$.

Чтобы найти $\int\limits_0^\infty\left(1-F(s)\right)ds$, рассмотрим

$$\int\limits_0^E\left(1-F(z)\right)dz=E-\int\limits_0^E F(z)dz=E-\int\limits_0^E e^{\mu(1-r)z}\int\limits_0^z e^{-\mu(1-r)s}\mu\left(r-A(s)\right)dsdz=$$

$$=E-\int\limits_0^E e^{-\mu(1-r)s}\mu\left(r-A(s)\right)\int\limits_s^E e^{\mu(1-r)z}dzds=$$

$$=E-\int\limits_0^E e^{-\mu(1-r)s}\mu\left(r-A(s)\right)\frac{1}{\mu(1-r)}\left[e^{\mu(1-r)E}-e^{\mu(1-r)s}\right]ds=$$

$$=E-\frac{1}{1-r}\left\{e^{\mu(1-r)E}\int\limits_0^E e^{-\mu(1-r)s}\left(r-A(s)\right)ds-\int\limits_0^E\left(r-A(s)\right)ds\right\}=$$

$$=E-\frac{1}{1-r}\left\{\frac{1}{\mu}F(E)-\int\limits_0^E\left(r-1+1-A(s)\right)ds\right\}=$$

$$=E-\frac{1}{1-r}\left\{\frac{1}{\mu}F(E)-(r-1)E-\int\limits_0^E\left(1-A(s)\right)ds\right\}=$$

$$=E-\frac{1}{\mu(1-r)}F(E)-E+\frac{1}{1-r}\int\limits_0^E\left(1-A(s)\right)ds=\frac{1}{1-r}\left(\int\limits_0^E\left(1-A(s)\right)ds-\frac{1}{\mu}F(E)\right).$$

Таким образом, при $E\to\infty$

$$\int\limits_0^\infty\left(1-F(z)\right)dz=\frac{1}{1-r}\left(\frac{1}{\lambda}-\frac{1}{\mu}\right)=\frac{1}{1-r}\frac{\mu-\lambda}{\lambda\mu}=\frac{1}{\lambda}\frac{1-\rho}{1-r},$$

следовательно,

$$\pi(1)=\frac{\pi(0)}{\mu\frac{1}{1-r}\left(\frac{1}{\lambda}-\frac{1}{\mu}\right)}=\frac{1-r}{\frac{\mu-\lambda}{\lambda}}\pi(0)=\frac{\rho(1-r)}{1-\rho}\pi(0).$$

Так как $\sum\limits_{i=0}^{\infty}\pi(i)=\pi(0)+\sum\limits_{i=1}^{\infty}\pi(1)r^{i-1}=\pi(0)+\pi(1)\frac{1}{1-r}=1$, то

$$\pi(0)\left[1+\frac{\rho(1-r)}{1-\rho}\frac{1}{1-r}\right]=\pi(0)\frac{1-\rho+\rho}{1-\rho}=\pi(0)\frac{1}{1-\rho}=1,$$

и поэтому

$$\pi(0)=1-\rho,$$

$$\pi(1)=\rho(1-r),$$

$$\pi(i)=\rho(1-r)r^{i-1},\quad i=1,2,3,\ldots \tag{4.25}$$

Здесь r определяется уравнением (4.22)

$$r=A^{*}\left(\mu(1-r)\right),$$

где $A^{*}(\alpha)=\int\limits_{0}^{\infty}e^{-\alpha x}dA(x)$.

Для системы $GI\mid M\mid 1\mid\infty$ распределение $\pi(i)$ в произвольный момент времени отличается от распределения π_i из (4.17) для моментов, предшествующих моментам поступления заявок в систему, и определяется равенством (4.25).

Глава 5. НЕМАРКОВСКИЕ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

Системы массового обслуживания, в которых число $i(t)$ заявок в системе является цепью Маркова с непрерывным временем, называются марковскими СМО. Системы массового обслуживания, в которых число $i(t)$ заявок в системе является полумарковским процессом, называются полумарковскими СМО.

Системы массового обслуживания, не являющиеся полумарковскими, будем называть немарковскими.

§ 5.1. Время ожидания в системе $GI \mid G \mid 1 \mid \infty$

5.1.1. Уравнение Линдли

Рассмотрим однолинейную СМО, на вход которой поступает рекуррентный поток, длины z_n интервалов между моментами поступления заявок, в котором имеют функцию распределения $A(z)$. Длительности η_n обслуживания заявок имеют функцию распределения $B(x)$. Последовательности $\{z_n\}$ и $\{\eta_n\}$ независимы. Блок-схема рассматриваемой СМО изображена на рис. 5.1.

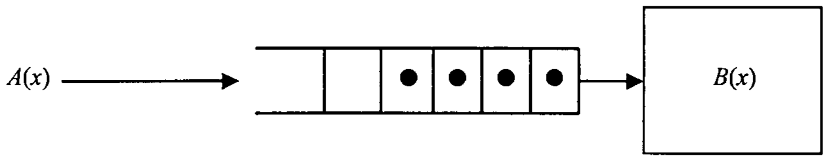


Рис. 5.1

В системе имеется буфер неограниченного объема. Заявка, заставшая прибор занятым, ожидает обслуживания в очереди, реализующей прямой порядок обслуживания. Такая СМО является немарковской системой массового обслуживания.

Обозначим W_n – время ожидания начала обслуживания n-й заявки. Очевидно, что $W_1 = 0$, так как первая заявка приходит в пустую систему.

Выразим W_n через W_{n-1} следующим образом:

$$W_n = \begin{cases} 0, \text{ если } W_{n-1} + \eta_{n-1} - z_n < 0, \\ W_{n-1} + \eta_{n-1} - z_n, \quad \text{если } W_{n-1} + \eta_{n-1} - z_n \geq 0. \end{cases} \tag{5.1}$$

В первом случае n-я заявка приходит в пустую систему, так как длина z_n интервала между моментами поступления $(n-1)$-й и n-й заявками больше суммарной величины времени ожидания W_{n-1} $(n-1)$-й заявки и ее времени обслуживания η_{n-1}.

Следовательно, последовательность W_1, W_2, W_3, ... образует марковский процесс с дискретным временем и непрерывным множеством состояний из полубесконечного интервала $[0, \infty)$.

Обозначим

$$F_n(x) = P(W_n < x).$$

Последовательность $F_n(x)$ удовлетворяет уравнению Чепмена – Колмогорова

$$F_{n+1}(x) = P(W_{n+1} < x) = P\left(W_n + \eta_n - z_{n+1} < x\right) =$$

$$= \int_{-\infty}^{x} P\left(W_n + \eta_n - z_{n+1} < x,\ y \leq \eta_n - z_{n+1} < y + dy\right) =$$

$$= \int_{-\infty}^{x} P\left(W_n < x - y\right) P\left(y \leq \eta_n - z_{n+1} < y + dy\right). \tag{5.2}$$

Обозначим $K(y)$ – функцию распределения разности двух независимых случайных величин $\eta_n - z_{n+1}$, то есть

$$P\left(\eta_n - z_{n+1} < y\right) = K(y),$$

тогда

$$F_{n+1}(x) = \begin{cases} \int\limits_{-\infty}^{x} F_n(x-y)dK(y), \ x > 0, \\ 0, \quad x \leq 0. \end{cases} \tag{5.3}$$

Если существует финальное распределение

$$F(x) = \lim_{n \to \infty} F_n(x),$$

то это распределение совпадает со стационарным и определяется уравнением

$$F(x) = \begin{cases} \int\limits_{-\infty}^{x} F(x-y)dK(y), & x > 0, \\ 0, & x \le 0, \end{cases} \tag{5.4}$$

которое называется интегральным уравнением Линдли. Здесь

$$K(y) = P\left(\eta_n - z_{n+1} < y\right) = \int\limits_0^{\infty} P(\eta_n - z_{n+1} < y,\ z \le z_{n+1} < z + dz) =$$

$$= \int\limits_0^{\infty} P(\eta_n < y + z)\ \ P(z \le z_{n+1} < z + dz) = \begin{cases} \int\limits_0^{\infty} B(y+z)dA(z) & \text{при} \quad y \ge 0, \\ \int\limits_{-y}^{\infty} B(y+z)dA(z) & \text{при} \quad y < 0. \end{cases}$$

Так как $B(y+z) = 0$ при $y + z \le 0$, то отсюда следует, что при $y < 0$

$$K(y) = \int\limits_{-y}^{\infty} B(y+z)dA(z).$$

Таким образом,

$$K(y) = \int\limits_{\max(0,-y)}^{\infty} B(y+z)dA(z); \tag{5.5}$$

$$F(x) = \begin{cases} \int\limits_{-\infty}^{x} F(x-y)dK(y), & x > 0, \\ 0, & x \le 0. \end{cases} \tag{5.6}$$

5.1.2. Метод Винера – Хопфа решения уравнения Линдли

Проблема решения уравнения Линдли заключается в том, что его правая часть не является сверткой функций, поэтому предлагается следующий метод его решения.

Введем

$$F_-(x)=\begin{cases}0 \quad \text{при} \quad x>0,\\ \int\limits_{-\infty}^{x} F(x-y)dK(y) \quad \text{при} \quad x\le 0,\end{cases} \tag{5.7}$$

тогда уравнение Линдли перепишем в виде

$$F_-(x)+F(x)=\int\limits_{-\infty}^{x} F(x-y)dK(y) \tag{5.8}$$

при всех $-\infty<x<\infty$, и в этом случае правая часть становится сверткой функций $F(x)$ и $K(y)$.

Определим следующие преобразования Лапласа:

$$\Phi_+(\alpha)=\int\limits_{0}^{\infty} e^{-\alpha x}F(x)dx,$$

$$\Phi_-(\alpha)=\int\limits_{-\infty}^{0} e^{-\alpha x}F_-(x)dx.$$

Функция $\Phi_+(\alpha)$ аналитична в области $\operatorname{Re}\alpha>0$. Аналитичность функции $\Phi_-(\alpha)$ требует дополнительных условий.

Пусть существуют плотности распределения

$$a(x)=A'(x), \quad b(x)=B'(x), \quad k(x)=K'(x).$$

Пусть существует такое $c>0$, что

$$\lim_{x\to\infty}\frac{a(x)}{e^{-cx}}<\infty,$$

то есть на бесконечности $a(x)$ ведет себя как e^{-cx}. Тогда

$$K(x)=\int\limits_{-x}^{\infty} B(x+z)dA(z)=\int\limits_{-x}^{\infty} B(x+z)a(z)dz=\int\limits_{0}^{\infty} B(y)a(y-x)dy\sim e^{cx}$$

при $x\to-\infty$.

Для функции $F_-(x)$ получим при $x\to-\infty$

$$F_-(x)=\int\limits_{-\infty}^{x} F(x-y)dK(y)=\int\limits_{-\infty}^{x} F(x-y)k(y)dy=\int\limits_{0}^{\infty} F(z)k(x-y)dz\sim e^{cx}.$$

Следовательно, преобразование Лапласа

$$\Phi_-(\alpha)=\int\limits_{-\infty}^{0} e^{-\alpha x}F(x)dx$$

будет аналитической функцией при $\operatorname{Re}\alpha < c$. Таким образом, обе функции $\Phi_+(\alpha)$ и $\Phi_-(\alpha)$ одновременно аналитичны в полосе

$$0 < \operatorname{Re}\alpha < c.$$

Применяя к уравнению (5.8)

$$F_-(x) + F(x) = \int_{-\infty}^{x} F(x-y)dK(y)$$

двухстороннее преобразование Лапласа, получим

$$\Phi_-(\alpha) + \Phi_+(\alpha) = \int_{-\infty}^{\infty} e^{-\alpha x}\left\{\int_{-\infty}^{x} F(x-y)dK(y)\right\}dx = \int_{-\infty}^{\infty}\int_{y}^{\infty} e^{-\alpha x} F(x-y)dxdK(y) =$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} e^{-\alpha(y+z)} F(z)dzdK(y) = \int_{-\infty}^{\infty} e^{-\alpha y}\Phi_+(\alpha)dK(y) =$$

$$= \Phi_+(\alpha)\int_{-\infty}^{\infty} e^{-\alpha y}dK(y) = \Phi_+(\alpha)K^*(\alpha).$$

Здесь $K^*(\alpha)$ – двухстороннее преобразование Лапласа – Стилтьеса функции распределения $K(x)$. Далее имеем

$$K^*(\alpha) = \int_{-\infty}^{\infty} e^{-\alpha y}dK(y) = e^{-\alpha y}K(y)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} K(y)de^{-\alpha y} =$$

$$= \begin{bmatrix} \lim\limits_{y\to\infty} e^{-\alpha y}K(y) = 0 \\ \lim\limits_{y\to-\infty} e^{-\alpha y}K(y) = 0, \text{так как } K(y) \sim e^{cy}, \text{ при y} \to -\infty \end{bmatrix} =$$

$$= \alpha\int_{-\infty}^{\infty} K(y)e^{-\alpha y}dy = \alpha\int_{-\infty}^{0} K(y)e^{-\alpha y}dy + \alpha\int_{0}^{\infty} K(y)e^{-\alpha y}dy =$$

$$= \alpha\int_{-\infty}^{0} e^{-\alpha y}\left\{\int_{-y}^{\infty} B(y+z)dA(z)\right\}dy + \alpha\int_{0}^{\infty} e^{-\alpha y}\left\{\int_{0}^{\infty} B(y+z)dA(z)\right\}dy =$$

$$= [\text{поменяем порядок интегрирования}] =$$

$$= \alpha\int_{0}^{\infty}\left\{\int_{-z}^{0} e^{-\alpha y}B(\underbrace{y+z}_{x})dy\right\}dA(z) + \alpha\int_{0}^{\infty}\left\{\int_{0}^{\infty} e^{-\alpha y}B(\underbrace{y+z}_{x})dy\right\}dA(z) =$$

$$=\alpha\int_0^\infty\left\{\int_0^z e^{-\alpha(x-z)}B(x)dx\right\}dA(z)+\alpha\int_0^\infty\left\{\int_z^\infty e^{-\alpha(x-z)}B(x)dx\right\}dA(z)=$$

$$=\alpha\int_0^\infty e^{\alpha z}\int_0^\infty e^{-\alpha x}B(x)dxdA(z)=\alpha\int_0^\infty e^{-\alpha x}B(x)dx\cdot\int_0^\infty e^{\alpha z}dA(z)=$$

$$=\alpha\int_0^\infty e^{\alpha z}a(z)dz\cdot\int_0^\infty B(x)d\left(-e^{-\alpha x}\right)=a^*(-\alpha)b^*(\alpha),$$

где $a^*(\alpha)$ и $b^*(\alpha)$ – преобразования Лапласа от плотностей $a(x)$ и $b(x)$; $a^*(-\alpha)$ существует, так как $a(x)\sim e^{-cx}$ при $\operatorname{Re}\alpha<c$.

Таким образом, в полосе

$$0<\operatorname{Re}\alpha<c$$

$$\Phi_-(\alpha)+\Phi_+(\alpha)=\Phi_+(\alpha)K^*(\alpha)=\Phi_+(\alpha)a^*(-\alpha)b^*(\alpha),$$

то есть

$$\Phi_-(\alpha)=\Phi_+(\alpha)\left(a^*(-\alpha)b^*(\alpha)-1\right). \qquad (5.9)$$

Основная идея метода Винера – Хопфа заключается в том, чтобы найти функции $\Psi_+(\alpha)$ и $\Psi_-(\alpha)$ такие, что

$$a^*(-\alpha)b^*(\alpha)-1=\frac{\Psi_+(\alpha)}{\Psi_-(\alpha)}.$$

При этом $\Psi_+(\alpha)$ должна быть аналитической в области $\operatorname{Re}\alpha>0$ и не иметь нулей в этой области, а $\Psi_-(\alpha)$ аналитична в области $\operatorname{Re}\alpha<c$ и также не иметь нулей в этой полуплоскости. Тогда будет выполнено уравнение

$$\Phi_-(\alpha)=\Phi_+(\alpha)\frac{\Psi_+(\alpha)}{\Psi_-(\alpha)} \text{ или } \Phi_-(\alpha)\Psi_-(\alpha)=\Phi_+(\alpha)\Psi_+(\alpha). \qquad (5.10)$$

Левая часть является функцией аналитической в полуплоскости $\operatorname{Re}\alpha>0$, правая – в полуплоскости $\operatorname{Re}\alpha<c$. Их общая область аналитичности является полосой $0<\operatorname{Re}\alpha<c$, в которой нет нулей этих функций и эти функции ограничены.

Их аналитические продолжения на всю комплексную плоскость являются ограниченной аналитической функцией на всей комплексной плоскости, не имеющей нулей. Следовательно, в силу теоремы Лиувилля, эта функция является константой, которую обозначим L:

$$\Phi_+(\alpha)\Psi_+(\alpha)=L \text{ или } \Phi_+(\alpha)=\frac{L}{\Psi_+(\alpha)}. \tag{5.11}$$

Здесь $\Psi_+(\alpha)$ – известная функция, а

$$\Phi_+(\alpha)=\int_0^\infty e^{-\alpha x}F(x)dx$$

– преобразование Лапласа искомой функции распределения $F(x)$.

Для однозначного определения функции $\Phi_+(\alpha)$ осталось найти константу L. Так как

$$\lim_{\alpha\to 0}\alpha\Phi_+(\alpha)=\lim_{x\to\infty}F(x)=1, \text{ то } \lim_{\alpha\to 0}\alpha\frac{L}{\Psi_+(\alpha)}=1,$$

следовательно,

$$L=\lim_{\alpha\to 0}\frac{\Psi_+(\alpha)}{\alpha}. \tag{5.12}$$

Зная преобразование Лапласа $\Phi_+(\alpha)$, можно найти функцию распределения $F(x)$, используя либо формулу обращения, либо таблицы обратного преобразования Лапласа.

Пример. Рассмотрим систему $M \mid M \mid 1 \mid \infty$, для которой $a(x)=\lambda e^{-\lambda x}$, $b(x)=\mu e^{-\mu x}$. Поэтому $a^*(-\alpha)=\dfrac{\lambda}{\lambda-\alpha}$, $b^*(\alpha)=\dfrac{\mu}{\mu+\alpha}$ и, следовательно,

$$a^*(-\alpha)b^*(\alpha)-1=\frac{\lambda}{\lambda-\alpha}\frac{\mu}{\mu+\alpha}-1=\frac{\lambda\mu-(\lambda-\alpha)(\mu+\alpha)}{(\lambda-\alpha)(\mu+\alpha)}=$$

$$=\frac{\alpha(\alpha+\mu-\lambda)}{(\lambda-\alpha)(\mu+\alpha)}=\frac{\Psi_+(\alpha)}{\Psi_-(\alpha)}.$$

Нули этой функции находятся в точках $\alpha=0$, $\alpha=\lambda-\mu<0$, а полюса – в точках $\alpha=\lambda$, $\alpha=-\mu$ (рис. 5.2).

Так как функция $\Psi_+(\alpha)$ должна быть аналитична при $\mathrm{Re}\,\alpha>0$, то ее следует взять в виде

$$\Psi_+(\alpha)=\frac{\alpha(\alpha+\mu-\lambda)}{(\mu+\alpha)}.$$

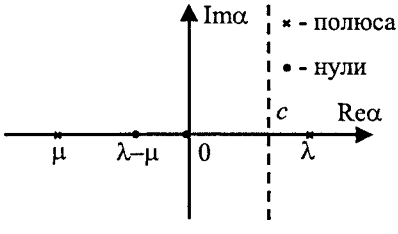


Рис.5.2

Так как функция $\Psi_-(\alpha)$ должна быть аналитична при $\mathrm{Re}\,\alpha < c$, то ее следует взять в виде $\Psi_-(\alpha) = \lambda - \alpha$. Тогда и будет выполнено соотношение

$$a^*(-\alpha)b^*(\alpha) - 1 = \frac{\Psi_+(\alpha)}{\Psi_-(\alpha)}.$$

Найдем L:

$$L = \lim_{\alpha\to 0}\frac{\Psi_+(\alpha)}{\alpha} = \frac{\mu-\lambda}{\mu} = 1-\rho,$$

следовательно,

$$\Phi_+(\alpha) = \frac{L}{\Psi_+(\alpha)} = (1-\rho)\frac{\mu+\alpha}{\alpha(\alpha+\mu-\lambda)} = \frac{A}{\alpha} + \frac{B}{\alpha+\mu-\lambda} = \frac{A(\alpha+\mu-\lambda)+B\alpha}{\alpha(\alpha+\mu-\lambda)} =$$

$$= \frac{\mu-\lambda+(1-\rho)\alpha}{\alpha(\alpha+\mu-\lambda)}.$$

Так как $A(\mu-\lambda) = \mu-\lambda$, то $A = 1$, так как $A + B = 1-\rho$, то $B = -\rho$. Поэтому

$$\Phi_+(\alpha) = \frac{1}{\alpha} - \frac{\rho}{\alpha+\mu-\lambda}.$$

Обратное преобразование Лапласа имеет вид

$$F(x) = 1-\rho\ \ e^{-(\mu-\lambda)x},$$

так как $\int\limits_0^\infty e^{-\alpha x}dx = \frac{1}{\alpha}$, $\int\limits_0^\infty e^{-\alpha x}e^{-(\mu-\lambda)x}dx = \frac{1}{\alpha+\mu-\lambda}$, что совпадает с полученным ранее результатом (3.18) для распределения времени ожидания в марковской системе $M\,|\,M\,|\,1\,|\,\infty$.

§ 5.2. Теорема Литтла

В системе массового обслуживания определим: ν_n – время пребывания в системе n-й заявки; $i(t)$ – число заявок в системе в момент времени t; $N(t)$ – число заявок, поступивших в систему за время t.

Теорема Литтла. Пусть выполнены условия

1) $P\left(\lim\limits_{T\to\infty}\dfrac{N(T)}{T}=\lambda\right)=1$;

2) процессы ν_n и $i(t)$ – эргодичны.

Тогда

$$M\{i(t)\}=\lambda\cdot M\{\nu_n\}.$$

Доказательство

Обозначим $\Delta(a,b,c)$ – суммарное время, проведенное в системе на интервале (b,c) теми заявками, которые поступили в систему на интервале (a,b).

Тогда имеем равенство

$$\int\limits_0^T i(t)dt=\Delta(-\infty,0,T)+\sum_{n=1}^{N(T)}\nu_n-\Delta(0,T,\infty).$$

Здесь $\int\limits_0^T i(t)dt$ – суммарное время, проведенное на интервале $(0,T)$ заявками, которые на этом интервале находились в системе; $\sum\limits_{n=1}^{N(T)}\nu_n$ – суммарное время пребывания в системе заявок, поступивших на интервале $(0,T)$.

Имеем

$$\frac{1}{T}\int\limits_0^T i(t)dt=\frac{\Delta(-\infty,0,T)}{T}+\frac{N(T)}{T}\frac{1}{N(T)}\sum_{n=1}^{N(T)}\nu_n-\frac{\Delta(0,T,\infty)}{T}.$$

В силу стационарности величины $\Delta(-\infty,0,T)$ и $\Delta(0,T,\infty)$ конечны, поэтому при $T\to\infty$ в силу эргодичности имеем

$$\frac{1}{T}\int\limits_0^T i(t)dt\to M\{i(t)\},\quad \frac{1}{N(T)}\sum_{n=1}^{N(T)}\nu_n=M\{\nu_n\},$$

следовательно, $M\{i(t)\}=\lambda\cdot M\{\nu_n\}$.

§ 5.3. Исследование немарковской однолинейной СМО при инверсионном порядке обслуживания с прерыванием и дообслуживанием вытесненных заявок

Рассмотрим однолинейную СМО с неограниченным бункером, на вход которой поступает простейший поток заявок с интенсивностью λ (рис. 5.3).

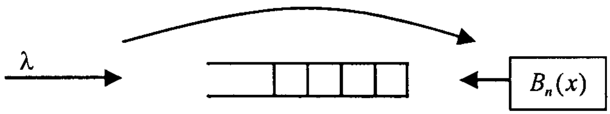


Рис. 5.3

Если в момент поступления заявки в системе находится $n-1$ требование (здесь n = 1, 2, 3, ...), то поступившей заявке присваивается номер n и она ставится на прибор, вытесняя обслуживаемую заявку в очередь. Поступившей заявке с номером n назначается время обслуживания случайной продолжительности с функцией распределения $B_n(x)$. Если за время обслуживания этой заявки другие требования в систему не поступят, то, завершив обслуживание, эта заявка покинет систему, а на прибор ставится заявка из очереди с максимальным $n-1$-м номером. Тем самым реализуется инверсионный порядок обслуживания (дисциплина *LIFO*). При вытеснении заявки с прибора ее обслуживание прерывается, а при возвращении этой заявки из очереди на прибор ее обслуживание начинается с прерванного момента, то есть реализуется процедура дообслуживания заявки.

При реализации такой дисциплины обслуживания принципиальным является то, что все заявки, поступающие в систему, проходят тест на продолжительность реализации своего времени обслуживания. Заявки с короткими реализациями успевают завершить обслуживание и покинуть систему до момента прерывания, связанного с приходом следующей заявки.

Обозначим через $n(t)$ число заявок в системе в момент времени t, $z_i(t)$ – продолжительность остаточного времени обслуживания заявки с номером i, находящейся в системе, и пусть

$$P(n(t)=n, z_1(t)<z_1,\ldots,z_n(t)<z_n) = P_n(z_1,z_2,\ldots,z_n,t).$$

Для распределения $P_n(z_1,z_2,...,z_n,t)$ можно записать

$$P_0(t+\Delta t)=(1-\lambda\Delta t)P_0(t)+P_1(\Delta t,t)+o(\Delta t),$$

$$P_1(z_1-\Delta t,t+\Delta t)=(1-\lambda\Delta t)\left[P_1(z_1,t)-P_1(\Delta t,t)\right]+$$
$$+\lambda\Delta tP_0(t)B_1(z_1)+P_2(z_1,\Delta t,t)+o(\Delta t),$$

$$P_2(z_1,z_2-\Delta t,t+\Delta t)=(1-\lambda\Delta t)\left[P_2(z_1,z_2,t)-P_2(z_1,\Delta t,t)\right]+$$
$$+\lambda\Delta tP_1(z_1,t)B_2(z_2)+P_3(z_1,z_2,\Delta t,t)+o(\Delta t),$$

....

$$P_n(z_1,...,z_{n-1},z_n-\Delta t,t+\Delta t)=(1-\lambda\Delta t)\left[P_n(z_1,...,z_n,t)-P_2(z_1,...,z_{n-1},\Delta t,t)\right]+$$
$$+\lambda\Delta tP_{n-1}(z_1,z_2,...,z_{n-1},t)B_n(z_n)+P_{n+1}(z_1,...,z_n,\Delta t,t)+o(\Delta t).$$

В стационарном режиме обозначим

$$P_n(z_1,z_2,...,z_n,t)=P_n(z_1,z_2,...,z_n)$$

и перепишем, выполнив несложные преобразования, последнюю систему в виде

$$\lambda P_0=\frac{\partial P_1(0)}{\partial z_1},$$

$$\lambda P_1(z_1)=\frac{\partial P_1(z_1)}{\partial z_1}-\frac{\partial P_1(0)}{\partial z_1}+\lambda P_0B_1(z_1)+\frac{\partial P_2(z_1,0)}{\partial z_2},$$

$$\lambda P_2(z_1,z_2)=\frac{\partial P_2(z_1,z_2)}{\partial z_2}-\frac{\partial P_2(z_1,0)}{\partial z_2}+\lambda P_1(z_1)B_2(z_2)+\frac{\partial P_3(z_1,z_2,0)}{\partial z_3},$$

....

$$\lambda P_n(z_1,...,z_{n-1},z_n)=\frac{\partial P_n(z_1,z_2,...,z_n)}{\partial z_n}-\frac{\partial P_n(z_1,z_2,...,0)}{\partial z_n}+$$
$$+\lambda P_{n-1}(z_1,z_2,...,z_{n-1})B_n(z_n)+\frac{\partial P_{n+1}(z_1,z_2,...,z_n,0)}{\partial z_{n+1}},\quad n=2,3,...$$

Решение полученной системы будем икать в мультипликативном виде

$$P_n(z_1,z_2,...,z_n)=P_0\prod_{i=1}^{n}f_i(z_i),\quad n\geq 1.$$

Тогда система примет вид

$$\lambda = \frac{\partial f_1(0)}{\partial z_1}$$

$$\lambda f_1(z_1) = \frac{\partial f_1(z_1)}{\partial z_1} - \frac{\partial f_1(0)}{\partial z_1} + \lambda B_1(z_1) + f_1(z_1)\frac{\partial f_2(0)}{\partial z_2},$$

$$\lambda f_2(z_2) = \frac{\partial f_2(z_2)}{\partial z_2} - \frac{\partial f_2(0)}{\partial z_2} + \lambda B_2(z_2) + f_2(z_2)\frac{\partial f_3(0)}{\partial z_3},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\lambda f_n(z_n) = \frac{\partial f_n(z_n)}{\partial z_n} - \frac{\partial f_n(0)}{\partial z_n} + \lambda B_n(z_n) + f_n(z_n)\frac{\partial f_{n+1}(0)}{\partial z_{n+1}}. \tag{5.13}$$

Положим системе (5.13) все z_i равными бесконечности, в результате получим

$$\frac{\partial f_1(0)}{\partial z_1} = \lambda, \quad \frac{\partial f_2(0)}{\partial z_2} = \lambda, \quad \dots, \quad \frac{\partial f_n(0)}{\partial z_n} = \lambda.$$

Следовательно, систему (5.13) можно записать в виде

$$\frac{\partial f_1(z_1)}{\partial z_1} = \lambda\left(1 - B_1(z_1)\right),$$

$$\frac{\partial f_2(z_2)}{\partial z_2} = \lambda\left(1 - B_2(z_2)\right),$$

$$\cdots\cdots\cdots\cdots\cdots,$$

$$\frac{\partial f_n(z_n)}{\partial z_n} = \lambda\left(1 - B_n(z_n)\right).$$

Отсюда находится вид функций $f_i(z_i)$:

$$f_i(z_i) = \lambda\int_0^{z_i}\left(1 - B_i(s)\right)ds.$$

Таким образом,

$$P_n(z_1, z_2, \dots, z_n) = P_0\prod_{i=1}^{n}\lambda\int_0^{z_i}\left(1 - B_i(s)\right)ds. \tag{5.14}$$

Полагая $z_1 = z_2 = \dots = \infty$, получим

$$P_n = P_0 \prod_{i=1}^{n} \rho_i, \quad P_0 = 1 \Big/ \left[1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} \rho_i \right], \tag{5.15}$$

где $\rho_i = \lambda b_i = \lambda \int\limits_0^{\infty} (1 - B_i(s))\,ds$.

Если все функции распределения $B_i(z)$ одинаковые, то, обозначив $B_i(z) = B(z)$, $b_i = b$, $\rho_i = \rho$, получим

$$P_n = P_0 \rho^n,$$

где $P_0 = 1 \Big/ \left[1 + \sum\limits_{n=1}^{\infty} \rho^n \right] = 1 - \rho$ и, следовательно, $P_n = (1-\rho)\rho^n$.

Таким образом, для рассматриваемой немарковской СМО стационарное распределение вероятностей P_n является, аналогично марковской системе, геометрическим и определяется лишь загрузкой ρ канала обслуживания. Здесь $\rho = \lambda b$ зависит лишь от интенсивности входящего потока и среднего времени обслуживания и не зависит от вида функции распределения $B(z)$ времени обслуживания.

§ 5.4. Исследование системы $M \mid G \mid \infty$

Рассмотрим систему $M|G|\infty$, то есть бесконечно линейную СМО, на вход которой поступает простейший поток с параметром λ. Обслуживание каждым прибором рекуррентное с функцией распределения $B(x)$ времени обслуживания (рис. 5.4).

Обозначим $i(t)$ – число заявок в системе в момент времени t. Очевидно, случайный процесс $i(t)$ – немарковский.

Для нахождения распределения вероятностей $\pi(i) = P(i(t) = i)$ в стационарном режиме поступим следующим образом.

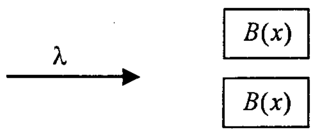


Рис. 5.4

Рассмотрим вспомогательную N линейную СМО, **состоящую из N независимых однолинейных** систем, каждая из которых имеет следующую структуру.

Имеется один источник заявок и один обслуживающий прибор.

Источник в течение случайного времени, распределенного по экспоненциальному закону с параметром $\frac{\lambda}{N-\lambda b}$, где $b=\int_0^{\infty}(1-B(x))dx$, генерирует заявку, которую в момент окончания генерирования отправляет на прибор, где она обслуживается в течение случайного времени с функцией распределения $B(x)$. Завершив обслуживание, заявка покидает систему, а источник начинает генерировать новую заявку.

Состояние рассматриваемой однолинейной СМО определим величиной r, принимающей только два значения: $r=0$, если источник генерирует заявку, и $r=1$, если заявка находится на обслуживании. Процесс $r(t)$ изменения во времени состояний r немарковский. Его исследование выполним методом дополнительной переменной, определив компоненту $z(t)$ как длину интервала от момента t до момента окончания обслуживания. Если в момент t заявка генерируется источником, то компонента $z(t)$ не определяется в силу экспоненциальноети времени генерирования заявки.

Естественно, что случайный процесс $\{r(t),z(t)\}$ – марковский.

Обозначим

$$P_0(t)=P(r(t)=0),$$

$$P_1(z,t)=P(r(t)=1,z(t)<z).$$

Для этого распределения можно записать

$$P_0(t+\Delta t)=\left(1-\frac{\lambda}{N-\lambda b}\Delta t\right)P_0(t)+P_1(\Delta t,t)+o(\Delta t),$$

$$P_1(z-\Delta t,t+\Delta t)=P_1(z,t)-P_1(\Delta t,t)+\frac{\lambda}{N-\lambda b}\Delta tP_0(t)B(z)+o(\Delta t),$$

откуда получим

$$P_0'(t)+\frac{\lambda}{N-\lambda b}P_0(t)=\frac{\partial P_1(0,t)}{\partial z},$$

$$\frac{\partial P_1(z,t)}{\partial t}-\frac{\partial P_1(z,t)}{\partial z}=-\frac{\partial P_1(0,t)}{\partial z}+\frac{\lambda}{N-\lambda b}P_0(t)B(z).$$

В стационарном режиме полученную систему, обозначив $P_0(t)=P_0$, $P_1(z,t)=P_1(z)$, перепишем в виде

$$\frac{\lambda}{N-\lambda b}P_0 = \frac{\partial P_1(0)}{\partial z},$$

$$\frac{\partial P_1(z)}{\partial z} = \frac{\partial P_1(0)}{\partial z} - \frac{\lambda}{N-\lambda b}P_0 B(z).$$

Отсюда очевидно следует, что

$$\frac{\partial P_1(z)}{\partial z} = \frac{\lambda}{N-\lambda b}P_0\left(1-B(z)\right),$$

то есть

$$P_1(z) = \frac{\lambda}{N-\lambda b}P_0\int_0^z \left(1-B(x)\right)dx.$$

При $z \to \infty$ получаем

$$P_1 = \frac{\rho}{N}, \quad P_0 = 1-\frac{\rho}{N}, \quad \text{где } \rho = \lambda b.$$

Распределение вероятностей $P_N(i)$ числа приборов, занятых в N-линейной вспомогательной СМО, определяется схемой Бернулли и имеет биноминальное распределение

$$P_N(i) = C_N^i P_1^i (1-P_1)^{N-i},$$

где $P_1 = \frac{\rho}{N}$, которое в силу предельной теоремы Пуассона при $N \to \infty$ имеет вид

$$P(i) = \lim_{N\to\infty} P_N(i) = \frac{\rho^i}{i!}e^{-\rho}.$$

На каждую однолинейную СМО во вспомогательной N-линейной системе поступает рекуррентный поток заявок, средняя длина интервалов между моментами поступления которых составляет

$$b + \frac{N-\lambda b}{\lambda} = \frac{N}{\lambda},$$

и, следовательно, его интенсивность равна $\frac{\lambda}{N}$.

Суммарная интенсивность поступления заявок в бесконечно линейную СМО равна λ и не зависит от N.

При $N \to \infty$ сумма независимых рекуррентных потоков, в силу теоремы Григелиониса, дает предельный поток, который является простейшим с параметром λ, равным интенсивности поступления заявок в N-линейную СМО.

Таким образом, при $N \to \infty$ вспомогательная N-линейная СМО переходит в исходную бесконечно линейную систему с простейшим входящим потоком и рекуррентным обслуживанием, поэтому для распределения вероятностей $\pi(i)$ можно записать

$$\pi(i) = P(i) = \lim_{N\to\infty} P_N(i) = \frac{\rho^i}{i!} e^{-\rho},$$

то есть распределение $\pi(i)$ числа приборов, занятых в системе $M \mid G \mid \infty$, является пуассоновским с параметром $\rho = \lambda b$, зависящим лишь от интенсивности λ и среднего значения $b = \int\limits_0^\infty \left(1 - B(x)\right) dx$ времени обслуживания, и не зависит от вида функции распределения $B(x)$.

§ 5.5. Выходящий поток системы $M \mid G \mid \infty$

Вновь рассмотрим вспомогательную N-линейную СМО, состоящую из N независимых однолинейных систем.

Очевидно, что выходящий поток каждой из этих СМО является рекуррентным, длины интервалов между моментами появления заявок в котором складываются из времени η обслуживания с функцией распределения $B(x)$ и экспоненциально распределенного времени ξ_N генерирования заявок с параметром $\dfrac{\lambda}{N - \lambda b}$. Функцию распределения длины такого интервала обозначим $A_N(x)$. Тогда

$$A_N(x) = P(\xi_N + \eta < x) = \int\limits_0^x P(\xi_N + \eta < x,\ y \le \xi_N < y + dy) =$$

$$= \int\limits_0^x P(\eta < x - y) P(y \le \xi_N < y + dy) = \int\limits_0^x B(x - y) \left(\frac{\lambda}{N - \lambda b}\right) e^{-\frac{\lambda}{N - \lambda b} y} dy.$$

Среднее значение длины такого интервала составит

$$a_N = b + \frac{N - \lambda b}{\lambda} = \frac{N}{\lambda}.$$

Обозначим τ_n – величину перескока выходящего потока для n-й однолинейной СМО. Запишем ее функцию распределения $G_N(z)$:

$$G_N(z) = \frac{\lambda}{N}\int_0^z (1 - A_N(x)dx) = P(\tau_n < z).$$

Все τ_n распределены одинаково и независимы.

Для суммарного выходящего потока величина τ_N перескока составит

$$\tau_N = \min_{1 \le n \le N} \tau_n.$$

Найдем

$$P(\tau_N > u) = P(\min_{1\le n\le N} \tau_n > u) = P^N(\tau_n > u) = (1 - G_N(u))^N =$$

$$= \left\{1 - \frac{\lambda}{N}\int_0^u (1 - A_N(x))dx\right\}^N = \exp\left\{N \cdot \ln\left[1 - \frac{\lambda}{N}\int_0^u (1 - A_N(x))dx\right]\right\}.$$

Предел выражения в фигурных скобках при $N \to \infty$ составит

$$\lim_{N\to\infty} N \cdot \ln\left[1 - \frac{\lambda}{N}\int_0^u (1 - A_N(x))dx\right] = \lim_{N\to\infty} N\left[-\frac{\lambda}{N}\int_0^u (1 - A_N(x))dx\right] =$$

$$= -\lambda \lim_{N\to\infty}\int_0^u (1 - A_N(x))dx = -\lambda\int_0^u (1 - \lim_{N\to\infty} A_N(x))dx = -\lambda u,$$

следовательно,

$$\lim_{N\to\infty} P(\tau_N < u) = 1 - e^{-\lambda u}.$$

Таким образом, предельное значение величины перескока для суммарного выходящего потока имеет экспоненциальное распределение с параметром λ, следовательно, выходящий поток для СМО $M|G|\infty$ является простейшим с тем же параметром λ, что и входящий.

Полученный результат достаточно очевидно следует из предельной теоремы Григелиониса о сумме независимых редких рекуррентных потоков, которая утверждает, что предельный поток является простейшим.

А совпадение интенсивностей входящего и выходящего потоков для СМО $M|G|\infty$ очевиден в силу закона сохранения, так как в этой системе обслуживаются и покидают систему все поступившие в нее заявки.

§ 5.6. Период занятости в системе $M \mid G \mid \infty$

Периодом занятости системы $M\,|G\,|\infty$ назовем интервал от момента поступления заявки в свободную систему до момента ее опустошения. Его длину обозначим T.

Рассмотрим $\mathrm{T}(t)$ – длину интервала от момента t, когда в системе имеются заявки, до момента ее опустошения.

Для нахождения распределения вероятностей величины $\mathrm{T}(t)$ определим $z(t)$ как продолжительность остаточного времени обслуживания, максимального для всех заявок, находящихся в системе в момент времени t.

Обозначим

$$g(\alpha,z,t)=M\left\{ e^{-\alpha\mathrm{T}(t)}\,|\,z(t)=z\right\}.$$

Рассмотрим

$$g(\alpha,z+\Delta t,t-\Delta t)=M\left\{ e^{-\alpha\mathrm{T}(t-\Delta t)}\,|\,z(t-\Delta t)=z+\Delta t\right\}=$$

$$=(1-\lambda\Delta t)M\left\{ e^{-\alpha(\Delta t+\mathrm{T}(t))}\,|\,z(t)=z\right\}+$$

$$+\lambda\Delta t\left[B(z)M\left\{e^{-\alpha\mathrm{T}(t)}\,|\,z(t)=z\right\}+\int\limits_z^\infty g(\alpha,x,t)dB(x)\right].$$

В стационарном режиме это равенство перепишется в виде $g(\alpha,z+\Delta t)=\left[1-(\lambda+\alpha)\Delta t)\right]g(\alpha,z)+\lambda\Delta t\left[B(z)g(\alpha,z)+\int\limits_z^\infty g(\alpha,x)dB(x)\right]$,

откуда, после несложных преобразований, получим

$$\begin{cases}\dfrac{\partial g(\alpha,z)}{\partial z}+\left[(\lambda+\alpha)-\lambda B(z)\right]g(\alpha,z)=\lambda\int\limits_z^\infty g(\alpha,x)dB(x), & (5.16)\\ g(\alpha,0)\equiv 1. & (5.17)\end{cases}$$

Полученная задача Коши для интегродифференциального уравнения (5.16) определяет условную характеристическую функцию длины $\mathrm{T}(t)$ интервала от момента t до момента опустошения системы при условии, что выполняется равенство $z(t)=z$.

Период занятости начинается поступлением заявки в пустую систему, и в этот момент величина $z(t)$ остаточного времени обслуживания совпадает с полным временем обслуживания заявки, имеющим функцию распределения $B(x)$, поэтому безусловная характеристическая функция $g(\alpha)$ длины T периода занятости имеет вид

$$g(\alpha)=M\{e^{-\alpha T}\}=\int_0^\infty g(\alpha,x)dB(x)\,. \tag{5.18}$$

Решение интегродифференциального уравнения (5.16) представляет определенные сложности, поэтому выполним следующие преобразования.

Дифференцируя (5.16) по z, получим

$$\frac{\partial^2 g(\alpha,z)}{\partial z^2}+[\lambda+\alpha-\lambda B(z)]\frac{\partial g(\alpha,z)}{\partial z}-\lambda\frac{dB(z)}{dz}g(\alpha,z)=-\lambda g(\alpha,z)\frac{dB(z)}{dz}\,,$$

откуда для функции $g(\alpha,z)$ получим обыкновенное дифференциальное уравнение

$$\frac{\partial^2 g(\alpha,z)}{\partial z^2}+[\lambda+\alpha-\lambda B(z)]\frac{\partial g(\alpha,z)}{\partial z}=0$$

второго порядка, в которое входят только производные функции $g(\alpha,z)$ по z. Поэтому его порядок можно понизить заменой

$$\frac{\partial g(\alpha,z)}{\partial z}=h(\alpha,z)\,.$$

Для функции $h(\alpha,z)$ получим уравнение

$$\frac{\partial h(\alpha,z)}{\partial z}+[\lambda+\alpha-\lambda B(z)]\;h(\alpha,z)=0\,,$$

решение которого имеет вид

$$h(\alpha,z)=C(\alpha)e^{-\int_0^z(\lambda+\alpha-\lambda B(x))dx}\,.$$

Учитывая начальное условие (5.17), функцию $g(\alpha,z)$ запишем в виде

$$g(\alpha,z)=1+C(\alpha)\int_0^z e^{-\int_0^y(\lambda+\alpha-\lambda B(x))dx}dy\,. \tag{5.19}$$

Так как при $\alpha=0$ $g(0,z)\equiv 1$, то $C(0)=0$. При $\alpha>0$ не зависящую от z функцию $C(\alpha)$ определим, подставив (5.19) в (5.16). Тогда получим тождественное по z равенство

$$C(\alpha)e^{-\int_0^z(\lambda+\alpha-\lambda B(x))dx}+\left[\lambda+\alpha-\lambda B(z)\right]\left\{1+C(\alpha)\int_0^z e^{-\int_0^y(\lambda+\alpha-\lambda B(x))dx}dy\right\}=$$

$$=\lambda\int_z^\infty\left\{1+C(\alpha)\int_0^v e^{-\int_0^y(\lambda+\alpha-\lambda B(x))dx}dy\right\}dB(v),$$

в котором выполним предельный переход при $z\to\infty$.

Учитывая, что при $\alpha>0$ имеет место равенство

$$\lim_{z\to\infty}e^{-\int_0^z(\lambda+\alpha-\lambda B(x))dx}=\lim_{z\to\infty}e^{-\alpha z}e^{-\lambda\int_0^z(1-B(x))dx}=0,$$

получим

$$C(\alpha)=\frac{-1}{\int_0^\infty e^{-\int_0^y(\lambda+\alpha-\lambda B(x))dx}dy}=\frac{-1}{\int_0^\infty e^{-\alpha y}e^{-\lambda\int_0^y(1-B(x))dx}dy}. \tag{5.20}$$

Далее, подставляя выражение (5.19) в (5.18), получим

$$g(\alpha)=\int_0^\infty\left\{1+C(\alpha)\int_0^z e^{-\int_0^y(\lambda+\alpha-\lambda B(x))dx}dy\right\}dB(z)=1+C(\alpha)\int_0^\infty\int_0^z e^{-\int_0^y(\lambda+\alpha-\lambda B(x))dx}dydB(z)=$$

$$=1+C(\alpha)\int_0^\infty(1-B(y))e^{-\int_0^y(\lambda+\alpha-\lambda B(x))dx}dy=$$

$$=1+\frac{C(\alpha)}{\lambda}\int_0^\infty(\lambda+\alpha-\lambda B(y)-\alpha)e^{-\int_0^y(\lambda+\alpha-\lambda B(x))dx}dy=$$

$$=1+\frac{C(\alpha)}{\lambda}\left\{\int_0^{\infty}(\lambda+\alpha-\lambda B(y))e^{-\int_0^y(\lambda+\alpha-\lambda B(x))dx}dy-\alpha\int_0^{\infty}e^{-\int_0^y(\lambda+\alpha-\lambda B(x))dx}dy\right\}=$$

$$=1+\frac{C(\alpha)}{\lambda}\left\{1-\alpha\int_0^{\infty}e^{-\alpha y}e^{-\lambda\int_0^y(1-B(x))dx}dy\right\}.$$

Подставляя сюда выражение (5.20) для $C(\alpha)$, получим вид характеристической функции $g(\alpha)$ продолжительности T периода занятости в бесконечно линейной СМО $M|G|\infty$

$$g(\alpha)=1-\frac{1-\alpha\int_0^{\infty}e^{-\alpha y}e^{-\lambda\int_0^y(1-B(x))dx}dy}{\lambda\int_0^{\infty}e^{-\alpha y}e^{-\lambda\int_0^y(1-B(x))dx}dy}=1+\frac{\alpha}{\lambda}-\frac{1}{\lambda}\frac{1}{\int_0^{\infty}e^{-\alpha y}e^{-\lambda\int_0^y(1-B(x))dx}dy}. \tag{5.21}$$

Обозначив

$$\varphi(\alpha)=\alpha\int_0^{\infty}e^{-\alpha y}e^{-\lambda\int_0^y(1-B(x))dx}dy, \tag{5.22}$$

функцию $g(\alpha)$ запишем в виде

$$g(\alpha)=1+\frac{\alpha}{\lambda}-\frac{1}{\lambda}\frac{\alpha}{\varphi(\alpha)}. \tag{5.23}$$

Например, если

$$B(x)=\begin{cases}0, & \text{при} \quad x\le b,\\ 1, & \text{при} \quad x>b,\end{cases}$$

то

$$\varphi(\alpha)=\alpha\int_0^{b}e^{-\alpha y}e^{-\lambda y}dy+\alpha\int_b^{\infty}e^{-\alpha y}e^{-\lambda b}dy=\frac{\alpha}{\alpha+\lambda}\left(1-e^{-(\alpha+\lambda)b}\right)+e^{-\lambda b}e^{-\alpha b}=$$

$$=\frac{\alpha}{\alpha+\lambda}+\frac{\lambda}{\alpha+\lambda}e^{-(\alpha+\lambda)b}$$

и, следовательно,

$$g(\alpha)=1+\frac{\alpha}{\lambda}-\frac{1}{\lambda}\frac{\alpha}{\dfrac{\alpha}{\alpha+\lambda}+\dfrac{\lambda}{\alpha+\lambda}e^{-(\alpha+\lambda)b}}=\frac{\alpha+\lambda}{\lambda}-\frac{(\alpha+\lambda)\alpha}{\lambda\left(\alpha+\lambda e^{-(\alpha+\lambda)b}\right)}=$$

$$=\frac{\alpha+\lambda}{\lambda}\frac{\alpha+\lambda e^{-(\alpha+\lambda)b}-\alpha}{\alpha+\lambda e^{-(\alpha+\lambda)b}}=\frac{\alpha+\lambda}{1}\frac{e^{-(\alpha+\lambda)b}}{\alpha+\lambda e^{-(\alpha+\lambda)b}}=\frac{\lambda+\alpha}{\lambda+\alpha e^{(\alpha+\lambda)b}}.$$

Таким образом, характеристическая функция $g(\alpha)$ продолжительности T периода занятости для СМО $M|G|\infty$ с детерминированным обслуживанием имеет вид

$$g(\alpha)=\frac{\lambda+\alpha}{\lambda+\alpha e^{(\alpha+\lambda)b}}.$$

Найдем среднее значение $M\{\mathrm{T}\}$ продолжительности периода занятости для произвольной функции распределения $B(x)$ времени обслуживания.

Из формулы (5.23) следует, что

$$g'(\alpha)=\frac{1}{\lambda}-\frac{1}{\lambda}\frac{\varphi(\alpha)-\alpha\varphi'(\alpha)}{\varphi^2(\alpha)}. \qquad (5.24)$$

Так как

$$\lim_{\alpha\to 0}\varphi(\alpha)=\lim_{\alpha\to 0}\alpha\int_0^{\infty}e^{-\alpha y}e^{-\lambda\int_0^y(1-B(x))dx}dy=\lim_{y\to\infty}e^{-\lambda\int_0^y(1-B(x))dx}=e^{-\lambda b}=e^{-\rho},$$

то

$$M\{\mathrm{T}\}=-g'(\alpha)\big|_{\alpha=0}=\frac{1}{\lambda}\frac{e^{-\rho}}{e^{-2\rho}}-\frac{1}{\lambda}=\frac{e^{\rho}-1}{\lambda}$$

и не зависит от вида функции $B(x)$.

Найдем второй момент $M\{\mathrm{T}^2\}$ продолжительности периода занятости. Из (5.24) получим

$$g''(\alpha)=-\frac{1}{\lambda}\frac{\left(\varphi'(\alpha)-\varphi'(\alpha)-\alpha\varphi''(\alpha)\right)\varphi(\alpha)-\left(\varphi(\alpha)-\alpha\varphi'(\alpha)\right)2\varphi'(\alpha)}{\varphi^3(\alpha)}=$$

$$=\frac{1}{\lambda}\frac{2\left(\varphi(\alpha)-\alpha\varphi'(\alpha)\right)\varphi'(\alpha)+\alpha\varphi''(\alpha)\varphi(\alpha)}{\varphi^3(\alpha)},$$

следовательно,

$$M\{\mathrm{T}^2\} = \lim_{\alpha\to 0} g''(\alpha) = \frac{2}{\lambda} e^{3\rho} e^{-\rho} \lim_{\alpha\to 0} \varphi'(\alpha) = \frac{2}{\lambda} e^{2\rho} \lim_{\alpha\to 0} \varphi'(\alpha) .$$

Из формулы (5.22) получим

$$\varphi(\alpha) = -\int_0^\infty e^{-\lambda\int_0^y (1-B(x))dx} de^{-\alpha y} = -e^{-\lambda\int_0^y (1-B(x))dx} e^{-\alpha y}\Bigg|_0^\infty + \int_0^\infty e^{-\alpha y} de^{-\lambda\int_0^y (1-B(x))dx} =$$

$$= 1 + \int_0^\infty e^{-\alpha y} de^{-\lambda\int_0^y (1-B(x))dx},$$

и, следовательно,

$$\varphi'(\alpha) = -\int_0^\infty e^{-\alpha y} y \; de^{-\lambda\int_0^y (1-B(x))dx} .$$

Поэтому

$$\lim_{\alpha\to 0}\varphi'(\alpha) = -\int_0^\infty y \; de^{-\lambda\int_0^y (1-B(x))dx} .$$

Найдем это значение для функции распределения детерминированной величины b. Имеем

$$\lim_{\alpha\to 0}\varphi'(\alpha) = -\int_0^b y \; de^{-\lambda y} = -y \; e^{-\lambda y}\Big|_0^b + \int_0^b e^{-\lambda y} \; dy =$$

$$= -be^{-\lambda b} + \frac{1}{\lambda}\left(1 - e^{-\lambda b}\right) = \frac{1 - e^{-\rho} - \rho e^{-\rho}}{\lambda}.$$

Следовательно, второй момент $M\{\mathrm{T}^2\}$ продолжительности периода занятости при детерминированном обслуживании составит

$$M\{\mathrm{T}^2\} = \frac{2e^{2\rho}}{\lambda^2}\left(1 - e^{-\rho} - \rho e^{-\rho}\right),$$

а для произвольной функции $B(x)$ запишем

$$M\left\{\mathrm{T}^2\right\}=-\frac{2e^{2\rho}}{\lambda}\int_0^{\infty} y d e^{-\lambda\int_0^y(1-B(x))dx}=2e^{2\rho}\int_0^{\infty} y(1-B(y))e^{-\lambda\int_0^y(1-B(x))dx}dy .$$

Очевидно, что второй момент $M\left\{\mathrm{T}^2\right\}$ зависит от вида функции $B(x)$.

§ 5.7. Формулы Энгсета для марковских СМО

Рассмотрим систему массового обслуживания, содержащую N источников и блока обслуживания произвольной структуры. Функционирование системы реализуется следующим способом: n-источник (то есть источник с номером n = 1, 2, ..., N) случайное время с функцией распределения $A_n(x)=1-e^{-\lambda_n x}$ генерирует заявку и отправляет ее в блок обслуживания. Если требование принимается к обслуживанию, то оно находится в блоке обслуживания случайное время с функцией распределения $B_n(x)=1-e^{-\mu_n x}$. Завершив обслуживание, источник вновь начинает генерировать заявку. Если требованию отказано в обслуживании, источник начинает генерировать заявку заново. Время формирования новой заявки не зависит от того, было ли успешным обслуживание предыдущей заявки, либо ей в этом отказано.

Определим состояние системы состоянием источников, для этого обозначим через r_n состояние n-источника, положив $r_n=0$, если n-источник генерирует заявку, и $r_n=1$, если она находится на обслуживании. Состояние системы определим N-мерным булевым вектором

$$R=(r_1,r_2,...,r_N),$$

значения которого образуют вершины единичного N-мерного гиперкуба.

Если множество G существенных состояний R рассматриваемой СМО включает все вершины гиперкуба (то есть все заявки принимаются к обслуживанию), то исследование системы тривиально, так как источники функционируют стохастически независимо и распределение вероятностей состояний вектора R является произведением распределений вероятностей его компонент, а распределение вероятностей $P_n(r)=P(r_n=r)$ его n-й компоненты удовлетворят уравнению

$$\lambda_n P_n(0)=\mu_n P_n(1)$$

и условию нормировки $P_n(0)+P_n(1)=1$. Поэтому

$$P_n(0)=\frac{1}{1+\rho_n},\quad P_n(1)=\frac{\rho_n}{1+\rho_n},\quad \text{где } \rho_n=\frac{\lambda_n}{\mu_n}.$$

Таким образом,

$$P_n(r)=\begin{cases}\dfrac{1}{1+\rho_n}, \text{ если } r=0,\\ \dfrac{\rho_n}{1+\rho_n}, \text{ если } r=1,\end{cases} \tag{5.25}$$

для всех $n=1,\ 2,\ ...,\ N$.

Выражение (5.25) можно записать в виде

$$P_n(r)=\frac{\rho_n^r}{1+\rho_n},$$

тогда распределение $P(R)=P(r_1,r_2,...,r_N)$ вектора R состояний системы примет вид

$$P(r_1,r_2,...,r_N)=\prod_{n=1}^{N}P_n(r_n)=\prod_{n=1}^{N}\frac{\rho_n^{r_n}}{1+\rho_n}=\prod_{n=1}^{N}\frac{1}{1+\rho_n}\prod_{n=1}^{N}\rho_n^{r_n}=P(0)\prod_{n=1}^{N}\rho_n^{r_n}. \tag{5.26}$$

Гораздо более сложная, но практически интересная ситуация возникает в системах, в которых заявка принимается к обслуживанию в зависимости от состояния системы и заполненности блока обслуживания. В этом случае функционирование источников стохастически зависимо и необходимо их рассматривать в совокупности, в отличие от предыдущего случая независимого функционирования источников.

Будем полагать, что существуют подмножества $G_n \subset G$ существенных состояний

$$R_{n0}=(r_1,r_2,...r_{n-1},0,r_{n+1}...,r_N),$$

таких, что состояния

$$R_{n1}=(r_1,r_2,...r_{n-1},1,r_{n+1}...,r_N)$$

являются несущественными, то есть в состояниях $R \in G_n$ заявкам n-источника отказывается в обслуживании.

Определим функции $\delta(R)$ в виде

$$\delta_n(R_{n0})=\begin{cases}1, & \text{если } R_{n0}\in G_n,\\ 0, & \text{если } R_{n0}\notin G_n.\end{cases}$$

Функция $\delta_n(R)$ является индикатором множества G_n. Структура множества G существенных состояний такова, что если $R \in G$, то $R_{n0} \in G$ для всех $n = 1, 2, ..., N$. Если $R \notin G$, то $R_{n1} \notin G$ для всех $n = 1, 2, ..., N$.

Найдем распределение вероятностей $P(R)$ для такой СМО.

Формулы для распределения вероятностей $P(R)$ назовем формулами Энгсета для марковской СМО.

Процесс $R(t)$ изменения во времени состояний R рассматриваемой СМО является дискретным марковским процессом, поэтому его распределение

$$P(R,t) = P(R(t) = R)$$

удовлетворят равенствам

$$P(R,t+\Delta t) = \left\{1 - \sum_{n=1}^{N} [\lambda_n(1-r_n) + \mu_n r_n]\Delta t\right\} P(R,t) + \sum_{n=1}^{N} \lambda_n r_n \Delta t P(R_{n0},t) +$$

$$+\sum_{n=1}^{N}(1-r_n)\left[P(R_{n1},t)\mu_n\Delta t + \delta_n(R_{n0})P(R_{n0},t)\lambda_n\Delta t\right] + o(\Delta t).$$

В стационарном режиме $P(R,t) = P(R)$.

Для распределения $P(R)$ получим систему уравнений

$$\sum_{n=1}^{N}\left\{-[\lambda_n(1-r_n) + \mu_n r_n]P(R) + r_n\lambda_n P(R_{n0}) +\right.$$

$$\left. +(1-r_n)[\ \mu_n P(R_{n1}) + \delta_n(R_{n0})\lambda_n P(R_{n0})\ ]\ \right\} = 0. \qquad (5.27)$$

Отметим, что R_{n1} может быть несущественным, тогда $P(R_{n1}) = 0$.

Каждое уравнение этой системы заменим на N уравнений, полагая каждое слагаемое в (5.27) равным нулю:

$$[\lambda_n(1-r_n) + \mu_n r_n]P(R) = r_n\lambda_n P(R_{n0}) + (1-r_n)[\mu_n P(R_{n1}) + \delta_n(R_{n0})\lambda_n P(R_{n0})]. \qquad (5.28)$$

Пусть $R_{n0} \in G_n$, тогда R_{n1} – несущественное состояние и его вероятность $P(R_{n1}) = 0$, а $\delta_n(R_{n0}) = 1$. Следовательно, система (5.28) при $r_n = 0$ примет вид тождества

$$\lambda_n P(R_{n0}) = \lambda_n P(R_{n0}).$$

Пусть $R_{n0} \notin G_n$. Тогда R_{n1} – существенное состояние, а $\delta_n(R_{n0}) = 0$. Поэтому система (4) примет вид

1) для $r_n = 0$

$$\lambda_n P(R_{n0}) = \mu_n P(R_{n1});$$

2) для $r_n = 1$

$$\mu_n P(R_{n1}) = \lambda_n P(R_{n0}).$$

Следовательно, для любых n и тех R, для которых R_{n1} – существенное состояние, выполняется равенство

$$P(R_{n1}) = \rho_n P(R_{n0}),$$

откуда

$$P(R) = P(0)\prod_{n=1}^{N} \rho_n^{r_n}.$$

Так как

$$1 = \sum_{R\in G} P(R) = P(0)\sum_{R\in G}\prod_{n=1}^{N} \rho_n^{r_n},$$

то формулы Энгсета для марковской СМО имеют вид

$$\begin{cases} P(0) = 1\Big/\sum\limits_{R\in G}\prod\limits_{n=1}^{N} \rho_n^{r_n}, \\ P(R) = P(0)\prod\limits_{n=1}^{N} \rho_n^{r_n}. \end{cases} \tag{5.29}$$

Отметим, что если G содержит все вершины N-мерного гиперкуба, то

$$\sum_{R\in G}\prod_{n=1}^{N} \rho_n^{r_n} = \prod_{n=1}^{N}(1+\rho_n) \tag{5.30}$$

и формулы Энгсета, совпадающие с (5.26), дают произведение одномерных распределений.

Докажем равенство (5.30).

1. При $N = 1$ получим

$$\sum_{R\in G}\prod_{n=1}^{1} \rho_1^{r_1} = \rho_1^0 + \rho_1^1 = 1 + \rho_1.$$

2. Пусть (5.30) верна для $N - 1$. Тогда для N получим

$$\sum_{R\in G}\prod_{n=1}^{N}\rho_n^{r_n}=\sum_{r_N=0}^{1}\sum_{R^{N-1}\in G^{N-1}}\prod_{n=1}^{N}\rho_n^{r_n}=\sum_{r_N=0}^{1}\rho_N^{r_N}\prod_{n=1}^{N-1}(1+\rho_n)=$$

$$=\prod_{n=1}^{N-1}(1+\rho_n)+\rho_N\prod_{n=1}^{N-1}(1+\rho_n)=(1+\rho_N)\prod_{n=1}^{N-1}(1+\rho_n)=\prod_{n=1}^{N}(1+\rho_n).$$

В общем случае, как следует из (5.29), компоненты r_n вектора R стохастически зависимы, ибо

$$P(R)\neq\prod_{n=1}^{N}P_n(r_n),$$

где $P_n(r_n)=\dfrac{\rho_n^{r_n}}{1+\rho_n}$.

Действительно,

$$P(R)=P(0)\prod_{n=1}^{N}\rho_n^{r_n},\quad \prod_{n=1}^{N}P_n(r_n)=\prod_{n=1}^{N}\frac{1}{1+\rho_n}\prod_{n=1}^{N}\rho_n^{r_n},$$

но $P(0)$ из (5.29) не равно $\prod\limits_{n=1}^{N}\dfrac{1}{1+\rho_n}$.

Пусть $N=2$, а состояние $R=(1,\ 1)$ несущественное, тогда $G=\{(0,\ 0),(1,\ 0),(0,\ 1)\}$,

$$P(0)=1\Big/\sum_{R\in G}\prod_{n=1}^{2}\rho_n^{r_n}=1/\{1+\rho_1+\rho_2\},\qquad \prod_{n=1}^{2}\frac{1}{1+\rho_n}=1/\{1+\rho_1+\rho_2+\rho_1\rho_2\},$$

откуда очевидно следует, что

$$P(0)\neq\prod_{n=1}^{2}\frac{1}{1+\rho_n}.$$

Таким образом, несмотря на то, что $P(R)$ в (5.29) имеет форму произведения $\prod\limits_{n=1}^{N}\rho_n^{r_n}$, тем не менее компоненты r_n вектора R стохастически зависимы, так как в (5.29) сомножители не являются распределениями вероятностей компонент.

§ 5.8. Формулы Энгсета для неоднородных немарковских СМО

Рассмотрим систему массового обслуживания, исследование которой приведено в предыдущем разделе, для произвольных функций распределения $A_n(x)$ и $B_n(x)$.

В этом случае процесс $R(t)$ является немарковским, поэтому определим вектор

$$Z(t)=\{z_1(t),z_2(t),...z_N(t)\},$$

где $z_n(t)$ – длина интервала от момента t до момента окончания текущего режима n-источника, то это есть либо остаточное время генерирования заявки, если $r_n(t)=0$, либо остаточное время обслуживания, если $r_n(t)=1$. Тогда $2N$-мерный случайный процесс $\{R(t),Z(t)\}$ является марковским.

Обозначим

$$P(R,X,t)=P(r_1(t)=r_1,...,r_N(t)=r_N,z_1(t)<x_1,...,z_N(t)<x_N),$$

единичный N-мерный вектор

$$E=\{1,\ 1,\ ...,\ 1\},$$

а также вектор вида

$$X_n=\{x_1,...,x_{n-1},\Delta t,x_{n+1},...,x_N\}.$$

Для распределения вероятностей $P(R,X,t)$ можно записать равенства

$$P(R,X-\Delta tE,t+\Delta t)=P(R,X,t)-\sum_{n=1}^{N}P(R,X_n,t)+\sum_{n=1}^{N}r_nB_n(x_n)P(R_{n0},X_n,t)+$$

$$+\sum_{n=1}^{N}(1-r_n)A_n(x_n)\left[P(R_{n1},X_n,t)+\delta_n(R_{n0})P(R_{n0},X_n,t)\right]+o(\Delta t).$$

В стационарном режиме $P(R,X,t)=P(R,X)$, и система уравнений для него имеет вид

$$\sum_{n=1}^{N}\left\{\frac{\partial P(R,X)}{\partial x_n}-\frac{\partial P(R,X)}{\partial x_n}\bigg|_{x_n=0}+r_nB_n(x_n)\frac{\partial P(R_{n0},X)}{\partial x_n}\bigg|_{x_n=0}+\right.$$

$$+(1-r_n)A_n(x_n)\left[\left.\frac{\partial P(R_{n1},X)}{\partial x_n}\right|_{x_n=0}+\delta_n(R_{n0})\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}\right]\Bigg\}=0. \quad (5.31)$$

Каждое уравнение этой системы заменим на N уравнений, полагая каждое слагаемое в (5.31) равным нулю:

$$\frac{\partial P(R,X)}{\partial x_n}=\left.\frac{\partial P(R,X)}{\partial x_n}\right|_{x_n=0}-r_nB_n(x_n)\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}-$$

$$-(1-r_n)A_n(x_n)\left[\left.\frac{\partial P(R_{n1},X)}{\partial x_n}\right|_{x_n=0}+\delta_n(R_{n0})\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}\right], \quad (5.32)$$

$$n=1,\ 2,\ \ldots,\ N.$$

Пусть $R_{n0}\in G_n$, тогда $P(R_{n1},X)\equiv 0$ и в системе (5.32) такие уравнения примут вид для R_{n0}

$$\frac{\partial P(R_{n0},X)}{\partial x_n}=\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}-A_n(x_n)\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}=$$

$$=\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}\left(1-A_n(x_n)\right). \quad (5.33)$$

Пусть $R_{n0}\notin G_n$. Тогда R_{n1} – существенное состояние, $\delta_n(R_{n0})=0$, и поэтому система (5.32) примет вид

1) для $r_n=0$

$$\frac{\partial P(R_{n0},X)}{\partial x_n}=\left.\frac{\partial P(R_{n1},X)}{\partial x_n}\right|_{x_n=0}-A_n(x_n)\left.\frac{\partial P(R_{n1},X)}{\partial x_n}\right|_{x_n=0}=$$

$$=\left.\frac{\partial P(R_{n1},X)}{\partial x_n}\right|_{x_n=0}\left(1-A_n(x_n)\right); \quad (5.34)$$

2) для $r_n=1$

$$\frac{\partial P(R_{n1},X)}{\partial x_n}=\left.\frac{\partial P(R_{n1},X)}{\partial x_n}\right|_{x_n=0}-B_n(x_n)\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}. \quad (5.35)$$

Таким образом, для любого состояния R в систему уравнений (5.32) входят уравнения трех типов (5.33) – (5.35).

Для уравнения (5.35) можно записать

$$P(R_{n1},X)=\int_0^{x_n}\left\{\left.\frac{\partial P(R_{n1},X)}{\partial x_n}\right|_{x_n=0}-B_n(y)\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}\right\}dy\,. \qquad (5.36)$$

Так как $P(R_{n1},X)$ имеет конечный предел при $x_n\to\infty$, то в силу необходимого условия сходимости несобственного интеграла имеет место равенство

$$\left.\frac{\partial P(R_{n1},X)}{\partial x_n}\right|_{x_n=0}=\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}, \qquad (5.37)$$

и, следовательно, уравнение (5.36) можно записать в виде

$$P(R_{n1},X)=\left.\frac{\partial P(R_{n1},X)}{\partial x_n}\right|_{x_n=0}\int_0^{x_n}(1-B_n(y))dy\,. \qquad (5.38)$$

Аналогично из равенства (5.34), с учетом (5.37), запишем

$$P(R_{n0},X)=\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}\int_0^{x_n}(1-A_n(y))dy\,. \qquad (5.39)$$

Полученное равенство является также и решением уравнения (5.33).

Из (5.38) и (5.39) получим распределение $P(R,X)$ в виде

$$P(R,X)=\left.\frac{\partial P(R_{n0},X)}{\partial x_n}\right|_{x_n=0}\left\{\int_0^{x_n}(1-B_n(y))dy\right\}^{r_n}\left\{\int_0^{x_n}(1-A_n(y))dy\right\}^{1-r_n}=$$

$$=C\prod_{n=1}^{N}\left\{\int_0^{x_n}(1-B_n(y))dy\right\}^{r_n}\left\{\int_0^{x_n}(1-A_n(y))dy\right\}^{1-r_n}, \qquad (5.40)$$

где C в силу (5.37) от R не зависит.

Как и ранее, $R=0$ означает вектор, все компоненты которого равны нулю.

Из равенства (5.40) получим

$$P(0,X)=C\prod_{n=1}^{N}\left\{\int_0^{x_n}(1-A_n(y))dy\right\},$$

откуда

$$C = P(0, X) \Bigg/ \prod_{n=1}^{N} \left\{ \int_0^{x_n} (1 - A_n(y)) dy \right\},$$

и окончательно

$$P(R, X) = P(0, X) \prod_{n=1}^{N} \left\{ \frac{\int_0^{x_n} (1 - B_n(y)) dy}{\int_0^{x_n} (1 - A_n(y)) dy} \right\}^{r_n} . \qquad (5.41)$$

Обозначив $P(R) = P(R, \infty)$ и положив в (5.41) $X = \infty$, получим равенство

$$P(R) = P(0) \prod_{n=1}^{N} \left\{ \frac{b_n}{a_n} \right\}^{r_n} , \qquad (5.42)$$

совпадающее с формулой (5.35) для $P(R)$ предыдущего раздела, если обозначить

$$\rho_n = \frac{b_n}{a_n} .$$

Из условия нормировки аналогично найдем

$$P(0) = 1 \Bigg/ \sum_{R \in G} \prod_{n=1}^{N} \rho_n^{r_n} ,$$

также совпадающее с $P(0)$ из формул (5.35) предыдущего раздела.

Отметим, что распределение вероятностей $P(R)$ состояний R рассматриваемой немарковской СМО не зависит от вида функций распределения времени генерирования $A_n(x)$ и времени обслуживания заявок $B_n(x)$, а определяется только их средними значениями

$$a_n = \int_0^{\infty} (1 - A_n(y)) dy , \quad b_n = \int_0^{\infty} (1 - B_n(y)) dy .$$

Равенство (5.42) назовем формулами Энгсета для немарковских СМО.

§ 5.9. Некоторые частные случаи формул Энгсета

Таким образом, для рассматриваемых систем массового обслуживания получены формулы (5.42)

$$\begin{cases} P(R) = P(0)\prod\limits_{n=1}^{N}\rho_n^{r_n}, \\ P(0) = 1\Big/\sum\limits_{R\in G}\prod\limits_{n=1}^{N}\rho_n^{r_n}. \end{cases} \tag{5.43}$$

определяющие распределение вероятностей $P(R)$ состояний системы $R=(r_1,r_2,...,r_N)$, где r_n – состояние n-го источника.

Рассмотрим некоторые следствия, вытекающие из формул (5.43).

5.9.1. Формулы Энгсета для СМО с одинаковыми источниками

Так как распределение (5.43) определяется только величинами ρ_n, то источники, для которых эти величины совпадают, будем называть одинаковыми.

Рассмотрим СМО, в которой N_1 источников первого типа, характеризуемые величиной $\rho_1=\gamma_1$, N_2 источников второго типа с параметром $\rho_2=\gamma_2$, N_k источников типа k с параметром $\rho_k=\gamma_k$, и пусть $k=1,2,...,K$, т.е. в системе K типов источников.

Состояние системы определим вектором

$$I=(i_1,i_2,...,i_k,\ ...,\ i_K),$$

где i_k – число k-источников, заявки которых в данный момент находятся на обслуживании.

От распределения $P(R)$, определяемого равенством (5.43), перейдем к распределению

$$\pi(I)=P(i_1(t)=i_1,i_2(t)=i_2,...,i_K(t)=i_K)=\sum_{\substack{R\in G\\ r_1+r_2+...+r_{N_1}=i_1\\ r_{N_1+1}+...+r_{N_1+N_2}=i_2\\ ...\\ r_{N_1+...+N_{k-1}+1}+...+r_{N_1+N_2+...+N_K}=i_K}}P(R)=$$

$$=P(0)\prod_{k=1}^{K}C_{N_k}^{i_k}\gamma_k^{i_k}.$$

Пусть $\gamma_k=\dfrac{\rho_k}{N_k}$, т.е. суммарная интенсивность поступления заявок от всей группы одинаковых k-источников не зависит от N_k – числа источников в группе, тогда

$$\pi(I)=\pi(0)\prod_{k=1}^{K}C_{N_k}^{i_k}\frac{\rho_k^{i_k}}{N_k^{i_k}}. \qquad (5.44)$$

5.9.2. Формулы Эрланга для однородных немарковских систем обслуживания

В равенстве (5.44) положим $k=1$ и обозначим $\rho_1=\rho$, $N_1=N$. Тогда получим

$$\pi(i)=\pi(0)C_N^i\frac{\rho^i}{N^i}. \qquad (5.45)$$

Пусть в блоке обслуживания имеется J полнодоступных приборов, на каждом из которых может обслуживаться только одна заявка. Когда заняты все J приборов, поступившая заявка получает отказ в обслуживании, так что $0\le i\le J$. Это равенство определяет множество G существенных состояний, поэтому

$$\pi(0)=1\Big/\sum_{i=0}^{J}C_N^i\frac{\rho^i}{N^i}. \qquad (5.46)$$

Рассмотрим в (5.45) и (5.46) предельное соотношение при $N\to\infty$:

$$\lim_{N\to\infty}C_N^i\frac{1}{N^i}=\lim_{N\to\infty}\frac{N!}{i!(N-i)!N^i}=\lim_{N\to\infty}\frac{1}{i!}\frac{N(N-1)(N-2)\ldots(N-i+1)}{N^i}=\frac{1}{i!}.$$

Следовательно,

$$\begin{cases}\pi(0)=1\Big/\left\{\sum\limits_{i=0}^{J}\rho^i/i!\right\},\\ \pi(i)=\dfrac{\rho^i/i!}{\sum\limits_{k=0}^{J}\rho^k/k!}.\end{cases} \qquad (5.47)$$

Полученные формулы определяют распределение вероятностей $\pi(i)$ числа занятых приборов в J-линейной немарковской СМО, на вход которой поступает простейший, в силу предельной теоремы Григелиониса для суммы редких потоков, поток заявок, то есть (5.47) дают формулы Эрланга для немарковской СМО.

5.9.3. Формулы Эрланга для неоднородных немарковских систем обслуживания

Рассмотрим формулы (5.44) и положим в них все $N_k \to \infty$. Пусть блок обслуживания, на вход которого поступает K простейших потоков заявок, аналогично предыдущему случаю содержит J полнодоступных обслуживающих приборов.

В этом случае (5.44) при $N_k \to \infty$ примет вид

$$\pi(I) = \pi(0)\prod_{k=1}^{K} \rho_k^{i_k} / i_k!,$$

$$\pi(0) = 1 \Big/ \sum_{0 \le i_1 + i_2 + \ldots + i_k \le J}^{J} \prod_{k=1}^{K} \rho_k^{i_k} / i_k!.$$

5.9.4. Формулы Эрланга для заявок, требующих случайное число приборов

Рассмотрим J-линейную СМО, на вход которой поступает простейший с параметром λ поток заявок, каждая из которых с вероятностью $q(j)$ требует для своего обслуживания j приборов, $j = 1,2,\ldots,J$. Если в системе недостаточно свободных приборов, то поступившая заявка получает отказ в обслуживании, в противном случае она занимает все выделенные приборы, на время обслуживания имеющие функцию распределения $B_j(x)$. Зависимость $B_j(x)$ от j может быть достаточно произвольной. Освобождение всех j приборов, занятых одной заявкой, происходит одновременно в момент окончания ее обслуживания.

Состояние системы определим вектором

$$I = (i_1, i_2, \ldots i_J),$$

где i_j – число обслуживаемых j-заявок, каждая из которых занимает j приборов.

Для нахождения распределения вероятностей $\pi(I)$ воспользуемся формулой (5.44). Пусть в системе имеется N_1 источник, заявки которого требуют для своего обслуживания по одному прибору, N_2 источников, для заявок которого необходимо по 2 прибора, N_j источников по j приборов. Здесь $j = 1,2,\ldots,J$.

В формулах (5.44) положим

$$\rho_j = \lambda q(j) b_j,$$

где b_j – среднее время обслуживания j-заявки. Равенство (5.44) перепишем в виде

$$\pi(I) = \pi(0) \prod_{j=1}^{J} C_{N_j}^{i_j} \frac{\rho_j^{i_j}}{N_j^{i_j}}$$

и рассмотрим $N_j \to \infty$ для всех j. Тогда получим

$$\pi(I) = \pi(0) \prod_{j=1}^{J} \rho_j^{i_j} \Big/ i_j!\,. \qquad (5.48)$$

Для нахождения $\pi(0)$ необходимо определить множество G существенных состояний I.

Очевидно, I будет существенным состоянием, если для его компонент выполнено условие

$$i_1 + 2i_2 + 3i_3 + \ldots + Ji_J \le N,$$

поэтому

$$\pi(0) = 1 \Bigg/ \sum_{\substack{i_1 i_2 i_J \\ i_1 + 2i_2 + \ldots + J\, i_J \le N}}^{J} \prod_{k=1}^{K} \rho_j^{i_j} \Big/ i_j!\,. \qquad (5.49)$$

Формулы (5.48) и (5.49) решают поставленную задачу. Их можно назвать формулами Эрланга для заявок, требующих случайное число обслуживающих приборов.

Глава 6. ПРИОРИТЕТНЫЕ СМО. ЭРГОДИЧЕСКИЕ МЕТОДЫ

§ 6.1. Дисциплины обслуживания

До сих пор мы рассматривали СМО с одним входящим потоком заявок. Теперь мы перейдем к рассмотрению СМО с несколькими входящими потоками заявок.

Пусть на СМО поступают n пуассоновских потоков заявок с интенсивностью $\lambda_1, \lambda_2, ..., \lambda_n$. Эти заявки разные, т.е. они требуют разного времени обслуживания. Будем считать, что функция распределения времени обслуживания заявки i-го типа есть $B_i(x)$, $i = \overline{1,n}$. Если обслуживающий прибор 1, то получающая СМО обозначается как $\vec{M}_n / \vec{G}_n / 1 / \infty$.

Прежде всего возникает вопрос, как быть с заявками разного типа. Простейший вариант ставить их в общую очередь. Однако он не самый лучший, потому что заявки разного типа могут иметь разную значимость для системы. Поэтому более естественно ставить заявки разного типа в разные очереди. Возникающая картина имеет следующий вид (рис. 6.1):

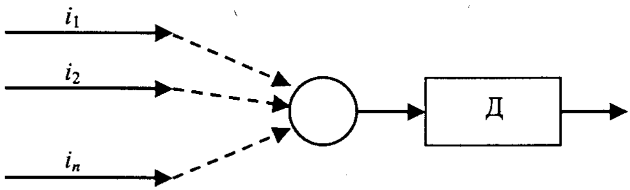


Рис. 6.1. Блок-схема приоритетной СМО

Теперь между прибором и очередями возникает фигура диспетчера который управляет поступлением заявок на прибор. Возникающие здесь **дисциплины диспетчеризации** очень разнообразны, но основную роль играют так называемые приоритетные дисциплины.

Пусть первый поток имеет приоритет 1, второй поток – приоритет 2,…, n-й поток – приоритет n. Основные дисциплины следующие.

Абсолютные приоритеты

Пусть на обслуживающем приборе находится заявка, имеющая k-й приоритет. Если в систему поступает заявка с приоритетом 1, 2, ..., (k–1) (как говорят, заявка с более высоким приоритетом), то обслуживание заявки **прерывается** и ее место занимает поступившая заявка. При этом возможны следующие основные варианты:

Абсолютные приоритеты с дообслуживанием

Когда заявки более высокого приоритета покинут систему, заявка, чье обслуживание было прервано, **дообслуживается**, то есть общее время пребывания заявки на приборе, несмотря на перерывы в обслуживании, все равно имеет функцию распределения $B_k(x)$.

Абсолютные приоритеты с обслуживанием заново

В этом случае после окончания обслуживания заявок более высокого приоритета обслуживание прерванной заявки начинается заново, так что заявка покинет систему только в том случае, если она будет обслуживаться непрерывно от начала до конца.

Есть и другие варианты абсолютных приоритетов, но приведенные выше, – основные.

Относительные приоритеты

В этом случае обслуживание заявки, находящейся на приборе, **не может быть прервано**, т.е. поступившая на прибор заявка обслуживается до конца. Но по окончании ее обслуживания на прибор берется заявка из очереди с **наименьшим номером**, то есть заявка самого высокого приоритета из имеющихся в наличии. Внутри очереди из заявок с одним приоритетом применяется дисциплина FIFO или LIFO.

Динамические приоритеты

В этом случае выбор на обслуживание заявки с тем или иным номером зависит от длин очередей ($i_1(t)$, $i_2(t)$, ..., $i_n(t)$), имеющихся в системе в момент освобождения прибора.

В данном курсе мы ограничимся только этими дисциплинами, хотя имеются и другие.

Прежде чем переходить к исследованию, заметим следующее. Системы с абсолютным и относительным приоритетами могут быть рассчитаны точно, то есть получены формулы для Z-преобразования от финальных вероятностей и преобразования Лапласа – Стилтьеса от времени ожидания. Однако эти точные формулы чрезвычайно громоздки, и их вывод также очень громоздок. Однако основные характеристики таких систем могут быть получены гораздо проще, хотя и недостаточно строго, исходя из так называемых **эргодических соображений,** в которых предполагается равенство средних по ансамблю и средних по времени. Именно этот путь и избран в настоящем курсе.

§ 6.2. Системы с общей очередью

Выведем прежде всего выражение для загрузки нашего прибора. Рассуждения здесь выглядят следующим образом.

Рассмотрим очень большой интервал времени T (точнее, рассматривается асимптотика $\mathrm{T} \to \infty$). За этот интервал в среднем в систему поступит $\lambda_i \mathrm{T}$ заявок i-го типа. Если b_i есть среднее время обслуживания заявки i-го типа ($b_i = \int_0^\infty x dB_i(x)$), то прибор будет занят обслуживанием этих заявок в среднем $\lambda_i \mathrm{T} \cdot b_i$ времени. Всего, если пренебречь «краевыми эффектами» (это можно сделать в силу того, что $\mathrm{T} \to \infty$), прибор занят обслуживанием заявок время $\sum_{i=1}^{n} \lambda_i b_i \mathrm{T}$. Так как весь интервал времени равен T, то вероятность того, что, входя в систему в произвольный момент времени, мы застанем прибор занятым обслуживанием заявки i-го типа, $\rho_i = \dfrac{\lambda_i \mathrm{T} \cdot b_i}{\mathrm{T}} = \lambda_i b_i$. Вероятность застать прибор занятым «вообще», то есть загрузка прибора равна $R_n = \sum_{i=1}^{n} \lambda_i b_i \mathrm{T} \Big/ \mathrm{T} = \sum_{i=1}^{n} \lambda_i b_i = \sum_{i=1}^{n} \rho_i$, а вероятность застать прибор свободным равна $1 - R_n$. Очевидно, что система будет работоспособной при условии $R_n < 1$, иначе очереди в системе будут нарастать до бесконечности.

Пусть теперь v_i – среднее время пребывания в системе заявки i-го типа; w_i – среднее время ожидания в очереди начала обслуживания заявок i-го типа; l_i – среднее число находящихся в системе заявок i-го типа; q_i – среднее число находящихся в очереди заявок i-го типа.

Тогда по формуле Литтла

$$l_i = \lambda_i v_i . \tag{6.1}$$

Далее, так как вероятность застать прибор занятым заявкой i-го типа равна ρ_i, то $l_i = q_i + \rho_i$. Очевидно, что среднее время пребывания в системе складывается из среднего времени ожидания и среднего времени обслуживания b_i, т.е. $v_i = w_i + b_i$. Поэтому

$$l_i = \lambda_i v_i = \lambda_i w_i + \lambda_i b_i = q_i + \rho_i , \tag{6.2}$$

то есть верна еще одна формула Литтла

$$q_i = \lambda_i w_i . \tag{6.3}$$

Рассмотрим еще вопрос о среднем времени дообслуживания заявки. Пусть, входя в систему, мы застаем на приборе заявку i-го типа. Так как время, прошедшее с момента нашего входа в систему до момента окончания обслуживания есть то, что мы называем «временем перескока», то время дообслуживания заявки имеет функцию распределения

$$\frac{1}{b_i}\int_0^x \left(1 - B_i(z)\right) dz .$$

Поэтому среднее время дообслуживания заявки равно

$$\sigma_i = \frac{1}{b_i}\int_0^\infty x d\left(\int_0^x \left(1 - B_i(z)\right) dz\right) = \frac{1}{b_i}\int_0^\infty x\left(1 - B_i(x)\right) dx =$$

$$= \frac{1}{b_i}\frac{x^2}{2}\left(1 - B_i(x)\right)\Big|_0^\infty + \frac{1}{2b_i}\int_0^\infty x^2 dB_i(x) = \frac{b_i^{(2)}}{2b_i} , \tag{6.4}$$

где было проделано интегрирование по частям и $b_i^{(2)} = \int_0^\infty x^2 dB_i(x)$.

Система с общей очередью

Чтобы стало ясно дальнейшее, рассмотрим сначала уже известную нам систему $M|G|1|\infty$. Пусть w – среднее время ожидания заявки в очереди. В момент поступления заявки в систему она застанет в очереди в среднем $q = \lambda w$ заявок, а на приборе в среднем ρ заявок.

Перед тем, как наша заявка поступит на прибор: 1) должна быть дообслужена заявка, находящаяся на приборе, что займет в среднем $\rho \cdot \sigma$

единиц времени; 2) должны быть обслужены все заявки, находящиеся в очереди, что займет в среднем $\lambda w \cdot b$ единиц времени. Поэтому

$$w = \rho \cdot \sigma + w \cdot \lambda b, \tag{6.5}$$

откуда

$$w = \frac{\rho \cdot \sigma}{1 - \lambda b} = \frac{\lambda b^{(2)}}{2(1 - \lambda b)}, \tag{6.6}$$

что мы в свое время получали и раньше, но гораздо более сложным путем.

Вернемся теперь к СМО $\vec{M}_n | \vec{G}_n | 1 | \infty$ с общей очередью. Поступающая в систему заявка с вероятностью ρ_i застанет прибор занятым обслуживанием заявки i-го типа, среднее время дообслуживания которой равно σ_i. С вероятностью R_n прибор будет занят. Так как среднее время ожидания не зависит от типа заявки и складывается из среднего времени дообслуживания и из среднего времени обслуживания заявок, стоящих в очереди впереди, то

$$w = \sum_{i=1}^{n} q_i b_i + \sum_{i=1}^{n} \rho_i \sigma_i = \sum_{i=1}^{n} \lambda_i w b_i + \sum_{i=1}^{n} \rho_i \sigma_i = w R_n + \sum_{i=1}^{n} \rho_i \sigma_i = w R_n + \frac{1}{2} \sum_{i=1}^{n} \lambda_i b_i^{(2)}, \tag{6.7}$$

откуда

$$w = \frac{\frac{1}{2} \sum_{i=1}^{n} \lambda_i b_i^{(2)}}{1 - R_n}. \tag{6.8}$$

По формулам Литтла отсюда получаем

$$q_i = \lambda_i w, \qquad v_i = w + b_i, \qquad l_i = q_i + \rho_i, \qquad i = \overline{1, n},$$

что и определяет важнейшие средние характеристики в рассматриваемой СМО.

§ 6.3. Системы с относительными приоритетами

В этом случае среднее время обслуживания заявки i-го типа равно b_i, так как, начавшись, обслуживание продолжается до его завершения. Поэтому вероятность застать на обслуживании заявку i-го типа равна по-прежнему $\rho_i = \lambda_i b_i$.

Пусть в СМО поступает заявка, имеющая j-й приоритет ($j=\overline{1,n}$). Проследим за движением этой заявки. Перед тем, как ей попасть на прибор: 1) должно быть закончено обслуживание той заявки, которая находится на приборе; 2) должны быть обслужены все имеющиеся заявки из очередей с номерами от 1 до j, а также 3) те заявки, которые имеют приоритет от 1 до $j-1$ и которые поступили за время ожидания в очереди поступившей заявки j-го типа. Соответствующие слагаемые будут иметь вид

$$w_j=\sum_{i=1}^{n}\rho_i\sigma_i+\sum_{i=1}^{j}\underbrace{\lambda_i w_i}_{q_i}b_i+w_j\sum_{i=1}^{j-1}\lambda_i b_i\,. \tag{6.9}$$

Последнее слагаемое выписывается из тех соображений, что за время ожидания w_j в СМО поступило в среднем $w_j\lambda_i$ заявок i-го типа, которые обслуживались время $w_j\lambda_i\cdot b_i$.

Беря $j=1$, получим

$$w_1=\sigma+\underbrace{w_1\lambda_1}_{q_1}\cdot b_1\,,$$

откуда

$$w_1=\frac{\sigma}{1-R_1}\,, \tag{6.10}$$

где $\sigma=\sum_{i=1}^{n}\rho_i\sigma_i=\frac{1}{2}\sum_{i=1}^{n}\lambda_i b_i^{(2)}$ и $R_j=\sum_{i=1}^{j}\lambda_i b_i$.

Докажем теперь, пользуясь методом математической индукции, соотношение (считая $R_0=0$)

$$w_j=\frac{\sigma}{(1-R_{j-1})(1-R_j)}\,. \tag{6.11}$$

Заметим, что для $j=1$ формула (6.11) верна. Предположим теперь, что (6.11) верно для всех $i<j$. Перепишем (6.9) в виде

$$w_j=\sum_{i=1}^{n}\rho_i\sigma_i+\sum_{i=1}^{j-1}w_i\rho_i+w_j\sum_{i=1}^{j}\lambda_i b_i\,,$$

откуда получим рекуррентное соотношение

$$w_j=\left(\sigma+\sum_{i=1}^{j-1}\rho_i w_i\right)\Big/(1-R_j)\,.$$

Подставим сюда вместо w_i выражение (6.11), считая $R_0 = 0$:

$$w_j = \frac{\sigma}{1-R_j}\left(1+\sum_{i=1}^{j-1}\frac{\rho_i}{(1-R_{i-1})(1-R_i)}\right). \tag{6.12}$$

Но так как $\rho_i = R_i - R_{i-1}$, то

$$\frac{\rho_i}{(1-R_{i-1})(1-R_i)} = \frac{(1-R_{i-1})-(1-R_i)}{(1-R_{i-1})(1-R_i)} = \frac{1}{1-R_i} - \frac{1}{1-R_{i-1}},$$

поэтому

$$\sum_{i=1}^{j-1}\frac{\rho_i}{(1-R_{i-1})(1-R_i)} = \sum_{i=1}^{j-1}\left[\frac{1}{1-R_i}-\frac{1}{1-R_{i-1}}\right] = \left(\frac{1}{1-R_1}-\frac{1}{1-R_0}\right)+$$

$$+\left(\frac{1}{1-R_2}-\frac{1}{1-R_1}\right)+\ldots+\left(\frac{1}{1-R_{j-1}}-\frac{1}{1-R_{j-2}}\right) = \frac{1}{1-R_{j-1}} - 1 = \frac{R_{j-1}}{1-R_{j-1}},$$

и формула (6.12) приобретет вид $w_j = \sigma/(1-R_j)(1-R_{j-1})$, что и доказывает формулу (6.11).

Выражение для w_j позволяет находить все остальные интересующие нас величины – v_i, q_i, l_i по формулам Литтла.

Из формулы (6.11) следует формула, которая получила название «закона сохранения работы». Вычислим величину $\sum_{i=1}^{n}\rho_i w_i$. Имеем

$$\sum_{i=1}^{n}\rho_i w_i = \sigma\sum_{i=1}^{n}\frac{\rho_i}{(1-R_j)(1-R_{j-1})} = \sigma\frac{R_n}{1-R_n} = \left(\frac{1}{2}\sum_{i=1}^{n}\lambda_i b_i^{(2)}\right)\frac{R_n}{1-R_n}.$$

Заметим, что правая часть этой формулы не зависит от того, какие приоритеты назначены заявкам того или иного типа. Поэтому

$$\sum_{i=1}^{n}\rho_i w_i = \text{const}, \tag{6.13}$$

где const понимается как независимость от назначаемых приоритетов. Так как $\rho_i = \lambda_i b_i$, а $\lambda_i w_i = q_i$, то этот же закон можно переписать в виде

$$\sum_{i=1}^{n} q_i b_i = \text{const}. \tag{6.14}$$

Он и называется «законом сохранения работы», так как слагаемое $q_i b_i$ имеет смысл времени (работы), которое надо потратить на обслуживание находящихся в системе заявок i-го типа.

Оптимальное назначение относительного приоритета

Рассмотрим теперь вопрос о назначении приоритетов, то есть о том, какой приоритет назначить заявкам того или иного типа. Здесь приходится начинать с вопроса о выборе критерия эффективности работы СМО.

Будем считать, что от пребывания в СМО заявки типа i общество несет потери, равные C_i в единицу времени. Тогда так как среднее число заявок i-го типа равно $\lambda_i v_i$, средние потери в единицу времени равны

$$L = \sum_{i=1}^{n} C_i \lambda_i v_i = \sum_{i=1}^{n} C_i \lambda_i w_i + \sum_{i=1}^{n} C_i \rho_i . \tag{6.15}$$

Начнем с системы $\vec{M}_2 | \vec{G}_2 | 1 | \infty$, где все наиболее ясно. В СМО поступает два потока заявок – первый поток интенсивности λ_1 со стоимостью пребывания заявки в СМО C_1 в единицу времени, и второй поток с интенсивностью λ_2 со стоимостью пребывания заявки C_2 в единицу времени.

Если первому потоку назначить высший приоритет, а второму низший, то средние очереди в СМО будут q_1 и q_2 и средние потери $L = C_1 q_1 + C_2 q_2$. Если приоритеты назначить наоборот, то длины очередей станут $\bar{q}_1$ и $\bar{q}_2$, а средние потери $\bar{L} = C_1 \bar{q}_1 + C_2 \bar{q}_2$. Найдем условия, при которых

$$C_1 q_1 + C_2 q_2 \le C_1 \bar{q}_1 + C_2 \bar{q}_2 . \tag{6.16}$$

В силу закона сохранения работы независимо от правила назначения приоритетов $b_1 q_1 + b_2 q_2 = \text{const}$, т.е. $b_1 q_1 + b_2 q_2 = b_1 \bar{q}_1 + b_2 \bar{q}_2$. Поэтому разделим левую часть (6.16) на $b_1 q_1 + b_2 q_2$, а правую – на $b_1 \bar{q}_1 + b_2 \bar{q}_2$:

$$\frac{C_1 q_1 + C_2 q_2}{b_1 q_1 + b_2 q_2} \le \frac{C_1 \bar{q}_1 + C_2 \bar{q}_2}{b_1 \bar{q} + b_2 \bar{q}_2} .$$

Приводя к общему знаменателю, после элементарных преобразований получим

$$b_2 C_1 (\bar{q}_1 q_2 - q_1 \bar{q}_2) \ge b_1 C_2 (q_2 \bar{q}_1 - q_1 \bar{q}_2) . \tag{6.17}$$

Ясно, что $q_1 < \bar{q}_1$ (ведь q_1 получается, когда первому потоку представляется приоритет), а $q_2 > \bar{q}_2$ (ведь $\bar{q}_2$ получается, когда второму потоку предоставляется приоритет). Поэтому $\overline{q}_1 q_2 > q_1 \overline{q}_2$ и выражение в скобках положительно. Поэтому условие (6.17) переходит в условие $C_1 b_2 \geq C_2 b_1$ или в условие $\dfrac{C_1}{b_1} \geq \dfrac{C_2}{b_2}$. Таким образом, в СМО с двумя потоками событий более высокий приоритет следует назначить тому потоку, у которого больше отношение C_i/b_i.

Покажем, что это правило верно и для СМО с произвольным числом входящих потоков, т.е. оптимальное правило выглядит следующим образом: все типы заявок упорядочить по убыванию отношения C_i/b_i и назначать приоритеты так, чтобы $C_1/b_1 \geq C_2/b_2 \geq C_3/b_3 \geq \ldots \geq C_n/b_n$.

Действительно, пусть мы имеем СМО, в которой потоки пронумерованы в порядке убывания относительных приоритетов и **которая минимизирует величину** L. Представим себе, что мы поменяли приоритеты лишь j-го и $j+1$-го потоков. Так как остальных потоков это не коснулось, то должно получиться

$$\sum_{i=1}^{j-1} C_i \lambda_i w_i + C_j \lambda_j w_j + C_{j+1} \lambda_{j+1} w_{j+1} + \sum_{i=j+2}^{n} C_i \lambda_i w_i \leq$$

$$\leq \sum_{i=1}^{j-1} C_i \lambda_i w_i + C_j \lambda_j \overline{w}_j + C_{j+1} \lambda_{j+1} \overline{w}_{j+1} + \sum_{i=j+2}^{n} C_i \lambda_i w_i. \qquad (6.18)$$

Уничтожая одинаковые слагаемые, получим

$$C_j q_j + C_{j+1} q_{j+1} \leq C_{j+1} \overline{q}_{j+1} + C_j \overline{q}_j. \qquad (6.19)$$

В силу закона сохранения работы

$$\sum_{i=1}^{j-1} b_i q_i + b_j q_j + b_{j+1} q_{j+1} + \sum_{i=j+2}^{n} b_i q_i = \sum_{i=1}^{j-1} b_i q_i + b_j \overline{q}_j + b_{j+1} \overline{q}_{j+1} + \sum_{i=j+2}^{n} b_i q_i, \qquad (6.20)$$

то есть снова верно условие $b_j q_j + b_{j+1} q_{j+1} = b_j \overline{q}_j + b_{j+1} \overline{q}_{j+1}$. Таким образом, все свелось к прежнему случаю, откуда и получится условие $C_j/b_j \geq C_{j+1}/b_{j+1}$.

Замечательно, что оптимальное правило назначения приоритетов не зависит от интенсивностей входящих потоков заявок λ_i.

Сравнение приоритетной СМО и СМО с общей очередью

Пусть $\dfrac{C_1}{b_1} \geq \dfrac{C_2}{b_2}$. Подсчитаем потери в приоритетной СМО и в СМО с общей очередью. Имеем

$$L_{\text{пр}} = C_1\lambda_1 w_1 + C_2\lambda_2 w_2 = \frac{C_1\lambda_1\sigma}{1-\rho_1} + \frac{C_2\lambda_2\sigma}{(1-\rho_1)(1-\rho_1-\rho_2)},$$

$$L_{\text{общ}} = C_1\lambda_1 w_1 + C_2\lambda_2 w_2 = \frac{C_1\lambda_1\sigma}{1-\rho_1-\rho_2} + \frac{C_2\lambda_2\sigma}{1-\rho_1-\rho_2}.$$

Отсюда получаем, что

$$L_{\text{общ}} - L_{\text{пр}} = \frac{C_1\lambda_1\sigma}{1-\rho_1-\rho_2} + \frac{C_2\lambda_2\sigma}{1-\rho_1-\rho_2} - \frac{C_1\lambda_1\sigma}{1-\rho_1} - \frac{C_2\lambda_2\sigma}{(1-\rho_1)(1-\rho_1-\rho_2)} =$$

$$= \frac{\sigma}{(1-\rho_1)(1-\rho_1-\rho_2)}\left[C_1\lambda_1(1-\rho_1) + C_2\lambda_2(1-\rho_1) - C_1\lambda_1(1-\rho_1-\rho_2) - C_2\lambda_2\right] =$$

$$= \frac{\sigma}{(1-\rho_1)(1-\rho_1-\rho_2)}\left[C_1\lambda_1 - C_1\lambda_1\rho_1 + C_2\lambda_2 - C_2\lambda_2\rho_1 - \right.$$

$$\left. - C_1\lambda_1 + C_1\lambda_1\rho_1 + C_1\lambda_1\rho_2 - C_2\lambda_2\right] = \frac{\sigma(C_1\lambda_1\rho_2 - C_2\lambda_2\rho_1)}{(1-\rho_1)(1-\rho_1-\rho_2)} =$$

$$= \frac{\sigma(C_1\lambda_1\lambda_2 b_2 - C_2\lambda_2\lambda_1 b_1)}{(1-\rho_1)(1-\rho_1-\rho_2)} = \frac{\sigma\lambda_1\lambda_2 b_1 b_2}{(1-\rho_1)(1-\rho_1-\rho_2)}\left[\frac{C_1}{b_1} - \frac{C_2}{b_2}\right] \geq 0.$$

Таким образом, приоритетная СМО всегда лучше СМО с общей очередью, разумеется, только в смысле выбранного критерия оптимальности. Общественные и иные аспекты мы не рассматриваем.

§ 6.4. Абсолютные приоритеты с дообслуживанием

Пусть в системе $\vec{M}_n \mid \vec{G}_n \mid 1 \mid \infty$ потоки занумерованы в порядке убывания их приоритетов. Будем говорить, что в системе действуют абсолютные приоритеты с дообслуживанием, если поступающая в СМО i-заявка замещает на приборе j-заявку при $i < j$. Будем считать, что j-заявка возвращается в этом случае в начало j-й очереди и вновь поступит на обслуживание, когда будут обслужены все заявки с номерами от

1 до $j-1$. Так как дообслуживание продолжается с прерванного места, то чистое время ее обслуживания не зависит от числа прерываний.

Выведем некоторые соотношения, необходимые нам для дальнейшего.

Структура цикла обслуживания

Пусть на обслуживание поступает заявка k-го типа со средним временем обслуживания b_k. Пусть (для общности) ее обслуживание может быть прервано заявками с приоритетом 1, 2, 3, ..., $i<k$.

Тогда цикл обслуживания этой заявки выглядит так (рис. 6.2):

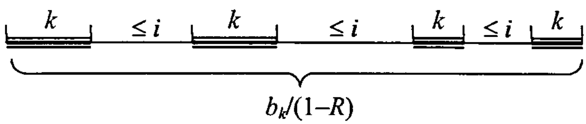


Рис. 6.2. Структура цикла обслуживания

где через обозначены периоды обслуживания самой k-заявки, а через $\le i$ – перерывы в обслуживании, вызванные приходом заявок с приоритетами от 1 до i. Легко догадаться, что эти перерывы есть периоды занятости системы этими приоритетными заявками. Рассчитаем их длину.

Пусть $\Lambda_i=\sum_{s=1}^{i}\lambda_s$, $R_i=\sum_{s=1}^{i}\rho_s$, T – большой интервал времени.

Расчет средней длины периода занятости

$\mathrm{T}R_i$ – время, в течение которого СМО была занята обслуживанием заявок с приоритетами $\overline{1,i}$;

$\mathrm{T}(1-R_i)$ – время, в течение которого на приборе не было этих заявок.

Так как Λ_i есть суммарная интенсивность потока заявок с приоритетами $\overline{1,i}$, то за время $\mathrm{T}(1-R_i)$ пришло $\mathrm{T}(1-R_i)\Lambda_i$ этих заявок, каждая из которых начала период занятости, т.е. за время T было всего $\mathrm{T}(1-R_i)\Lambda_i$ периодов занятости.

Поэтому средняя длительность периода $\le i$ равна

$$\frac{\mathrm{T}R_i}{\mathrm{T}(1-R_i)\Lambda_i}=\frac{R_i}{(1-R_i)\Lambda_i} \qquad (6.21)$$

– средняя длительность периода занятости $\le i$ заявок.

Расчет средней длины периода обслуживания

Так как на интервале длиной b_k пришло в среднем $b_k\Lambda_i$ заявок типа $\overline{1,i}$, то суммарная длина периодов $\underline{\ \le i\ }$ на одну k-заявку составила

$$b_k\Lambda_i\frac{R_i}{(1-R_i)\Lambda_i}=\frac{b_kR_i}{1-R_i} \qquad (6.22)$$

– суммарная длина i-периодов на одну k-заявку.

Поэтому средняя длительность цикла обслуживания заявки k-го типа есть

$$b_k+\frac{b_kR_i}{1-R_i}=\frac{b_k}{1-R_i} \qquad (6.23)$$

– средняя длина цикла обслуживания k-заявки с прерыванием $1-i$-заявками.

Связь среднего времени ожидания и среднего времени пребывания заявки в СМО

Пусть w_j есть среднее время ожидания в очереди j-заявки до начала обслуживания, а v_j – ее среднее время пребывания в системе. Тогда, так как обслуживание j-заявки прерывается заявками с приоритетами от 1 до $j-1$, то

$$v_j=w_j+\frac{b_j}{1-R_{j-1}}. \qquad (6.24)$$

Отсюда $l_j=\lambda_j q_j$, $q_j=\lambda_j w_j$, и нам осталось найти лишь w_j.

Среднее время ожидания в очереди

Обозначим $\sigma_i=b_i^{(2)}/2b_i$, $\Delta_j=\sum\limits_{i=1}^{j}\rho_i\sigma_i$.

Рассмотрим движение заявки j-го типа, поступающей в СМО. Ее среднее время ожидания в очереди складывается из:

А) Среднего времени дообслуживания заявки, находящейся на приборе. Так как при этом не стоит учитывать заявки с приоритетами, большими j (они будут вытеснены этой заявкой), то это среднее время равно Δ_j.

Б) Среднего времени обслуживания заявок приоритета $\le j$, которые были «вытеснены» обратно в очередь той заявкой, которая находится на обслуживании сейчас. Обозначим это время через η_j.

В) Среднего времени обслуживания тех заявок, которые находятся в очереди с приоритетами $\overline{1, j}$.

Г) Среднего времени обслуживания тех заявок с приоритетами $\overline{1, j-1}$, которые придут в систему за время ожидания в очереди заявки j-типа.

Окончательно

$$w_j = \Delta_j + \eta_j + \sum_{i=1}^{j} \lambda_i w_i b_i + w_j \sum_{i=1}^{j-1} \lambda_i b_i . \qquad (6.25)$$

От СМО с относительными приоритетами это выражение отличается слагаемым η_j. Вычислим его. Для этого найдем вероятность $p_{i\,k}$ – вероятность того, что в системе находится заявка k-типа, вытесненная с прибора заявкой i-типа ($i < k$). Пусть $P_{i\,k}$ есть вероятность того, что в системе находится заявка k-типа, вытесненная заявками с приоритетом $\overline{1, i}$.

Пусть имеется большой интервал T. За это время пришло $\lambda_k \mathrm{T}$ заявок k-типа. Суммарное время i-периодов $\lfloor\, \le i \,\rfloor$ составило $\lambda_k \mathrm{T} \dfrac{b_k}{1 - R_i}$.

Поэтому

$$P_{ik} = \lambda_k \mathrm{T} \frac{b_k R_i}{1 - R_i} \Big/ \mathrm{T} = \frac{\lambda_k b_k R_i}{1 - R_i} = \frac{\rho_k R_i}{1 - R_i} . \qquad (6.26)$$

Отсюда легко находится и $p_{i\,k}$:

$$p_{i\,k} = P_{i\,k} - P_{i-1,\,k} = \lambda_k b_k \left(\frac{R_i}{1 - R_i} - \frac{R_{i-1}}{1 - R_{i-1}} \right) = \frac{\lambda_k b_k \rho_i}{(1 - R_i)(1 - R_{i-1})} . \qquad (6.27)$$

Поэтому

$$\eta_j=\sum_{i=1}^{j-1}\sum_{k=i+1}^{j}p_{ik}\sigma_k=\sum_{i=1}^{j-1}\frac{\rho_i}{(1-R_i)(1-R_{i-1})}\sum_{k=i+1}^{j}\rho_k\sigma_k=\sum_{i=1}^{j-1}\frac{\rho_i(\Delta_j-\Delta_i)}{(1-R_i)(1-R_{i-1})}=$$

$$=\Delta_j\frac{R_{j-1}}{1-R_{j-1}}-\sum_{i=1}^{j-1}\frac{\rho_i\Delta_i}{(1-R_{i-1})(1-R_i)}.$$

Поэтому уравнение (6.25) приобретает вид

$$w_j=\frac{\Delta_j}{1-R_{j-1}}-\sum_{i=1}^{j-1}\frac{\rho_i\Delta_i}{(1-R_{i-1})(1-R_i)}+w_j\sum_{i=1}^{j}\lambda_i b_i+\sum_{i=1}^{j-1}\lambda_i w_i b_i. \qquad (6.28)$$

При $j=1$ имеем (считая $R_0=0$)

$$w_1=\Delta_1+w_1R_1,$$

откуда

$$w_1=\frac{\Delta_1}{1-R_1}=\frac{\Delta_1}{(1-R_0)(1-R_1)}.$$

Прямой подстановкой в (6.28) можно показать, что w_j имеет вид

$$w_j=\frac{\Delta_j}{(1-R_{j-1})(1-R_j)}, \qquad (6.29)$$

что по форме очень напоминает выражение для w_j при относительных приоритетах (отличие лишь в одном – вместо σ стоит Δ_j).

Что касается вопроса о назначении абсолютных приоритетов, то здесь можно сказать следующее. Для системы $\vec{M}_n\,|\,\vec{M}_n\,|1|\infty$ оптимальное правило назначения абсолютных приоритетов совпадает с правилом назначения относительных приоритетов. В общем случае системы $\vec{M}_n\,|\,\vec{G}_n\,|1|\infty$ правило назначения абсолютных приоритетов зависит от всех параметров $\left\{\lambda_i,b_i,b_i^{(2)},c_i,i=\overline{1,n}\right\}$ и не может быть сформулировано в простом виде.

Что касается сравнения всех этих систем, то для систем $\vec{M}_n\,|\,\vec{M}_n\,|1|\infty$ они располагаются в таком порядке: лучшая – СМО с абсолютными приоритетами, худшая – СМО с общей очередью. В общем случае СМО $\vec{M}_n\,|\,\vec{G}_n\,|1|\infty$ система с общей очередью всегда хуже приоритетных СМО, однако какая СМО – с абсолютными или относительными приоритетами – лучше зависит от параметров $\left\{ \lambda_i, b_i, b_i^{(2)}, c_i \right\}$.

§ 6.5. Системы с динамическими приоритетами

В приведенных выше ситуациях приоритеты заявкам каждого типа назначались заранее. Поэтому они называются еще статическими приоритетами.

Однако вполне возможна ситуация, когда выбор на обслуживание заявки того или иного типа зависит от ситуации, складывающейся в данный момент в СМО, например, от длин очередей, от времен ожидания в очередях и т.п. Такой выбор называется ситуационным или динамическим приоритетом.

Рассмотрим основные идеи исследования подобных систем на примере системы $\bar{M}_2 | M | 1 | N_1, N_2$, блок-схема которой изображена на рис. 6.3.

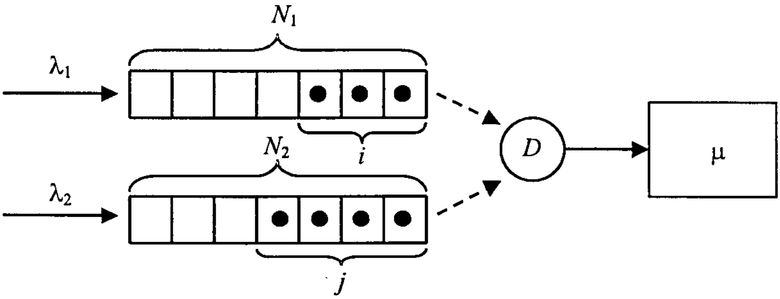


Рис. 6.3

Пусть в систему поступают два пуассоновских потока заявок с интенсивностями λ_1 и λ_2, которые становятся в отдельные очереди. Для простоты будем считать обслуживание экспоненциальным с одинаковой интенсивностью μ. Будем считать, что если в некоторый момент времени t в очередях находится i заявок первого типа и j заявок второго типа, то система в единицу времени несет потери, равные $F(i, j)$. Качество функционирования СМО описывается величиной $L = M\{F(i, j)\}$, где усреднение ведется по стационарному распределению вероятностей $\pi(i, j)$ состояний (i, j) рассматриваемой СМО.

Пусть в момент освобождения прибора в очередях было i заявок первого типа и j заявок второго типа. Обозначим через $\delta(i,j)$ вероятность взятия на обслуживание заявки **первого** типа в описанной выше ситуации. Очевидно, что вероятность взятия на обслуживание заявки **второго** типа равна $1-\delta(i,j)$. Величины $\delta(i,j)$ и определяют искомые динамические приоритеты.

Очевидно, что средние (в единицу времени) потери $L=M\{F(i,j)\}$ зависят от величин $\delta(i,j)$. Поэтому задача оптимизации рассматриваемой СМО приобретает вид

$$L=M\{F(i,j)\}\Rightarrow\min_{\{\delta(i,j)\}}. \tag{6.30}$$

Для решения этой задачи применяются в основном два метода.

Метод линейного программирования

Рассмотрим СМО $\tilde{M}_2|M|1|N_1,N_2$.

Обозначим через $\pi(i,j)$ финальную вероятность того, что в очередях находятся i заявок первого типа и j заявок второго типа. Тогда граф переходов соответствующих состояний с i, $j>0$ имеет вид, изображенный на рис 6.4.

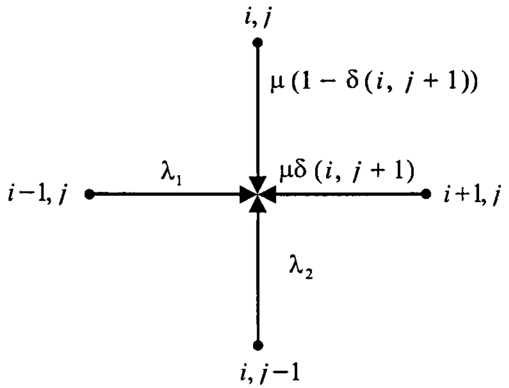


Рис. 6.4. Граф переходов в рассматриваемой СМО

Соответствующее ему уравнение для финальных вероятностей имеет вид

$$\lambda_1\pi(i-1,j)+\lambda_2\pi(i,j-1)+\mu\pi(i+1,j)\delta(i+1,j)+$$

$$+\mu\pi(i,j+1)\left[1-\delta(i,j+1)\right]-(\lambda_1+\lambda_2+\mu)\pi(i,j)=0. \quad (6.31)$$

Можно легко выписать и все уравнения для $\pi(i,j)$ для граничных состояний (то есть с $i=0$ или N_1, $j=0$ или N_2), но мы этого делать не будем. Сюда же следует добавить условие нормировки $\sum\limits_{i,j}\pi(i,j)=1$.

Средние потери тогда составят величину

$$L=\sum_{i,j}F(i,j)\pi(i,j)\Rightarrow\min_{\delta(i,j)}, \quad (6.32)$$

которая должна быть минимизирована по $\delta(i,j)$, удовлетворяющих ограничениям $0\le\delta(i,j)\le 1$.

Перейдем к величинам

$$x(i,j)=\pi(i,j)\delta(i,j),$$

$$y(i,j)=\pi(i,j)\left[1-\delta(i,j)\right], \quad (6.33)$$

так что $\pi(i,j)=x(i,j)+y(i,j)$; очевидно, что $x(i,j)\ge 0$, $y(i,j)\ge 0$. Тогда условия (6.31) – (6.33) примут вид

$$\sum_{i,j}F(i,j)\left[x(i,j)+y(i,j)\right]\Rightarrow\min_{\{x(i,j),y(i,j)\}},$$

$$\lambda_1\left[x(i-1,j)+y(i-1,j)\right]+\lambda_2\left[x(i,j-1)+y(i,j-1)\right]+$$

$$+\mu x(i+1,j)+\mu y(i,j+1)-(\lambda_1+\lambda_2+\mu)\left[x(i,j)+y(i,j)\right]=0,$$

$$\sum_{i,j}\left[x(i,j)+y(i,j)\right]=1,$$

$$x(i,j)\ge 0,\ y(i,j)\ge 0. \quad (6.34)$$

Мы получим типичную задачу линейного программирования. Если решить ее и найти $x(i,j)$ и $y(i,j)$, то легко найти и $\delta(i,j)$:

$$\delta(i,j)=x(i,j)/\left[x(i,j)+y(i,j)\right]. \quad (6.35)$$

Однако, пользуясь теорией линейного программирования, можно показать, что $\delta(i,j)$ принимает лишь значения 0 или 1.

Главная трудность задачи (6.34) заключается в ее большой размерности, равной $(N_1+1)(N_2+1)$. Для численного решения подобных задач, учитывающих специфику СМО, разрабатывались специальные численные методы (работы В.В. Мовы, Л.А. Понамаренко, А.М. Калиновского).

Метод динамического программирования

Другой метод численного исследования таких СМО разрабатывался А.А. Назаровым.

Рассмотрим ту же СМО $\vec{M}_2 \,|\, M \,|\, 1 \,|\, N_1, N_2$ и введем величину

$$Z(i,j,\alpha)=\alpha\int_t^\infty e^{-\alpha(\tau-t)}M\left\{F(i(\tau),j(\tau))\,|\,i(t)=i,j(t)=j\right\}d\tau\,. \quad (6.36)$$

В силу стационарности режима $Z(i,j,\alpha)$ от t не зависит. Кроме того, в силу эргодичности

$$\lim_{\alpha\to 0}Z(i,j,\alpha)=L\,, \quad (6.37)$$

где L есть те самые средние потери, которые нас и интересуют.

Пусть $0<i<N_1$, $0<j<N_2$. Рассмотрим малый интервал времени $(t,t+\Delta t)$. Тогда можно записать

$$Z(i,j,\alpha)=\alpha\int_t^{t+\Delta t} e^{-\alpha(\tau-t)}M\left\{F(i(\tau),j(\tau))\,|\,i(t)=i,j(t)=j\right\}d\tau+$$

$$+e^{-\alpha\Delta t}\alpha\int_{t+\Delta t}^{\infty} e^{-\alpha(\tau-t-\Delta t)}M\left\{F(i(\tau),j(\tau))\,|\,i(t)=i,j(t)=j\right\}d\tau\,. \quad (6.38)$$

Для того, чтобы выразить второй интеграл через функцию $Z(i,j,\alpha)$, нам необходимо условное математическое ожидание при условии, накладываемом в момент t, выразить через условное математическое ожидание при условии, накладываемом в момент времени $t+\Delta t$.

За время Δt с точностью до $o(\Delta t)$ система из состояния (i,j) с вероятностью

$\lambda_1\Delta t$ перейдет в состояние $i+1,j$;

$\lambda_2\Delta t$ перейдет в состояние $i,j+1$;

$\mu\Delta t\delta(i,j)$ перейдет в состояние $i-1,j$;

$\mu\Delta t\left[1-\delta(i,j)\right]$ перейдет в состояние $i,j-1$;

$1-(\lambda_1+\lambda_2+\mu)\Delta t$ вообще не изменит своего состояния.

Поэтому

$$M\left\{F(i(\tau),j(\tau))\,|\,i(t)=i,j(t)=j\right\}=$$

$$=\lambda_1\Delta tM\left\{F(i(\tau),j(\tau)\,|\,i(t+\Delta t)=i+1,j(t+\Delta t)=j\right\}+$$

$$+\lambda_2\Delta tM\left\{F(i(\tau),j(\tau))\,|\,i(t+\Delta t)=i,j(t+\Delta t)=j+1\right\}+$$

$$+\mu\Delta t\delta(i,j)M\left\{F(i(\tau),j(\tau))\,|\,i(t+\Delta t)=i-1,j(t+\Delta t)=j\right\}+ \qquad (6.39)$$

$$+\mu\Delta t\left[1-\delta(i,j)\right]M\left\{F(i(\tau),j(\tau))\,|\,i(t+\Delta t)=i,j(t+\Delta t)=j-1\right\}+$$

$$+(1-(\lambda_1+\lambda_2+\mu)\Delta t)M\left\{F(i(\tau),j(\tau)\,|\,i(t+\Delta t)=i,j(t+\Delta t)=j\right\}.$$

Учитывая, что $e^{-\alpha\Delta t}=1-\alpha\Delta t+o(\Delta t)$ и

$$\int\limits_t^{t+\Delta t} e^{-\alpha(\tau-t)}M\left\{F(i(\tau),j(\tau))\,|\,i(t)=i,j(t)=j\right\}d\tau=\Delta tF(i,j)+o(\Delta t),$$

не выписывая слагаемые с $o(\Delta t)$, получим

$$Z(i,j,\alpha)=\alpha\Delta tF(i,j)+\left[\lambda_1\Delta tZ(i+1,j,\alpha)+\lambda_2\Delta tZ(i,j+1,\alpha)+\right.$$

$$+\mu\Delta t\delta(i,j)Z(i-1,j,\alpha)+\mu\Delta t\left[1-\delta(i,j)\right]Z(i,j-1,\alpha)+$$

$$\left.+(1-(\lambda_1+\lambda_2+\mu)\Delta t)Z(i,j,\alpha)\right](1-\alpha\Delta t). \qquad (6.40)$$

Уничтожая $Z(i,j,\alpha)$, деля на Δt и делая предельный переход $\Delta t\to 0$, получим

$$(\lambda_1+\lambda_2+\mu+\alpha)Z(i,j,\alpha)=\alpha F(i,j)+\lambda_1 Z(i+1,j,\alpha)+\lambda_2 Z(i,j+1,\alpha)+$$

$$+\mu\delta(i,j)Z(i-1,j,\alpha)+\mu\left[1-\delta(i,j)\right]\ Z(i,j-1,\alpha). \qquad (6.41)$$

Глядя на структуру этого уравнения, можно, не делая предыдущих выкладок, выписать уравнения для $Z(i,j,\alpha)$ при различных значениях i и j.

Решение этой системы будем искать в виде

$$Z(i,j,\alpha)=L+\alpha l(i,j)+o(\alpha) \qquad (6.42)$$

в силу того, что $\lim\limits_{\alpha\to 0}Z(i,j,\alpha)=L$. Тогда, подставляя это решение в (6.41) и оставляя лишь члены со степенью α не выше первой, получим после сокращения на α и предельного перехода $\alpha\to 0$

$$L+(\lambda_1+\lambda_2+\mu)l(i,j)=F(i,j)+\lambda_1 l(i+1,j)+\lambda_2 l(i,j+1)+$$

$$+\mu\delta(i,j)l(i-1,j)+\mu\left[1-\delta(i,j)\right]\ l(i,j-1) \qquad (6.43)$$

или

$$L=F(i,j)+\lambda_1 l(i+1,j)+\lambda_2 l(i,j+1)-(\lambda_1+\lambda_2+\mu)l(i,j)+$$

$$+\mu l(i,j-1)+\mu\delta(i,j)\left[l(i-1,j)-l(i,j-1)\right]. \qquad (6.44)$$

Аналогичные уравнения можно выписать и для граничных значений i и j. Если известно $\delta(i,j)$, то можно, решив эту систему уравнений, найти $l(i,j)$ и L. Но так как мы стремимся сделать $L \Rightarrow \min\limits_{\delta(i,j)}$, то мы должны добиться того, что

$$\delta(i,j)\left[l(i-1,j)-l(i,j-1)\right] \Rightarrow \min_{\delta(i,j)}.$$

Отсюда очевидно, что

$$\delta(i,j)=\begin{cases}1, & \text{если } l(i-1,j)<l(i,j-1),\\ 0, & \text{если } l(i-1,j)>l(i,j-1),\\ \text{любое}, & \text{если } l(i-1,j)=l(i,j-1).\end{cases} \qquad (6.45)$$

На самом деле решение систем (6.44) и (6.45) осуществляется итеративно. Пусть на какой-то итерации известны $\delta^{(k)}(i,j)$. Подставив их вместо $\delta(i,j)$ в систему (6.44), можно найти $l_k(i,j)$ и L_k, а затем найти $\delta^{(k+1)}(i,j)$ по правилу

$$\delta^{(k+1)}(i,j)=\begin{cases}1, & \text{если } l_k(i-1,j)<l_k(i,j-1),\\ 0, & \text{если } l_k(i-1,j)>l_k(i,j-1).\end{cases}$$

Как показал один из авторов (А.А. Назаров), такой итеративный процесс является сходящимся, причем в силу того, что $\delta(i,j)$ равны 0 или 1 и их конечное число, то поэтому процесс сходится за конечное число шагов.

Основная трудность при численном решении системы (6.44) – это ее большая размерность. Им же были разработаны эффективные методы понижения размерности системы (6.44), позволяющие существенно сократить время ее решения.

Заметим еще, что $l_k(i,j)$ определятся с точностью до произвольной константы, и поэтому одну из них можно задавать произвольно.

§ 6.6. Асимптотические методы в СМО с динамическими приоритетами

В общем случае нахождение оптимальных динамических приоритетов возможно, по-видимому, лишь численно. Аналитические результаты можно получить лишь для случая больших загрузок.

Рассмотрим чуть более общую систему $\vec{M}_2|\vec{M}_2|1|\infty$, где интенсивности входящих потоков равны λ_1 и λ_2, а интенсивность обслуживания – μ_1 и μ_2.

Пусть $\rho_1 = \lambda_1/\mu_1$ и $\rho_2 = \lambda_2/\mu_2$ есть загрузки прибора первого и второго типов, $\rho = \rho_1 + \rho_2$ – общая загрузка, i и j – длины очередей заявок первого и второго типов.

При $\rho \to 1$ длины очередей неограниченно возрастают. Вводя стабилизирующий множитель $\varepsilon = 1 - \rho$, перейдем к новым переменным $x = (1-\rho)i$, $y = (1-\rho)j$ и выполним исследование СМО с динамическими приоритетами в условиях большой загрузки $\rho \to 1$. Предварительно уточним предельное поведение слагаемых ρ_1 и ρ_2 в сумме $\rho_1 + \rho_2 = \rho$ при $\rho \to 1$.

Будем считать, что при $\rho \to 1$ для каждого слагаемого выполняется соотношение $\rho_k \to s_k$, где $s_1 + s_2 = 1$. Для этого будем полагать, что значения μ_1 и μ_2 фиксированы, а при $\rho_k \to s_k$ меняются значения λ_k следующим образом:

$$\lambda_k \to s_k \mu_k, \quad k = 1,2. \tag{6.46}$$

В этих условиях докажем следующую теорему.

Теорема 1. Для СМО с динамическими приоритетами при $\varepsilon \to 0$ случайные процессы $x(\tau) = \varepsilon i(\tau/\varepsilon)$, $y(\tau) = \varepsilon j(\tau/\varepsilon)$, где $\tau = \varepsilon t$, по распределению сходятся к детерминированным функциям.

Доказательство

Состояние рассматриваемой СМО определим трехмерным вектором $\{k, i, j\}$, где $k = 1$, если прибор занят заявкой первого типа, и $k = 2$, если он занят заявкой второго типа, и $k = 0$, если прибор свободен.

Так как трехмерный случайный процесс $\{k(t), i(t), j(t)\}$ является цепью Маркова, то распределение вероятностей

$$P_k(i,j,t) = P(k(t)=k,\ \ i(t)=i,\ \ j(t)=j)$$

удовлетворяет системе дифференциальных уравнений Колмогорова

$$\frac{\partial P_1(i,j,t)}{\partial t} + (\lambda_1+\lambda_2+\mu_1)P_1(i,j,t) = \lambda_1 P_1(i-1,j,t) + \lambda_2 P_1(i,j-1,t) +$$
$$+\delta(i+1,j)\left\{\ \mu_1 P_1(i+1,j,t) + \mu_2 P_2(i+1,j,t)\right\}\ ,$$

$$\frac{\partial P_2(i,j,t)}{\partial t} + (\lambda_1+\lambda_2+\mu_2)P_2(i,j,t) = \lambda_1 P_2(i-1,j,t) + \lambda_2 P_2(i,j-1,t) +$$
$$+(1-\delta(i,j+1))\left\{\ \mu_1 P_1(i,j+1,t) + \mu_2 P_2(i,j+1,t)\right\}\ .$$

В этой системе выполним замены

$$t\varepsilon=\tau,\quad i\varepsilon=x,\quad j\varepsilon=y,\quad \frac{1}{\varepsilon}P_k(i,\ \ j,\ \ t)=\pi_k(x,y,\tau,\varepsilon)\,.$$

Тогда получим

$$\varepsilon\frac{\partial \pi_1(x,y,\tau,\varepsilon)}{\partial \tau} + \left(\lambda_1+\lambda_2+\mu_1\right)\pi_1(x,y,\tau,\varepsilon) = \lambda_1\pi_1(x-\varepsilon,y,\tau,\varepsilon) +$$
$$+\lambda_2\pi_1(x,y-\varepsilon,\tau,\varepsilon) + \delta(x+\varepsilon,y)\left\{\mu_1\pi_1(x+\varepsilon,y,\tau,\varepsilon) + \mu_2\pi_2(x+\varepsilon,y,\tau,\varepsilon)\right\},$$

$$\varepsilon\frac{\partial \pi_2(x,y,\tau,\varepsilon)}{\partial \tau} + \left(\lambda_1+\lambda_2+\mu_2\right)\pi_2(x,y,\tau,\varepsilon) = \lambda_1\pi_2(x-\varepsilon,y,\tau,\varepsilon) + \qquad (6.47)$$
$$+\lambda_2\pi_2(x,y-\varepsilon,\tau,\varepsilon) + (1-\delta(x,y+\varepsilon))\left\{\mu_1\pi_1(x,y+\varepsilon,\tau,\varepsilon) + \mu_2\pi_2(x,y+\varepsilon,\tau,\varepsilon)\right\}.$$

Эту систему будем решать в два этапа.

Этап 1. Полагая, что в системе (6.47) функции непрерывны и существуют конечные пределы

$$\lim_{\varepsilon\to 0}\pi_k(x,y,\tau,\varepsilon) = \pi_k(x,y,\tau)\,,$$

выполним в ней предельный переход $\varepsilon \to 0$. Тогда она примет вид

$$\mu_1\pi_1(x,y,\tau) = \delta(x,y)\left\{\mu_1\pi_1(x,y,\tau) + \mu_2\pi_2(x,y,\tau)\right\},$$

$$\mu_2\pi_2(x,y,\tau) = (1-\delta(x,y))\left\{\mu_1\pi_1(x,y,\tau) + \mu_2\pi_2(x,y,\tau)\right\}.$$

Отсюда получаем

$$\mu_2\delta(x,y)\pi_2(x,y,\tau) = \mu_1(1-\delta(x,y))\pi_1(x,y,\tau)\,. \qquad (6.48)$$

Обозначим

$$\pi(x,y,\tau) = \pi_1(x,y,\tau) + \pi_2(x,y,\tau)\,.$$

Тогда из (6.48) получим

$$\pi_k(x,y,\tau) = R_k(x,y)\pi(x,y,\tau)\,, \tag{6.49}$$

где

$$\begin{aligned} R_1(x,y) &= \frac{\mu_2\delta(x,y)}{\mu_2\delta(x,y)+\mu_1(1-\delta(x,y))}\,, \\ R_2(x,y) &= \frac{\mu_1(1-\delta(x,y))}{\mu_2\delta(x,y)+\mu_1(1-\delta(x,y))}\,. \end{aligned} \tag{6.50}$$

Этап 2. В системе (6.47) разложим функции в ряд Тейлора по приращениям аргументов x и y с точностью до $o(\varepsilon)$. Тогда получим следующие равенства:

$$\begin{aligned} &\varepsilon\frac{\partial\pi_1}{\partial\tau} + (\lambda_1+\lambda_2+\mu_1)\pi_1 = \lambda_1\pi_1 - \varepsilon\lambda_1\frac{\partial\pi_1}{\partial x} + \lambda_2\pi_1 - \varepsilon\lambda_2\frac{\partial\pi_1}{\partial y} + \\ &\qquad +\delta(\mu_1\pi_1+\mu_2\pi_2) + \varepsilon\frac{\partial}{\partial x}\left[\delta(\mu_1\pi_1+\mu_2\pi_2)\right] + o(\varepsilon)\,, \\ &\varepsilon\frac{\partial\pi_2}{\partial\tau} + (\lambda_1+\lambda_2+\mu_2)\pi_2 = \lambda_1\pi_2 - \varepsilon\lambda_1\frac{\partial\pi_2}{\partial x} + \lambda_2\pi_2 - \varepsilon\lambda_2\frac{\partial\pi_2}{\partial y} + \\ &\qquad +(1-\delta)(\mu_1\pi_1+\mu_2\pi_2) + \varepsilon\frac{\partial}{\partial y}\left[(1-\delta)(\mu_1\pi_1+\mu_2\pi_2)\right] + o(\varepsilon)\,. \end{aligned} \tag{6.51}$$

Сложив эти равенства и выполнив несложные преобразования, в пределе $\varepsilon \to 0$ с учетом (6.46) из (6.51) получим

$$\begin{aligned} \frac{\partial\pi(x,y,\tau)}{\partial\tau} &= -\frac{\partial}{\partial x}\left\{\left[s_1\mu_1 - \delta(x,y)\left(\mu_1R_1(x,y)+\mu_2R_2(x,y)\right)\right]\pi(x,y,\tau)\right\} - \\ &-\frac{\partial}{\partial y}\left\{\left[s_2\mu_2 - (1-\delta(x,y))\left(\mu_1R_1(x,y)+\mu_2R_2(x,y)\right)\right]\pi(x,y,\tau)\right\}, \end{aligned} \tag{6.52}$$

которое является вырожденным уравнением Фоккера – Планка для плотности вероятностей $\pi(x,y,\tau)$ значений некоторого двумерного диффузионного процесса $\{x(\tau),\ y(\tau)\}$ с нулевыми коэффициентами диффузии. Следовательно, функции $x(\tau)$ и $y(\tau)$ являются детерминированными.

Теорема доказана.

Следствие. Функции $x(\tau)$ и $y(\tau)$ являются решением системы двух обыкновенных дифференциальных уравнений

$$\begin{cases} x'(\tau) = s_1\mu_1 - \delta(x,y)[\mu_1 R_1(x,y) + \mu_2 R_2(x,y)] \ , \\ y'(\tau) = s_2\mu_2 - (1-\delta(x,y))[\mu_1 R_1(x,y) + \mu_2 R_2(x,y)] \ , \end{cases} \tag{6.53}$$

точками покоя которой являются точки (x, y) множества

$$Q = \left\{ (x,y):\ \delta(x,y) = \frac{s_1\mu_1}{s_1\mu_1 + s_2\mu_2} \right\}.$$

Доказательство

Система дифференциальных уравнений (6.53) является следствием уравнения (6.52) и определения коэффициентов переноса диффузионных процессов. Точки покоя системы (6.53) определяются системой алгебраических уравнений

$$\begin{cases} \delta(x,y)[\mu_1 R_1(x,y) + \mu_2 R_2(x,y)] = s_1\mu_1, \\ (1-\delta(x,y))[\mu_1 R_1(x,y) + \mu_2 R_2(x,y)] = s_2\mu_2. \end{cases}$$

Подставляя сюда (6.50), получим

$$\delta \frac{\mu_1\mu_2}{\mu_2\delta + \mu_1(1-\delta)} = s_1\mu_1 ,$$

$$(1-\delta)\frac{\mu_1\mu_2}{\mu_2\delta + \mu_1(1-\delta)} = s_2\mu_2 .$$

Отсюда следует, что

$$\frac{\mu_2\delta}{\mu_2\delta + \mu_1(1-\delta)} = s_1 , \quad \frac{\mu_1(1-\delta)}{\mu_2\delta + \mu_1(1-\delta)} = s_2 ,$$

что и дает равенство

$$\delta(x,\ y) = \frac{s_1\mu_1}{s_1\mu_1 + s_2\mu_2} . \tag{6.54}$$

Следствие доказано.

Смысл доказанных утверждений заключается в том, что в допредельной модели точка с координатами $\{i(t),\ j(t)\}$ почти детерминировано движется к точке множества (6.54), достигнув которое, продолжает движение в этом множестве, но движение там становится уже стохастическим. На этом множестве при $\rho<1$ процесс $\{i(t),\ j(t)\}$ переходит в стационарный режим.

Найдем это стационарное распределение в условиях большой загрузки из системы (6.47), полагая в которой δ постоянным и равным (6.54), а распределение $\pi_k(x,y,\tau,\varepsilon)$ не зависящем от τ, то есть $\pi_k(x,y,\tau,\varepsilon)=\pi_k(x,y,\varepsilon)$.

Запишем систему (6.47) в виде

$$
\begin{gathered}
(\lambda_1+\lambda_2+\mu_1)\pi_1(x,y,\varepsilon)=\lambda_1\pi_1(x-\varepsilon,y,\varepsilon)+\lambda_2\pi_1(x,y-\varepsilon,\varepsilon)+\\
+\frac{s_1\mu_1}{s_1\mu_1+s_2\mu_2}\left[\mu_1\pi_1(x+\varepsilon,y,\varepsilon)+\mu_2\pi_2(x+\varepsilon,y,\varepsilon)\right],\\
(\lambda_1+\lambda_2+\mu_2)\pi_2(x,y,\varepsilon)=\lambda_1\pi_2(x-\varepsilon,y,\varepsilon)+\lambda_2\pi_2(x,y-\varepsilon,\varepsilon)+\\
+\frac{s_2\mu_2}{s_1\mu_1+s_2\mu_2}\left[\mu_1\pi_1(x,y+\varepsilon,\varepsilon)+\mu_2\pi_2(x,y+\varepsilon,\varepsilon)\right].
\end{gathered}
\tag{6.55}
$$

Найдем траектории системы (6.53) в окрестности точек покоя (6.54). Из системы (6.53) имеем

$$
\frac{dy}{dx}=\frac{s_2\mu_2-(1-\delta(x,y))[\mu_1R_1(x,y)+\mu_2R_2(x,y)]}{s_1\mu_1-\delta(x,y)[\mu_1R_1(x,y)+\mu_2R_2(x,y)]}.
$$

Используя равенства (6.50), получим

$$
\frac{dy}{dx}=\frac{s_2\mu_2[\mu_2\delta+\mu_1(1-\delta)]-(1-\delta)\mu_1\mu_2}{s_1\mu_1[\mu_2\delta+\mu_1(1-\delta)]-\delta\mu_1\mu_2}.
$$

Предел этого выражения при $\delta\to s_1\mu_1/(s_1\mu_1+s_2\mu_2)$ находится по правилу Лопиталя. Он равен

$$
\frac{dy}{dx}=\frac{s_2\mu_2[\mu_2-\mu_1]+\mu_1\mu_2}{s_1\mu_1[\mu_2-\mu_1]-\mu_1\mu_2}=-\frac{\mu_2}{\mu_1},
$$

следовательно, имеет место соотношение

$$
\mu_1y+\mu_2x=\text{const}.
$$

Обозначим

$$
\mu_1y+\mu_2x=z. \tag{6.56}
$$

Пусть множество точек покоя системы (6.53), на котором рассматривается решение, образует некоторую замкнутую область. Тогда в допредельной модели краевые условия на решение системы (6.55) определяются главным образом лишь значениями величины z из (6.56). Поэтому в условиях большой загрузки $\rho \to 1$ можно доказать следующую теорему.

Теорема 2. В рассматриваемой области при $\varepsilon \to 0$ стационарное распределение $\pi(x,y) = \pi_1(x,y) + \pi_2(x,y)$, определяемое системой (6.55), зависит лишь от переменной z и имеет вид

$$\pi(x,y) = Ce^{-az},$$

где

$$a = 1/(s_2\mu_1 + s_1\mu_2).$$

Доказательство

В системе (6.55) выполним замену $\pi_k(x,y,\varepsilon) = H_k(z,\varepsilon)$. Тогда получим

$$(\lambda_1+\lambda_2+\mu_1)H_1(z,\varepsilon) = \lambda_1 H_1(z-\varepsilon\mu_2,\varepsilon) + \lambda_2 H_1(z-\varepsilon\mu_1,\varepsilon) + \\ + \frac{s_1\mu_1}{s_1\mu_1+s_2\mu_2}\left[\mu_1 H_1(z+\varepsilon\mu_2,\varepsilon) + \mu_2 H_2(z+\varepsilon\mu_2,\varepsilon)\right],$$

$$(\lambda_1+\lambda_2+\mu_2)H_2(z,\varepsilon) = \lambda_1 H_2(z-\varepsilon\mu_2,\varepsilon) + \lambda_2 H_2(z-\varepsilon\mu_1,\varepsilon) + \\ + \frac{s_2\mu_2}{s_1\mu_1+s_2\mu_2}\left[\mu_1 H_1(z+\varepsilon\mu_2,\varepsilon) + \mu_2 H_2(z+\varepsilon\mu_2,\varepsilon)\right]. \tag{6.57}$$

Эту систему будем решать в три этапа.

Этап 1. Полагая

$$\lim_{\varepsilon\to 0} H_k(z,\varepsilon) = H_k(z),$$

выполним в системе (6.57) предельный переход $\varepsilon \to 0$. Получаем

$$\mu_1 H_1(z) = \frac{s_1\mu_1}{s_1\mu_1+s_2\mu_2}[\mu_1 H_1(z) + \mu_2 H_2(z)],$$

$$\mu_2 H_2(z) = \frac{s_2\mu_2}{s_1\mu_1+s_2\mu_2}[\mu_1 H_1(z) + \mu_2 H_2(z)],$$

откуда следует соотношение

$$s_2 H_1(z) = s_1 H_2(z). \tag{6.58}$$

Обозначив $H(z) = H_1(z) + H_2(z)$, из (6.58) получим

$$H_k(z) = s_k H(z). \tag{6.59}$$

Этап 2. Будем искать решение системы (6.57) в виде

$$H_k(z,\varepsilon)=s_kH(z)+\varepsilon h_k(z)+o(\varepsilon). \tag{6.60}$$

Для этого в системе (6.57) функции $H_k(z\pm\varepsilon\mu,\ \varepsilon)$ разложим в ряд Тейлора по приращениям аргумента z с точностью до $o(\varepsilon)$. Получаем

$$\mu_1\frac{s_2\mu_2}{s_1\mu_1+s_2\mu_2}H_1-\mu_2\frac{s_1\mu_1}{s_1\mu_1+s_2\mu_2}H_2=-\varepsilon(\mu_2\lambda_1+\mu_1\lambda_2)\frac{\partial H_1}{\partial z}+$$

$$+\varepsilon\mu_2\frac{s_1\mu_1}{s_1\mu_1+s_2\mu_2}\frac{\partial}{\partial z}\{\mu_1H_1+\mu_2H_2\}+o(\varepsilon),$$

$$\mu_2\frac{s_1\mu_1}{s_1\mu_1+s_2\mu_2}H_2-\mu_1\frac{s_2\mu_2}{s_1\mu_1+s_2\mu_2}H_1=-\varepsilon(\mu_2\lambda_1+\mu_1\lambda_2)\frac{\partial H_2}{\partial z}+$$

$$+\varepsilon\mu_1\frac{s_1\mu_2}{s_1\mu_1+s_2\mu_2}\frac{\partial}{\partial z}\{\mu_1H_1+\mu_2H_2\}+o(\varepsilon).$$

Выполнив несложные преобразования и подставляя сюда разложения (6.60), будем иметь

$$\begin{aligned}&s_2\{s_1H(z)+\varepsilon h_1(z)\}-s_1\{s_2H(z)+\varepsilon h_2(z)\}=\\&=-\varepsilon\{(\rho_1+\rho_2-1)s_1(\mu_1s_1+\mu_2s_2)\}H'(z)+o(\varepsilon),\\&s_1\{s_2H(z)+\varepsilon h_2(z)\}-s_2\{s_1H(z)+\varepsilon h_1(z)\}=\\&=-\varepsilon\{(\rho_1+\rho_2-1)s_2(\mu_1s_1+\mu_2s_2)\}H'(z)+o(\varepsilon),\end{aligned} \tag{6.61}$$

откуда при $\varepsilon\to 0$ получим систему

$$s_2h_1(z)-s_1h_2(z)=0,$$

$$s_1h_2(z)-s_2h_1(z)=0.$$

Обозначив $h(z)=h_1(z)+h_2(z)$, аналогично (6.59) можно записать $h_k(z)=s_kh(z)$. Следовательно, разложение (6.60) имеет вид

$$H_k(z,\varepsilon)=s_kH(z)+\varepsilon s_kh(z)+o(\varepsilon). \tag{6.62}$$

Этап 3. На этом этапе функции $H_k(z\pm\varepsilon\mu,\ \varepsilon)$ в системе (6.57) разложим в ряд Тейлора по приращениям аргумента z с точностью до $o(\varepsilon^2)$. Имеем

$$\mu_1H_1=-\varepsilon(\lambda_1\mu_2+\lambda_2\mu_1)\frac{\partial H_1}{\partial z}+\frac{\varepsilon^2}{2}(\lambda_1\mu_2^2+\lambda_2\mu_1^2)\frac{\partial^2H_1}{\partial z^2}+\frac{s_1\mu_1}{s_1\mu_1+s_2\mu_2}\times$$

$$\times\left\{\mu_1H_1+\mu_2H_2+\varepsilon\mu_2\frac{\partial}{\partial z}[\mu_1H_1+\mu_2H_2]+\frac{\varepsilon^2}{2}\mu_2^2\frac{\partial^2}{\partial z^2}[\mu_1H_1+\mu_2H_2]\right\}+o(\varepsilon^2),$$

$$\mu_2H_2=-\varepsilon(\lambda_1\mu_2+\lambda_2\mu_1)\frac{\partial H_2}{\partial z}+\frac{\varepsilon^2}{2}(\lambda_1\mu_2^2+\lambda_2\mu_1^2)\frac{\partial^2 H_2}{\partial z^2}+\frac{s_2\mu_2}{s_1\mu_1+s_2\mu_2}\times$$

$$\times\left\{\mu_1H_1+\mu_2H_2+\varepsilon\mu_1\frac{\partial}{\partial z}[\mu_1H_1+\mu_2H_2]+\frac{\varepsilon^2}{2}\mu_1^2\frac{\partial^2}{\partial z^2}[\mu_1H_1+\mu_2H_2]\right\}+o(\varepsilon^2).$$

Сложив уравнения этой системы, получим следующее равенство:

$$-\varepsilon\frac{\partial}{\partial z}\left\{(\lambda_1\mu_2+\lambda_2\mu_1)H-\frac{\mu_1\mu_2}{s_1\mu_1+s_2\mu_2}\left(\mu_1H_1+\mu_2H_2\right)\right\}+$$

$$+\frac{\varepsilon^2}{2}\frac{\partial^2}{\partial z^2}\left\{(\lambda_1\mu_2^2+\lambda_2\mu_1^2)H+\mu_1\mu_2\frac{s_1\mu_2+s_2\mu_1}{s_1\mu_1+s_2\mu_2}\left(\mu_1H_1+\mu_2H_2\right)\right\}=o(\varepsilon^2).$$

Подставив сюда разложение (6.62), получим

$$-\varepsilon(\rho_1+\rho_2-1)H'(z)-\varepsilon^2(\rho_1+\rho_2-1)h'(z)+$$

$$+\frac{\varepsilon^2}{2}(\rho_1\mu_2+\rho_2\mu_1+s_1\mu_2+s_2\mu_1)H''(z)=o(\varepsilon^2).$$

Здесь $\rho_1+\rho_2=\rho=1-\varepsilon$. Поэтому, деля на ε^2 и переходя к пределу $\varepsilon\to 0$, получим, что $H(z)$ удовлетворяет уравнению

$$H'(z)+(s_1\mu_2+s_2\mu_1)H''(z)=0,$$

то есть обыкновенному дифференциальному уравнению с постоянными коэффициентами.

Решение этого уравнения имеет вид

$$H(z)=Ce^{-az}, \tag{6.63}$$

где

$$a=1/(s_1\mu_2+s_2\mu_1),$$

а константа C определяется из условия нормировки.

Теорема доказана.

Как и ранее, определим область Q следующим образом:

$$Q=\left\{(x,\ y):\ \delta(x,y)=\frac{s_1\mu_1}{s_1\mu_1+s_2\mu_2}\right\}.$$

Тогда в четверти плоскости $\{x \geq 0,\ \ y \geq 0\}$ стационарная плотность вероятностей $\pi(x, y)$ имеет вид

$$\pi(x,\ y) = \begin{cases} 0, & \text{если } (x,y) \notin Q, \\ Ce^{-a(\mu_1 y + \mu_2 x)}, & \text{если } (x,y) \in Q. \end{cases}$$

Константа C найдется из условия нормировки

$$C\iint\limits_Q e^{-a(\mu_2 x + \mu_1 y)} dxdy = 1. \tag{6.64}$$

Для любой функции $G(x,\ y)$ ее математическое ожидание $L = M\{G(x, y)\}$ имеет вид

$$L = \iint\limits_Q G(x, y) e^{-a(\mu_2 x + \mu_1 y)} dxdy \Big/ \iint\limits_Q e^{-a(\mu_2 x + \mu_1 y)} dxdy\,.$$

Как показано в предыдущем разделе, оптимальные динамические приоритеты являются нерандомизированными. Возникающая ситуация изображена на рис. 6.5.

Вся плоскость $(x,\ y)$ разбивается кривой Q на две области. В области I выбираются на обслуживание заявки первого типа, то есть $\delta(x,y) = 1$. В области II выбираются на обслуживание заявки второго типа, то есть $\delta(x,y) = 0$. И лишь на кривой Q остается рандомизированный выбор на обслуживание, эта кривая и есть та самая область Q, о которой говорилось выше.

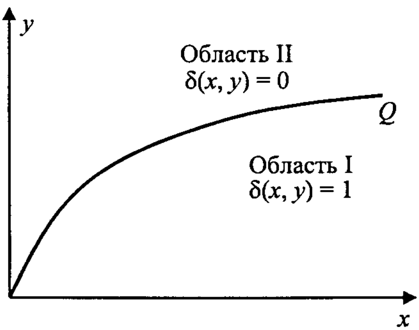


Рис. 6.5

Зададим кривую Q в параметрическом виде. В качестве параметра выберем величину $z = \mu_2 x + \mu_1 y$, и будем задавать кривую Q в форме

$$x = f(z)/\mu_2\,, \quad y = (z - f(z))/\mu_1\,.$$

Тогда соотношение $z = \mu_2 x + \mu_1 y$ будет выполнено автоматически.

Из вида $\pi(x, y)$ следует, что $\pi(z) = ae^{-az}$. Так как x и y определяются величиной z однозначно, то для совместной плотности вероятностей величин x, y и z можем записать

$$\pi(x,y,z) = ae^{-az}\delta\left(x - \frac{f(z)}{\mu_2}\right)\delta\left(y - \frac{z-f(z)}{\mu_1}\right), \qquad (6.65)$$

где $\delta(.)$ – δ-функция Дирака.

Задачей является нахождение такой функции $f(z)$, для которой $L = M\{F(i,j)\}$ минимально. Дадим решение этой задачи в условиях большой загрузки $\rho \to 1$ и найдем вид функции $f(z)$, минимизирующий значение функционала

$$L = M\left\{F\left(\frac{x(\tau)}{1-\rho}, \frac{y(\tau)}{1-\rho}\right)\right\}. \qquad (6.66)$$

Тогда для (6.66) получим

$$L = a\int_0^\infty e^{-az}F\left(\frac{f(z)}{\mu_2(1-\rho)}, \frac{z-f(z)}{\mu_1(1-\rho)}\right)dz. \qquad (6.67)$$

Очевидно, что при фиксированном z

$$F\left(\frac{f(z)}{\mu_2(1-\rho)}, \frac{z-f(z)}{\mu_1(1-\rho)}\right) \ge \min_{0\le u\le z} F\left(\frac{u}{\mu_2(1-\rho)}, \frac{z-u}{\mu_1(1-\rho)}\right),$$

и поэтому для любой функции $f(z)$ имеет место неравенство

$$L = a\int_0^\infty e^{-az}F\left(\frac{f(z)}{\mu_2(1-\rho)}, \frac{z-f(z)}{\mu_1(1-\rho)}\right)dz \ge a\int_0^\infty e^{-az} \min_{0\le u\le z} F\left(\frac{u}{\mu_2(1-\rho)}, \frac{z-u}{\mu_1(1-\rho)}\right)dz, \qquad (6.68)$$

которое обращается в равенство, если взять $f(z) = f^*(z)$, где $f^*(z)$ определяется условием

$$F\left(\frac{f^*(z)}{\mu_2(1-\rho)}, \frac{z-f^*(z)}{\mu_1(1-\rho)}\right) = \min_{0\le u\le z} F\left(\frac{u}{\mu_2(1-\rho)}, \frac{z-u}{\mu_1(1-\rho)}\right). \qquad (6.69)$$

При этом минимальное значение L составит величину

$$L^* = a\int_0^\infty e^{-az} \min_{0\le u\le z} F\left(\frac{u}{\mu_2(1-\rho)}, \frac{z-u}{\mu_1(1-\rho)}\right)dz. \qquad (6.70)$$

Таким образом, (6.69) и (6.70) решают задачу нахождения оптимальных динамических приоритетов (функция $f^*(z)$) и соответствующего минимального значения L^* величины потерь L.

Рассмотрим примеры.

Пример 1

Пусть $F(i,j)=(1-\rho)C_1 i+(1-\rho)C_2 j$. Тогда

$$\min_{0\le u\le z}\left\{\frac{C_1}{\mu_2}u+\frac{C_2}{\mu_1}(z-u)\right\}=\begin{cases}\dfrac{C_2}{\mu_1}, & \text{если } C_1\mu_1>C_2\mu_2,\\[2mm] \dfrac{C_1}{\mu_2}, & \text{если } C_1\mu_1<C_2\mu_2.\end{cases}$$

Следовательно, при линейной функции потерь оптимальными в классе динамических являются статические приоритеты. При $C_1\mu_1>C_2\mu_2$ высший приоритет имеют заявки первого типа, при $C_1\mu_1<C_2\mu_2$ – второго.

Заметим, что так как $\mu_i=1/b_i$, где b_i – среднее время обслуживания заявки i-го типа, то полученное правило назначения приоритетов совпадает с правилом, сформулированным в § 6.3.

Пример 2

Пусть $F(i,j)=(1-\rho)^2[C_1 i^2+C_2 j^2]$. Тогда

$$\min_{0\le u\le z}\left\{\frac{C_1}{\mu_2^2}u^2+\frac{C_2}{\mu_1^2}(z-u)^2\right\}=\frac{C_1C_2}{C_1\mu_1^2+C_2\mu_2^2}z^2.$$

При этом границей раздела является прямая, уравнение которой в параметрической форме имеет вид

$$x=\frac{C_2\mu_2}{C_1\mu_1^2+C_2\mu_2^2}z,\quad y=\frac{C_1\mu_1}{C_1\mu_1^2+C_2\mu_2^2}z,$$

или в явной форме

$$y=\frac{C_1\mu_1}{C_2\mu_2}x.$$

Минимальное значение функционала потерь составляет

$$L^*=\int_0^\infty ae^{-az}\frac{C_1C_2}{C_1\mu_1^2+C_2\mu_2^2}z^2dz=\frac{C_1C_2}{C_1\mu_1^2+C_2\mu_2^2}\frac{2}{a^2}.$$

Пусть заявкам первого типа назначен статический высший приоритет. Это соответствует тому, что $f(z)=0$ и значение функционала потерь в этом случае равно

$$L_1 = \int\limits_0^\infty ae^{-az} C_2 \frac{z^2}{\mu_1^2} dz = \frac{C_2}{\mu_1^2} \frac{2}{a^2}.$$

При этом выигрыш от применения динамических приоритетов равен

$$\frac{L_1}{L^*} = \frac{C_1\mu_1^2 + C_2\mu_2^2}{C_1\mu_1^2} > 1,$$

то есть динамические приоритеты в этом случае всегда лучше статических.

Пример 3

Пусть $F(i,\ j) = (1-\rho)^2 C_1 i^2 + (1-\rho) C_2 j$. Тогда задача принимает вид

$$\frac{C_1}{\mu_2^2} u^2 + \frac{C_2}{\mu_1}(z-u) \Rightarrow \min_{0 \le u \le z}. \tag{6.71}$$

Рассмотрим функцию

$$G(u) = \frac{C_1}{\mu_2^2} u^2 + \frac{C_2}{\mu_1}(z-u).$$

Очевидно, что это – парабола, минимум которой находится в точке $u^* = C_2\mu_2^2 / 2C_1\mu_1$. Поэтому, если $z < C_2\mu_2^2/2C_1\mu_1$, то минимум в (6.71) достигается при $u = z$, если же $z \ge C_2\mu_2^2/2C_1\mu_1$, то минимум в (6.71) достигается при $u = C_2\mu_2^2/2C_1\mu_1$. Таким образом,

$$f^*(z) = \begin{cases} \dfrac{\mu_2^2 C_2}{2\mu_1 C_1}, & \text{если } z \ge \dfrac{\mu_2^2 C_2}{2\mu_1 C_1}, \\ z, & \text{если } z < \dfrac{\mu_2^2 C_2}{2\mu_1 C_1}. \end{cases} \tag{6.72}$$

Отсюда определяется и граница раздела Q. При $z \ge C_2\mu_2^2/2C_1\mu_1$ получаем

$$x = \frac{f^*(z)}{\mu_2} = \frac{\mu_2 C_2}{2\mu_1 C_1} = x^*, \quad 0 \le y < +\infty, \tag{6.73}$$

а при $z < C_2\mu_2^2/2C_1\mu_1$

$$0 \le x \le x^*, \qquad y = 0. \tag{6.74}$$

Таким образом, правило выбора заявок из очереди имеет вид:

- если $0 \le x \le x^*$, то на обслуживание берется заявка из второй очереди;

- если $x > x^*$, то на обслуживание берется заявка из первой очереди.

Вид соответствующих областей приведен на рис. 6.6.

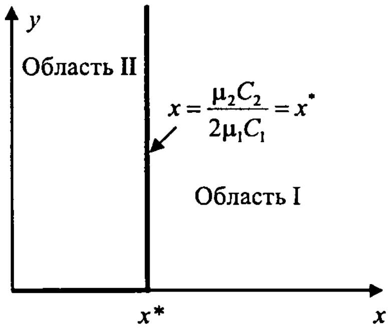


Рис. 6.6

При переходе с случаю произвольной загрузки это означает следующее. Имеется два потока заявок – заявки первого и второго типа, которые становятся в отдельные очереди. Пусть в очередях i заявок первого типа и j заявок второго типа соответственно. Устанавливается некоторое пороговое значение i^* и, при освобождении обслуживающего прибора, если выполнено условие $i \ge i^*$, берется на обслуживание заявка первого типа, а при выполнении условия $i < i^*$ – второго типа (естественно, если они имеются). Математически это означает, что

$$\delta(i,j) = \begin{cases} 0, & \text{если } i < i^* \wedge j > 0, \\ 1, & \text{в противном случае;} \end{cases}$$

$$\delta(0,\ 0) \text{ не определена.}$$

Эта дисциплина обслуживания обобщает дисциплину статических приоритетов и переходит в нее при $i^* = 1$.

Глава 7. СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ЦИКЛИЧЕСКИМ ОБСЛУЖИВАНИЕМ

В качестве математических моделей локальных вычислительных сетей с маркерным доступом довольно часто используют системы массового обслуживания с циклическим обслуживанием очередей. Рассмотрим некоторые методы их исследования.

§ 7.1. СМО с циклическим обслуживанием очередей (cyclic queue)

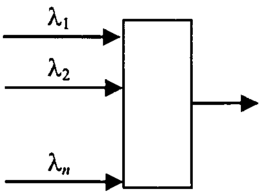


Рис. 7.1

Имеется один обслуживающий прибор, на который поступают n потоков заявок с интенсивностями λ_1, λ_2,..., λ_n. Время обслуживания заявок из i-й очереди имеет функцию распределения $B_i(x)$. Очереди обходятся в циклическом порядке, время на перестройку с i-й очереди на $i+1$-ю очередь имеет функцию распределения $A_i(x)$ (рис. 7.1).

СМО с прогулками (vacances) обслуживающего прибора (vacation queues)

Имеется СМО $M|G|1$, в которой прибор время от времени уходит на прогулку, время которой случайное с функцией распределения $A(x)$ (рис. 7.2).

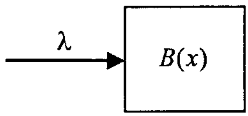


Рис. 7.2

Исследование СМО с циклическим обслуживанием n очередей выполняется следующим образом: систему декомпозируют на n систем с прогулками обслуживающего прибора и выбором функции распределения времени прогулки в зависимости от дисциплины переключения. Затем каким-то образом все эти системы «завязываются» в одну систему, хотя бы приближенно эквивалентную исходной.

Здесь надо еще уточнить правило, по которому прибор переключается с i-й очереди на $i+1$-ю (в cyclic queue) или уходит на прогулку (в vacation queues). Дисциплины здесь могут быть самые разнообразные. Наиболее часто встречаются следующие:

1. k-limited (k-ограниченная). Переключившись на i-ю очередь (или вернувшись с прогулки), прибор обслуживает там k заявок (если очередь будет исчерпана ранее, – то меньше) и затем переключается на следующую очередь (или уходит на прогулку). Дисциплина 1-limited называется **ординарной.**

2. Исчерпывающая (exhaustive). Подключившись к i-й очереди, прибор обслуживает ее до тех пор, пока в ней не останется ни одной заявки (то есть до полного ее исчерпания), а затем переключается на другую очередь.

3. С сокращением (semi-exhaustive или decrementing). Прибор уходит, оставляя в очереди на одну заявку меньше, чем в момент своего прихода (то есть, если в очереди в момент его прихода было i заявок, он уйдет, когда в момент его освобождения в очереди будет $i-1$ заявка).

4. Вентильная (gated). Подключаясь к очереди, прибор замечает, какие заявки в ней находятся, и уходит, когда все заявки, бывшие в очереди в момент его прихода, будут обслужены.

Заметим, что теория СМО с циклическим обслуживанием гораздо сложнее теории СМО с прогулками обслуживающего прибора.

Рассмотрим для примера СМО с прогулками и дисциплиной 1-limited.

Итак, имеется СМО $M|G|1|\infty$ с функцией распределения времени обслуживания $B(x)$. Прибор уходит на прогулку, функция распределения времени которой есть $A(x)$.

Пусть t_n – моменты начала или конца прогулки. Обозначим: Q_i – вероятность того, что в момент времени t_n+0 идет прогулка и в системе i заявок; P_i – вероятность того, что в момент времени t_n+0 в системе идет обслуживание и в системе i заявок; α_i – вероятность того, что за время прогулки в систему пришло i заявок; β_i – вероятность того, что за время обслуживания в систему пришло i заявок.

Тогда имеем следующую систему уравнений:

$$Q_0 = P_1\beta_0 + \alpha_0 Q_0, \quad P_0 = 0,$$

$$Q_i = \sum_{k=0}^{i} \beta_k P_{i-k+1}, \quad P_i = \sum_{k=0}^{i} \alpha_k Q_{i-k}. \tag{7.1}$$

Введем производящие функции

$$F(z)=\sum_{k=0}^{\infty}P_kz^k\,,\quad G(z)=\sum_{k=0}^{\infty}Q_kz^k\,,$$

$$\alpha(z)=\sum_{k=0}^{\infty}\alpha_kz^k\,,\quad \beta(z)=\sum_{k=0}^{\infty}\beta_kz^k\,.$$

Тогда стандартным путем получаем

$$G(z)=\beta_0P_1+\alpha_0Q_0+\sum_{i=1}^{\infty}z^i\sum_{k=0}^{i}\beta_kP_{i-k+1}=\alpha_0Q_0+\sum_{i=0}^{\infty}z^i\sum_{k=0}^{i}\beta_kP_{i-k+1}=$$

$$=\alpha_0Q_0+\sum_{k=0}^{\infty}\beta_k\sum_{i=k}^{\infty}z^iP_{i-k+1}=\alpha_0Q_0+\sum_{k=0}^{\infty}\beta_kz^k\sum_{j=0}^{\infty}z^jP_{j+1}=$$

$$=\alpha_0Q_0+\beta(z)\frac{1}{z}\left[F(z)-P_0\right]=\alpha_0Q_0+\frac{\beta(z)}{z}F(z)\,.$$

Итак, имеем систему

$$\begin{cases}G(z)-\dfrac{\beta(z)}{z}F(z)=\alpha_0Q_0,\\ \alpha(z)G(z)-F(z)=\alpha_0Q_0.\end{cases}\tag{7.2}$$

Ее решение

$$F(z)=\alpha_0Q_0\frac{(1-\alpha(z))z}{\alpha(z)\beta(z)-z}\,,$$

$$G(z)=\alpha_0Q_0\frac{\beta(z)-z}{\alpha(z)\beta(z)-z}\,,\tag{7.3}$$

откуда

$$F(z)+G(z)=\alpha_0Q_0\frac{\beta(z)-z\alpha(z)}{\alpha(z)\beta(z)-z}\,.\tag{7.4}$$

Заметим, что $F(z)+G(z)$ имеет смысл производящей функции вероятностей того, что в системе находится i заявок в выбранные нами моменты t_n. Поэтому, в силу условия нормировки, должно быть $F(1)+G(1)=1$. Отсюда находится константа Q_0. По правилу Лопиталя имеем

$$F(1)+G(1)=\alpha_0 Q_0 \lim_{z\to 1}\frac{\beta'(z)-\alpha(z)-z\alpha'(z)}{\alpha'(z)\beta(z)+\alpha(z)\beta'(z)-1}, \tag{7.5}$$

но, как это было в системе $M \mid G \mid 1 \mid \infty$,

$$\alpha(z)=A^*(\lambda-\lambda z),$$

$$\alpha'(z)=-\lambda A^{*\prime}(\lambda-\lambda z),$$

$$\alpha'(1)=a\lambda,$$

$$\beta(z)=B^*(\lambda-\lambda z),$$

$$\beta'(z)=-\lambda B^{*\prime}(\lambda-\lambda z),$$

$$\beta'(1)=b\lambda,$$

и поэтому

$$F(1)+G(1)=\alpha_0 Q_0\frac{b\lambda-1-a\lambda}{a\lambda+b\lambda-1}=\alpha_0 Q_0\frac{1-(b-a)\lambda}{1-(a+b)\lambda}=1. \tag{7.6}$$

Отсюда получается условие существования стационарного режима в рассматриваемой СМО

$$1-(a+b)\lambda>0, \text{ т.е. } \lambda(a+b)<1,$$

что вполне логично и могло бы быть написано из эргодических соображений. Окончательно

$$F(z)=\frac{1-(a+b)\lambda}{1-(b-a)\lambda}\cdot\frac{z\left(1-\alpha(z)\right)}{\alpha(z)\beta(z)-z},$$

$$G(z)=\frac{1-(a+b)\lambda}{1-(b-a)\lambda}\cdot\frac{\beta(z)-z}{\alpha(z)\beta(z)-z}.$$

А теперь надо сделать самое главное – произвести возврат к циклической СМО. Для этого надо найти $A_i(x)$ в циклической СМО для какой-то i-й очереди.

Обозначим: τ_i – время перестройки после i-й очереди; ν_i – время, которое прибор проводит, посещая i-ю очередь; ξ_i – время прогулки для i-й очереди.

Тогда

$$\xi_i=\sum_{j=1}^{N}\tau_j+\sum_{\substack{j=1\\ j\neq i}}^{N}\nu_j. \tag{7.7}$$

Но

$$\nu_i = \begin{cases} \eta_i, & \text{если в момент подключения прибора в очереди были заявки,} \\ 0, & \text{если этих заявок не было.} \end{cases}$$

Заметим (и это главное), что ν_i, вообще говоря, – зависимые случайные величины.

Из наших рассуждений следует, что вероятность застать очередь по возвращению из прогулки пустой равна $\alpha_0 Q_0$, так как это произойдет тогда, когда прибор оставил ее пустой, и за время его прогулки в нее не поступило ни одной заявки. Так как нам нужна **условная** вероятность ухода из пустой очереди (ведь t_n есть не только моменты ухода, но и моменты начала работы прибора), то

$$\pi_0 = \alpha_0 Q_0 \Big/ \sum_{i=0}^{\infty} Q_i = \alpha_0 Q_0 / G(1).$$

Беря выражение для $G(z)$ и переходя к пределу $z \to 1$, получим

$$G(1) = \frac{1-(a_i+b_i)\lambda_i}{1-(b_i-a_i)\lambda_i} \cdot \frac{\beta_i'(z)-z}{\alpha_i'(z)\beta_i(z)+\alpha_i(z)\beta_i'(z)-z} =$$

$$= \frac{1-(a_i+b_i)\lambda_i}{1-(b_i-a_i)\lambda_i} \cdot \frac{\lambda_i b_i - 1}{\lambda_i a_i + \lambda_i b_i - 1} = \frac{1-\lambda_i b_i}{1-(b_i-a_i)\lambda_i}, \tag{7.8}$$

и поэтому

$$\pi_0 = \frac{1-(a_i+b_i)\lambda_i}{1-(b_i-a_i)\lambda_i} \cdot \frac{\lambda_i a_i + \lambda_i b_i - 1}{1-\lambda_i b_i} = \frac{1-(a_i+b_i)\lambda_i}{1-\lambda_i b_i},$$

$$1-\pi_0 = \frac{a_i \lambda_i}{1-\lambda_i b_i}. \tag{7.9}$$

Так как $M\{\xi_i\} = a_i$, то, усредняя (7.7), получим

$$a_i = \mathrm{T} + \sum_{\substack{j=1 \\ j \neq i}}^{N} b_j \frac{a_j \lambda_j}{1-\lambda_j b_j},$$

где $\mathrm{T} = \sum_{i=1}^{N} \overline{\tau}_i$. Обозначим $\rho_j = \lambda_j b_j$. Тогда

$$a_i = \mathrm{T} + \sum_{j=1}^{N} a_j \frac{\rho_j}{1-\rho_j} - a_i \frac{\rho_i}{1-\rho_i} = \mathrm{T} + A - a_i \frac{\rho_i}{1-\rho_i},$$

где через A обозначено $A=\sum\limits_{j=1}^{N} a_j \dfrac{\rho_j}{1-\rho_j}$. Отсюда

$$a_i\left(1+\frac{\rho_i}{1-\rho_i}\right)=\frac{a_i}{1-\rho_i}=\mathrm{T}+A,$$

то есть $a_i=(A+\mathrm{T})(1-\rho_i)$.

Подставляя это обратно в выражение для A, получим

$$A=\sum_{j=1}^{N}(A+\mathrm{T})\rho_j=(A+\mathrm{T})\rho, \quad \rho=\sum_j \rho_j.$$

Отсюда окончательно

$$A=\frac{\mathrm{T}\rho}{1-\rho}, \quad a_i=\mathrm{T}\frac{1-\rho_i}{1-\rho}. \tag{7.10}$$

Любопытно, что:

а) вероятность того, что i-я очередь не пуста в момент подключения прибора, равна

$$\frac{a_i\lambda_i}{1-\lambda_i b_i}=\frac{\lambda_i\mathrm{T}}{1-\rho}; \tag{7.11}$$

б) среднее время, которое прибор проводит за 1 визит у i-й очереди, равно

$$\nu_i=\frac{\lambda_i\mathrm{T}}{1-\rho}b_i=\frac{\rho_i\mathrm{T}}{1-\rho}; \tag{7.12}$$

в) среднее время цикла равно

$$C=\frac{\rho_i\mathrm{T}}{1-\rho}+\frac{(1-\rho_i)\mathrm{T}}{1-\rho}=\frac{\mathrm{T}}{1-\rho}. \tag{7.13}$$

Теперь можно найти и функцию распределения времени прогулки $A_i(x)$. Точно решить задачу трудно, так как ν_i зависимы. Однако при $n \gg 1$ можно приближенно считать, что ν_i независимы, и тогда для преобразований Лапласа от соответствующих функций распределений получим

$$A_i^*(s)=\mathrm{T}^*(s)\prod_{\substack{j=1\\ j\neq i}}^{n}\left(\left(1-\frac{\lambda_j\mathrm{T}}{1-\rho}\right)\cdot 1+\frac{\lambda_j\mathrm{T}}{1-\rho}B_j^*(s)\right), \tag{7.14}$$

где $T^*(s)$ – преобразование Лапласа от суммарного времени перестройки $\sum_{j=1}^{N}\tau_j$.

Знание функции распределения для времени прогулки i-го прибора позволяет приближенно найти все характеристики i-й очереди.

§ 7.2. Эргодические соображения для циклических СМО

Пусть $\overline{\tau}_i$ – среднее время перестройки после i-й очереди и $T=\sum_{i=1}^{n}\overline{\tau}_i$; $b_i=M\{\eta_i\}$ и $b_i^{(2)}=M\{\eta_i^2\}$, где η_i – время обслуживания в i-й очереди; λ_i – интенсивность потока заявок i-й очереди; $\rho_i=\lambda_i b_i$ – загрузка i-й очереди и $\rho=\sum_{i=1}^{n}\rho_i$.

Определим время цикла как время между двумя соседними (последовательными) подключениями прибора к i-й станции. Среднее время цикла обозначим через C (оно, очевидно, от i не зависит).

Тогда

- за время C в систему придет в среднем $\lambda_i C$ заявок i-го типа, и они будут обслуживаться в среднем время $\lambda_i C b_i=\rho_i C$;

- суммарное время обслуживания всех заявок, поступивших за время цикла, займет в среднем $C\sum_{i=1}^{n}\rho_i=C\cdot\rho$ времени.

С другой стороны, время цикла складывается из времени обслуживания и времени переключения.

Поэтому

$$C=\rho\cdot C+T,$$

откуда

$$C=T/(1-\rho).$$

Пусть $\overline{v}_i$ – среднее время визита в i-ю очередь. Чтобы в системе существовал стационарный режим, надо, чтобы за среднее время визита обслуживались те заявки, которые пришли за время цикла. Так как таких заявок $\lambda_i C$, то должно быть $\lambda_i C b_i=\overline{v}_i$, откуда $\overline{v}_i=\dfrac{\rho_i T}{1-\rho}$.

Рассмотрим еще вопрос о существовании стационарного режима. Ясно, что должно выполняться условие $\rho < 1$, и этого достаточно для исчерпывающей «вентильной» дисциплины обслуживания. Однако для других дисциплин нужны дополнительные ограничения. Так, для ординарной дисциплины должно быть еще

$$\lambda_i C = \lambda_i T/(1-\rho) < 1,$$

так как за время цикла должно приходить в среднем меньше одной заявки. Для semi-exhaustive-дисциплины должно быть еще

$$\lambda_i(C - \bar{v}_i) = \frac{\lambda_i(1-\rho_i)T}{1-\rho} < 1,$$

так как для нее среднее число поступлений заявок за межвизитное время должно быть меньше 1 для каждой очереди.

§ 7.3. Законы сохранения для циклических СМО

Пусть имеется система $\vec{M}_n \mid \vec{G}_n \mid 1 \mid \infty$, i-й входящий поток является пуассоновским потоком интенсивности λ_i и функцией распределения времени обслуживания $B_i(x)$. Прибор обслуживает очереди в циклическом порядке.

Пусть время на перестройку равно 0. Обозначим незавершенную работу в момент времени t через $u(t)$. Тогда

$$u(t) = \xi_0 + \sum_{k=1}^{n} \sum_{i=1}^{N_k(t)} \xi_{i\,k},$$

где ξ_0 – остаточное время обслуживания требования, стоящего на приборе в момент времени t; $N_k(t)$ – число заявок k-го потока в системе в момент времени t; $\xi_{i\,k}$ – время обслуживания i-й заявки k-го потока.

Берем математическое ожидание

$$M\{u(t)\} = M\xi_0 + \sum_{k=1}^{n}\sum_{s=1}^{\infty} P\{N_k(t) = s\} \sum_{i=1}^{s} M\xi_{i\,k},$$

и устремим $t \to \infty$. Тогда $M\{u(t)\} = \bar{u}$, $P\{N_k(t) = s\} \to \pi_{s\,k}$ (финальные вероятности), $M\xi_{i\,k} = b_k$. Поэтому

$$\bar{u} = M\xi_0 + \sum_{k=1}^{n}\sum_{s=1}^{\infty} \pi_{s\,k} \cdot s \cdot b_k = M\xi_0 + \sum_{k=1}^{n} b_k \sum_{s=1}^{\infty} s \quad \pi_{s\,k} = M\xi_0 + \sum_{k=1}^{n} b_k \bar{n}_k ,$$

где $\bar{n}_k$ – средняя длина k-й очереди. По формулам Литтла $n_k = \lambda_k w_k$, и поэтому

$$\bar{u} = M\xi_0 + \sum_{k=1}^{n} b_k \lambda_k w_k = M\xi_0 + \sum_{k=1}^{n} \rho_k w_k .$$

Найдем теперь $M\xi_0$. С вероятностью ρ_k на приборе обслуживается заявка k-го типа. Среднее время ее дообслуживания равно $b_k^{(2)}/2b_k$. Поэтому

$$M\xi_0 = \sum_{k=1}^{n} \rho_k \frac{b_k^{(2)}}{2b_k} = \frac{1}{2}\sum_{k=1}^{n} \lambda_k b_k^{(2)} = w_0 .$$

Таким образом,

$$\bar{u} = w_0 + \sum_{k=1}^{n} \rho_k w_k .$$

Назовем «corresponding» систему $M|G|1|\infty$ с входным потоком интенсивности $\Lambda = \sum_i \lambda_i$ и функцией распределения времени обслуживания $\sum_{i=1}^{n} \frac{\lambda_i}{\Lambda} B_i(x)$, $\rho = \sum_{i=1}^{n} \rho_i$. Если $T_i = 0$, то в смысле незавершенной работы эти СМО эквивалентны. Но в «corresponding» СМО

$$\bar{u}_c = \frac{\Lambda b^{(2)}}{2(1-\rho)} = \frac{\Lambda}{2(1-\rho)} \sum_{i=1}^{n} \frac{\lambda_i}{\Lambda} b_i^{(2)} = \frac{w_0}{1-\rho} ,$$

и поэтому

$$w_0 + \sum_{k=1}^{n} \rho_k w_k = \frac{w_0}{1-\rho} .$$

Откуда получается закон сохранения незавершенной работы

$$\sum_{k=1}^{n} \rho_k w_k = \frac{\rho}{1-\rho} w_0 = \frac{\rho}{1-\rho} \cdot \frac{1}{2} \sum_{k=1}^{n} \lambda_k b_k^{(2)} .$$

Он дает хоть что-то относительно времен ожидания.

В случае, когда $T_i \neq 0$, показывается, что между $\bar{u}$ и $\bar{u}_c$ существует соотношение

$$\overline{u} = \overline{u}_c + M\{Y\},$$

где $M\{Y\}$ – среднее количество незавершенной работы на периодах переключения. Отсюда получается закон сохранения

$$\sum_{k=1}^{n} \rho_k w_k = \frac{\rho}{2(1-\rho)} \sum_{k=1}^{n} \lambda_k b_k^{(2)} + M\{Y\}.$$

Величину $M\{Y\}$ можно рассчитать. Она зависит от дисциплины ухода на обслуживание другой очереди.

Глава 8. ИССЛЕДОВАНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ КОМПЬЮТЕРНЫХ СЕТЕЙ СВЯЗИ, УПРАВЛЯЕМЫХ ПРОТОКОЛАМИ СЛУЧАЙНОГО МНОЖЕСТВЕННОГО ДОСТУПА. МЕТОД АСИМПТОТИЧЕСКОГО АНАЛИЗА

§ 8.1. Математическая модель сети случайного доступа

Компьютерные сети связи, управляемые протоколами случайного множественного доступа, реализуются, как правило, на топологиях шины (рис. 8.1, *а*) либо звезды (рис. 8.1, *б*). Звезда используется при созда-

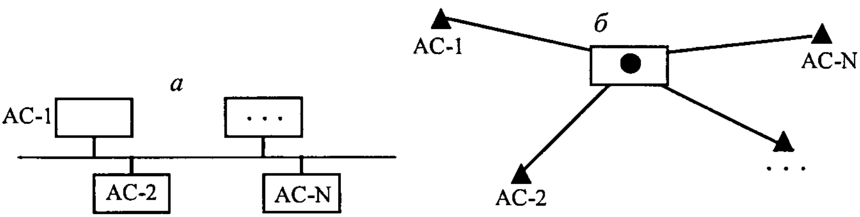


Рис. 8.1. Блок-схема топологии шина (*а*) и топология звезда (*б*)

нии космических (спутниковых) сетей связи (рис. 8.2), в которых спутник-ретранслятор исполняет роль центрального узла сети связи. Топология шины применятся при создании офисных, корпоративных или региональных сетей связи.

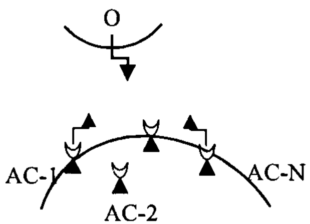


Рис. 8.2. Схема космической связи

В сетях случайного доступа обязательно наличие общего ресурса (центрального узла в звезде или моноканала в шине), который совместно используется всеми абонентскими станциями (АС) на правах конкуренции за захват общего ресурса.

Случайный множественный доступ в сетях связи реализуется следующим образом. Абонентская станция, сформировав сообщение, отправляет его в общий ресурс сети (центральный узел или моноканал), где происходит его обработка, если ресурс свободен. Если ресурс занят другим сообщением, то оба сообщения искажаются (говорят, что попадают в конфликт) и возвращаются на свои абонентские станции, где осуществляется их задержка случайной длительности, после которой АС повторяет попытку захвата общего ресурса. Далее процедура повторяется. В случае успешной обработки (передачи) сообщения в общем ресурсе сети считается, что сеанс связи реализован успешно.

Математическую модель сети случайного доступа определим в виде однолинейной СМО с источником повторных вызовов, на вход которой поступает поток заявок (рис. 8.3).

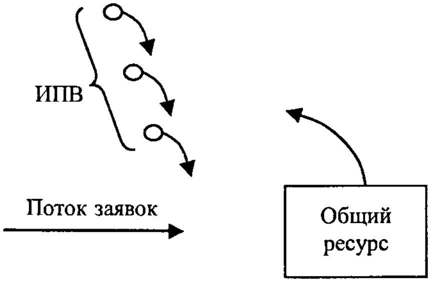


Рис. 8.3

Здесь в обозначениях Кендалла $\alpha|\beta|\gamma|\delta$ для сетей связи случайного доступа имеем:

α) входящий поток заявок – это суммарный поток сообщений от всех АС;

β) обслуживающий прибор – общий ресурс сети (центральный узел или моноканал);

γ) структура сети, заданная рис. 8.3;

δ) дисциплина обслуживания, определяемая процедурой повторения попыток захвата общего ресурса при возникновении конфликтов (протокол случайного множественного доступа).

Возможны различные модификации протокола случайного множественного доступа при наличии технических возможностей аппаратуры сети либо соответствующего алгоритмического и программного обеспечения.

1. Сигнал оповещения о конфликте (сигнал заглушки). При возникновении конфликта рассылается сигнал оповещения всех АС о том, что произошло искажение сообщений в результате их наложений в общем ресурсе сети.
2. Резервирование. Для уменьшения вероятности возникновения конфликта абонентская станция, сформировав сообщение, посылает сигнал-запрос на резервирование общего ресурса. При резервировании возможны конфликты с другими запросами на

резервирование. Если резервирование прошло успешно, то передача сообщения реализуется без конфликтов.

8.1.1. Математическая модель сети случайного доступа с оповещением о конфликте

Рассмотрим однолинейную СМО, на вход которой поступает простейший поток заявок с параметром ρ (рис. 8.4). Если в момент поступления заявки прибор свободен, то заявка начинает обслуживаться. Продолжительность обслуживания случайная с функцией распределения $B(x)$. Если за время обслуживания другие заявки не поступали, то полагают, что обслуживание прошло успешно и заявка покидает систему. Если во время обслуживания поступает другая заявка, то обе заявки считаются искаженными и переходят в источник повторных вызовов. От этого момента начинает рассылаться сигнал оповещения о конфликте случайной продолжительности с функцией распределения $A(x)$.

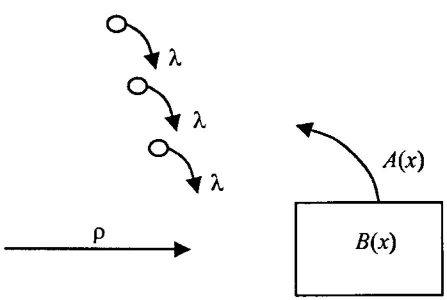


Рис. 8.4

Заявки, поступившие во время рассылки сигнала оповещения, также переходят в ИПВ. Продолжительность задержки в ИПВ случайная и имеет экспоненциальное распределение с одним и тем же параметром γ для всех заявок. Времена задержки разных заявок стохастически независимы. После случайной задержки в ИПВ заявка вновь обращается к прибору с повторной попыткой его захвата.

Будем различать три класса протоколов доступа в компьютерных сетях связи:

1. Если параметр γ является постоянной величиной, не зависящей от состояний системы, то протокол будем называть статическим.

2. Если параметр γ имеет вид $\gamma = \gamma_1/i$, где i – число заявок в ИПВ, то протокол назовем динамическим. Отметим, что в этом случае распределение вероятностей времени задержки заявки в ИПВ будет не экспоненциальным и определяется состоянием системы, но поток заявок, выходящих из ИПВ, принадлежит классу марковских потоков.

3. Если параметр γ имеет вид $\gamma = \gamma_1 / T$, где T – состояние адаптера (эту конструкцию опишем ниже), то протокол будем называть адаптивным.

Ниже проведем исследование всех трех классов компьютерных сетей случайного доступа, управляемых статическим, динамическим либо адаптивным протоколом доступа.

Если функции распределения $A(x)$ и $B(x)$ произвольные, то построенная СМО называется немарковской моделью сети случайного доступа с оповещением о конфликте. Если $A(x)$ и $B(x)$ экспоненциальные, то модель называется марковской.

Далее будем полагать

$$B(x) = 1 - e^{-\mu x}, \quad \mu = 1,$$

$$A(x) = 1 - e^{-\mu_1 x}, \quad \mu_1 = 1/a,$$

то есть рассматривать марковскую модель.

Аналогично, методом дополнительной переменной, можно провести исследование немарковской модели рассмотренной сети случайного доступа.

Для рассматриваемой модели состояния определим вектор $(k,\ i)$, где i – число заявок в ИПВ, а k характеризует состояние канала:

$$k = \begin{cases} 0, & \text{если прибор свободен,} \\ 1, & \text{если прибор занят обслуживанием,} \\ 2, & \text{если прибор находится в режиме оповещения.} \end{cases}$$

Процесс $\{k(t), i(t)\}$ изменения во времени состояний системы для марковской модели является марковским, а для немарковской модели – немарковским, и в этом случае необходимо определить дополнительную компоненту $z(t)$, равную длине интервала от момента t до момента окончания текущего режима функционирования прибора, если $k(t) = 1$ или 2, тогда процесс $\{k(t), i(t), z(t)\}$ является марковским.

8.1.2. Математическая модель сети случайного доступа с протоколом Алоха

В отличие от предыдущей модели, заявки, попавшие в конфликт, остаются на приборе и продолжают обслуживание в искаженном режиме, завершив которое, переходят в ИПВ. Заявки, поступившие во время обслуживания искаженных заявок, также искажаются и остаются на приборе до завершения их обслуживания, после чего также переходят в ИПВ. Таким образом, роль сигнала оповещения о конфликте для сети Алоха играет обслуживание искаженных заявок. Очевидно, его продолжительность зависит от числа заявок в ИПВ, так как чем больше i, тем выше интенсивность поступления заявок из источника повторных вызовов и, следовательно, продолжительность периода занятости прибора искаженными заявками увеличивается. Такая процедура разрешения конфликта значительно усложняет методы исследования ее математических моделей как марковской, так и немарковской.

8.1.3. Математическая модель сети случайного доступа Ethernet

В сети случайного доступа Ethernet реализуется протокол доступа с резервированием канала связи и оповещением о конфликте. Следовательно, канал может находиться в одном из четырех состояний, которые определим следующим образом:

$$k = \begin{cases} 0, & \text{если прибор свободен,} \\ 1, & \text{если реализуется этап резервирования,} \\ 2, & \text{если прибор занят обслуживанием,} \\ c, & \text{если прибор находится в режиме оповещения.} \end{cases}$$

Заявками являются сигнал-запросы на резервирование, во время реализации которого возможны конфликты. От момента наступления конфликта начинает распространяться сигнал оповещения о конфликте.

Заявки, поступившие на этапах обслуживания заявки (передачи сообщения) и оповещения, переходят в ИПВ, не искажая текущих режимов на приборе. Отличие модели сети Ethernet от модели сети с оповещением незначительное и заключается в увеличении числа состояний прибора еще одним, характеризующим режим запроса на резервирование.

Из рассмотренных шести моделей наиболее простой является марковская модель сети с оповещением о конфликте. Именно исследованию этой модели и будет посвящено дальнейшее изложение.

Остальные модели можно исследовать аналогично, выполнив соответствующую модификацию, наиболее принципиальной из которых является введение дополнительной компоненты для немарковских моделей.

§ 8.2. Общий подход к исследованию марковских моделей сетей случайного доступа, управляемых статическими протоколами

Рассмотрим математическую модель сети случайного доступа в виде однолинейной марковской СМО с ИПВ вида, изображенного на рис. 8.5.

Обозначим состояния прибора:

$$k = \begin{cases} 0 - \text{свободен}, \\ 1 - \text{занят}, \\ 2 - \text{оповещение}. \end{cases}$$

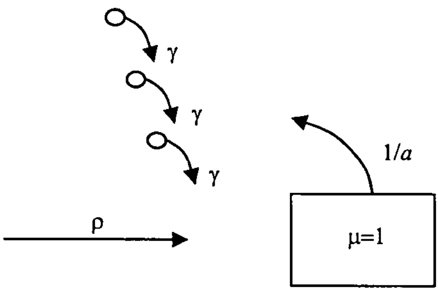


Рис. 8.5

Двумерный случайный процесс $\{k(t), i(t)\}$ – марковский, поэтому его распределения вероятностей

$$P_k(i,t) = P(k(t) = k, i(t) = i)$$

удовлетворяет системе уравнений

$$\frac{\partial P_0(i,t)}{\partial t} = -(\rho + i\gamma)P_0(i,t) + P_1(i,t) + \frac{1}{a}P_2(i,t),$$

$$\frac{\partial P_1(i,t)}{\partial t} = -(\rho + i\gamma + 1)P_1(i,t) + \rho P_0(i,t) + (i+1)\gamma P_0(i+1,t), \quad (8.1)$$

$$\frac{\partial P_2(i,t)}{\partial t} = -\left(\rho + \frac{1}{a}\right)P_2(i,t) + \rho P_2(i-1,t) + (i-1)\gamma P_1(i-1,t) + \rho P_1(i-2,t).$$

Обозначим вектор-столбец

$$P(i,t) = (P_0(i,t), P_1(i,t), P_2(i,t))^T,$$

и матрицы $A_0(i)$, $A_1(i)$, $B_1(i)$, B_2 определим таким образом, что систему (8.1) представим в виде

$$\frac{\partial P(i,t)}{\partial t} = A_0(\gamma i)P(i,t) + A_1(\gamma(i+1))P(i+1,t) + \\ +B_1(\gamma(i-1))P(i-1,t) + B_2 P(i-2,t). \tag{8.2}$$

Обозначим $\gamma = \varepsilon^2$, $\varepsilon^2 t = \tau$ и покажем, что

$$x(\tau) = \lim_{\varepsilon \to 0} \varepsilon^2 i(\tau/\varepsilon^2)$$

является детерминированной функцией,

$$y(\tau) = \lim_{\varepsilon \to 0} \frac{\varepsilon^2 i(\tau/\varepsilon^2) - x(\tau)}{\varepsilon}$$

– диффузионным процессом авторегрессии. Процесс изменения состояний канала $k(\tau/\varepsilon^2)$ при $\varepsilon \to 0$ является дискретным марковским процессом, независимым от процесса $y(\tau)$.

Используя предельные процессы $x(\tau)$ и $y(\tau)$, для достаточно малых ε рассмотрим процесс

$$z(\tau) = x(\tau) + \varepsilon y(\tau),$$

который с точностью до $o(\varepsilon)$ совпадает с процессом $\varepsilon^2 i(\tau/\varepsilon^2)$ и характеризует процесс изменения числа заявок в ИПВ.

Покажем, что процесс $z(\tau)$ является диффузионным, и найдем его коэффициенты переноса и диффузии.

Сформулированные результаты получим с помощью исследования системы (8.2), которая имеет место не только для рассматриваемой СМО, но также и для других сетей случайного доступа, математические модели которых можно представить в виде марковских СМО с источником повторных вызовов и конечным числом состояний канала.

Для получения указанных результатов в системе (8.2) выполним замены

$$\varepsilon^2 t = \tau, \; \varepsilon^2 i = x(\tau) + \varepsilon y, \; \frac{1}{\varepsilon} P(i,t) = H(y,\tau,\varepsilon), \tag{8.3}$$

где $x(\tau)$ – некоторая функция, вид которой будет определен ниже. Тогда систему (8.2) перепишем следующим образом:

$$\varepsilon^2 \frac{\partial H(y,\tau,\varepsilon)}{\partial \tau} - \varepsilon x'(\tau) \frac{\partial H(y,\tau,\varepsilon)}{\partial y} = A_0(x+\varepsilon y)H(y,\tau,\varepsilon) +$$

$$+A_1(x+\varepsilon(y+\varepsilon))H(y+\varepsilon,\tau,\varepsilon)+B_1(x+\varepsilon(y-\varepsilon))H(y-\varepsilon,\tau,\varepsilon)+ \\ +B_2H(y-2\varepsilon,\tau,\varepsilon). \tag{8.4}$$

Систему (8.4) будем решать в четыре этапа.

На первом этапе найдем распределение вероятностей значений процесса $k(\tau)$. Для этого в системе (8.4) перейдем к пределу при $\varepsilon\to 0$ и, обозначив $H(y,\tau,0)=H(y,\tau)$, получим относительно вектора $H(y,\tau)$ однородную систему линейных алгебраических уравнений

$$K(x)H(y,\tau)=0, \tag{8.5}$$

где матрица $K(x)$ имеет вид

$$K(x)=A_0(x)+A_1(x)+B_1(x)+B_2 \tag{8.6}$$

и является инфинитезимальной матрицей интенсивностей переходов случайного процесса $k(\tau)$. Из свойств таких матриц следует, что их строки линейно зависимы, так как

$$E^T K(x)=0, \tag{8.7}$$

где E – единичный вектор-столбец. Следовательно, система (8.5) имеет нетривиальное решение, которое представим в виде

$$H(y,\tau)=R(x)F(y,\tau), \tag{8.8}$$

где $F(y,\tau)$ – скалярная функция, а вектор $R(x)$ определяется аналогично (8.5) однородной системой линейных алгебраических уравнений

$$K(x)R(x)=0. \tag{8.9}$$

Положим, что вектор $R(x)$ удовлетворяет условию нормировки

$$E^T R(x)=1. \tag{8.10}$$

Тогда $R(x)$ имеет смысл распределения вероятностей значений процесса $k(\tau)$, а $F(y,\tau)$ является плотностью распределения вероятностей значений процесса $y(\tau)$, ее вид будет определен ниже.

Отметим, что в силу равенства (8.8) процессы $k(\tau)$ и $y(\tau)$ стохастически независимы.

Покажем, что решение $R(x)$ системы (8.9), удовлетворяющее условию нормировки (8.10), существует и единственно.

Теорема 1. Решение $R(x)$ системы (8.9), удовлетворяющее условию нормировки (8.10), существует и единственно.

Доказательство

Для доказательства теоремы формально воспользуемся теоремой эргодичности.

По инфинитезимальной матрице построим матрицу переходных вероятностей вложенной цепи Маркова с дискретным временем. Это реализуется исключением диагональных элементов матрицы $K(x)$ и нормировкой ее строк. Полученная матрица неприводима и непериодична, поэтому выполнены условия теоремы эргодичности для цепей Маркова с конечным числом состояний. Следовательно, существует единственное эргодическое распределение, которое для процесса с непрерывным временем определяется системой (8.9) и условием нормировки (8.10).

Таким образом, решение $R(x)$ однородной системой линейных алгебраических уравнений (8.9), удовлетворяющих условию (8.10), существует и единственно.

Теорема доказана.

Еще раз подчеркнем, что приведенное доказательство использует теорему эргодичности не по существу, а чисто формально, так как распределение $R(x)$ зависит от значений функции $x(\tau)$ и, следовательно, не является, вообще говоря, стационарным.

Решение системы (8.9), удовлетворяющее условию (8.10), можно найти достаточно просто, хотя оно имеет довольно громоздкий вид, а записать его в явном виде возможно следующим образом.

Одно из уравнений системы (8.9), например первое, заменим на условие (8.10). Тогда систему (8.9) – (8.10) перепишем в виде

$$\tilde{K}(x)R(x)=\theta_1, \tag{8.11}$$

где θ_1 – вектор, в котором первая компонента равна 1, а остальные нулю, в матрице $\tilde{K}(x)$ первая строка является единичным вектором, а остальные элементы совпадают с элементами матрицы $K(x)$. В силу доказанной теоремы матрица $\tilde{K}(x)$ невырождена, поэтому из (8.11) следует, что для нее существует обратная матрица. В результате

$$R(x)=\tilde{K}^{-1}(x)\theta_1. \tag{8.12}$$

В частности, для рассматриваемой системы (8.1) компоненты $R_k(x)$ вектора $R(x)$ имеют вид

$$R_0(x)=\frac{\rho+x+1}{a(\rho+x)^2+2(\rho+x)+1},\quad R_1(x)=\frac{\rho+x}{a(\rho+x)^2+2(\rho+x)+1},$$

$$R_2(x)=\frac{a(\rho+x)^2}{a(\rho+x)^2+2(\rho+x)+1}.$$

На втором этапе решения системы (8.4) найдем вид функции $x=x(\tau)$.

Теорема 2. Функция $x=x(\tau)$ является решением обыкновенного дифференциального уравнения

$$x'(\tau)=E^T V(x)R(x), \tag{8.13}$$

где матрица $V(x)$ имеет вид

$$V(x)=B_1(x)+2B_2-A_1(x). \tag{8.14}$$

Доказательство

Функции в правой части системы (8.4) разложим в ряд по приращениям аргумента y с точностью до $o(\varepsilon)$ и перепишем эту систему в виде

$$-\varepsilon x'(\tau)\frac{\partial H(y,\tau,\varepsilon)}{\partial y}=K(x+\varepsilon y)H(y,\tau,\varepsilon)-\varepsilon V(x)\frac{\partial H(y,\tau,\varepsilon)}{\partial y}+o(\varepsilon), \tag{8.15}$$

где матрица $V(x)$, очевидно, имеет вид (8.14).

Просуммируем все уравнения системы (8.15) и, учитывая свойство (8.7) матрицы $K(x)$, получим

$$-\varepsilon x'(\tau)E^T\frac{\partial H(y,\tau,\varepsilon)}{\partial y}=-\varepsilon E^T V(x)\frac{\partial H(y,\tau,\varepsilon)}{\partial y}+o(\varepsilon).$$

Поделив левую и правую части этого равенства на ε и полагая $\varepsilon\to 0$, запишем

$$x'(\tau)E^T\frac{\partial H(y,\tau)}{\partial y}=E^T V(x)\frac{\partial H(y,\tau)}{\partial y}.$$

Подставляя в это равенство $H(y,\tau)$ в виде (8.8), получим

$$x'(\tau)E^T R(x)\frac{\partial F(y,\tau)}{\partial y}=E^T V(x)R(x)\frac{\partial F(y,\tau)}{\partial y},$$

откуда, учитывая условие нормировки (8.10), получим обыкновенное дифференциальное уравнение

$$x'(\tau) = E^T V(x) R(x), \tag{8.16}$$

определяющее вид функции $x = x(\tau)$.

Теорема доказана.

На третьем этапе найдем разложение функции $H(y,\tau,\varepsilon)$ в виде

$$H(y,\tau,\varepsilon) = R(x)F(y,\tau) + \varepsilon h(y,\tau) + o(\varepsilon). \tag{8.17}$$

Теорема 3. Вектор $h(y,\tau)$ является решением неоднородной системы линейных алгебраических уравнений

$$K(x)h(y,\tau) = \left[V(x) - x'(\tau)I\right]R(x)\frac{\partial F(y,\tau)}{\partial y} - K'(x)R(x)yF(y,\tau), \tag{8.18}$$

где I – диагональная единичная матрица, а $x'(\tau)$ определяется равенством (8.16).

Доказательство

Систему (8.15) перепишем в виде

$$K(x+\varepsilon y)H(y,\tau,\varepsilon) = \varepsilon\left[V(x) - x'(\tau)I\right]\frac{\partial H(y,\tau,\varepsilon)}{\partial y} + o(\varepsilon)$$

и подставим в это равенство разложение матрицы $K(x+\varepsilon y)$ в виде

$$K(x+\varepsilon y) = K(x) + \varepsilon y K'(x).$$

Тогда получим

$$K(x)H(y,\tau,\varepsilon) + \varepsilon y K'(x)H(y,\tau,\varepsilon) = \varepsilon\left[V(x) - x'(\tau)I\right]\frac{\partial H(y,\tau,\varepsilon)}{\partial y} + o(\varepsilon).$$

Подставив в это равенство разложение (8.17), получим

$$K(x)R(x)F(y,\tau) + \varepsilon K(x)h(y,\tau) + \varepsilon y K'(x)R(x)F(y,\tau) =$$

$$= \varepsilon\left[V(x) - x'(\tau)I\right]R(x)\frac{\partial F(y,\tau)}{\partial y} + o(\varepsilon).$$

Учитывая равенство (8.9) и выполнив несложные преобразования, последнее равенство перепишем в виде

$$K(x)h(y,\tau) = \left[V(x) - x'(\tau)I\right]R(x)\frac{\partial F(y,\tau)}{\partial y} - K'(x)R(x)yF(y,\tau),$$

совпадающем с (8.18).

Матрица $K(x)$ неоднородной системы (8.18) имеет определитель, равный нулю, поэтому данная система линейных алгебраических уравнений имеет решение только тогда, когда ранг собственной матрицы $K(x)$ системы совпадает с рангом расширенной матрицы.

В силу теоремы 1 ранг матрицы $K(x)$ на единицу меньше ее размерности. Покажем, что уравнения неоднородной системы (8.18) линейно зависимы.

Суммируя все уравнения системы (8.18), получим

$$E^T K(x)h(y,\tau) = E^T\left[V(x) - x'(\tau)I\right]R(x)\frac{\partial F(y,\tau)}{\partial y} - E^T K'(x)R(x)yF(y,\tau) =$$

$$= \left[E^T V(x)R(x) - x'(\tau)E^T R(x)\right]\frac{\partial F(y,\tau)}{\partial y} - \frac{\partial}{\partial x}\left\{E^T K(x)\right\}yF(y,\tau).$$

В силу равенства (8.7) левая часть равна нулю. В силу равенств (8.7), (8.10) и (8.16) правая часть также равна нулю, следовательно, ранги соответствующих матриц совпадают, а система (8.18) имеет решение, определяемое с точностью до однопараметрического семейства векторов вида $R(x)C$, где C – произвольная скалярная величина.

Следствие 1. Решение $h(y,\tau)$ системы (8.18) имеет вид

$$h(y,\tau) = h^{(1)}(x)\frac{\partial F(y,\tau)}{\partial y} + R'(x)yF(y,\tau) + R(x)C, \qquad (8.19)$$

где вектор $h^{(1)}(x)$ является решением системы

$$K(x)h^{(1)}(x) = \left[V(x) - x'(\tau)I\right]R(x). \qquad (8.20)$$

Доказательство

Выражение (8.19) подставим в систему (8.18) и получим

$$K(x)h^{(1)}(x)\frac{\partial F(y,\tau)}{\partial y} + K(x)R'(x)yF(y,\tau) + K(x)R(x)C =$$

$$= \left[V(x) - x'(\tau)I\right]R(x)\frac{\partial F(y,\tau)}{\partial y} - K'(x)R(x)yF(y,\tau).$$

В силу (8.9) и (8.20) это равенство можно переписать в виде

$$K(x)R'(x)yF(y,\tau) = -K'(x)R(x)yF(y,\tau),$$

откуда получим равенство

$$\left(K(x)R(x)\right)' = 0,$$

которое выполняется тождественно в силу равенства (8.9).

Следствие доказано.

Следствие 2. Частное решение $h^{(1)}(x)$ системы (8.20) имеет вид

$$h^{(1)}(x)=\tilde{K}^{-1}(x)\left[V(x)-x'(\tau)I\right]R(x)\,, \tag{8.21}$$

где $\tilde{K}(x)$ – та же матрица, что и в (8.12).

Доказательство

Аналогично (8.18), система (8.20) имеет бесконечное множество решений, из которых, в силу (8.19), достаточно выделить единственное частное решение, наложив одно дополнительное условие, например положив, что $E^T h(y,\tau)$ равно первой компоненте вектора свободных членов системы (8.20). Тогда это частное решение определяется системой уравнений

$$\tilde{K}(x)h^{(1)}(x)=\left[V(x)x'(\tau)I\right]R(x)$$

с невырожденной матрицей $\tilde{K}(x)$. Поэтому

$$h^{(1)}(x)=\tilde{K}^{-1}(x)\left[V(x)-x'(\tau)I\right]R(x)\,.$$

Следствие доказано.

На четвертом этапе найдем вид функции $F(y,\tau)$.

Теорема 4. Функция $F(y,\tau)$ является плотностью распределения вероятностей значений диффузионного процесса авторегрессии и определяется уравнением Фоккера – Планка вида

$$\frac{\partial F(y,\tau)}{\partial\tau}=-\left(E^T V(x)R(x)\right)'\frac{\partial(yF(y,\tau))}{\partial y}+$$

$$+\frac{1}{2}E^T\left\{\,D(x)-2\left(V(x)-x'(\tau)I\right)\tilde{K}^{-1}(x)\left(V(x)-x'(\tau)I\right)\right\}\,R(x)\frac{\partial^2 F(y,\tau)}{\partial y^2}\,, \tag{8.22}$$

где матрица $D(x)$ имеет вид

$$D(x)=B_1(x)+4B_2+A_1(x)\,. \tag{8.23}$$

Доказательство

Разлагая в ряд функции в правой части системы (8.4) по приращениям аргумента y с точностью до $o(\varepsilon^2)$, получим

$$\varepsilon^2 \frac{\partial H(y,\tau,\varepsilon)}{\partial \tau} - \varepsilon x'(\tau)\frac{\partial H(y,\tau,\varepsilon)}{\partial y} = K(x+\varepsilon y)H(y,\tau,\varepsilon) -$$

$$-\varepsilon\frac{\partial}{\partial y}\{V(x+\varepsilon y)H(y,\tau,\varepsilon)\} + \frac{\varepsilon^2}{2}D(x)\frac{\partial^2 H(y,\tau,\varepsilon)}{\partial y^2} + o(\varepsilon^2),$$

где матрица $D(x)$, очевидно, имеет вид (8.23).

Сложив все уравнения этой системы и используя свойство матрицы K, получим

$$\varepsilon^2 E^T \frac{\partial H(y,\tau,\varepsilon)}{\partial \tau} - \varepsilon x'(\tau)E^T \frac{\partial H(y,\tau,\varepsilon)}{\partial y} = -\varepsilon E^T \frac{\partial}{\partial y}\{V(x+\varepsilon y)H(y,\tau,\varepsilon)\} +$$

$$+\frac{\varepsilon^2}{2}E^T D(x)\frac{\partial^2 H(y,\tau,\varepsilon)}{\partial y^2} + o(\varepsilon^2).$$

Используя разложение

$$V(x+\varepsilon y) = V(x) + \varepsilon y V'(x),$$

перепишем это равенство следующим образом:

$$\varepsilon^2 E^T \frac{\partial H(y,\tau,\varepsilon)}{\partial \tau} - \varepsilon x'(\tau)E^T \frac{\partial H(y,\tau,\varepsilon)}{\partial y} = -\varepsilon E^T V(x)\frac{\partial H(y,\tau,\varepsilon)}{\partial y} -$$

$$-\varepsilon^2 E^T V'(x)\frac{\partial\left(yH(y,\tau,\varepsilon)\right)}{\partial y} + \frac{\varepsilon^2}{2}E^T D(x)\frac{\partial^2 H(y,\tau,\varepsilon)}{\partial y^2} + o(\varepsilon^2).$$

Подставляя в это равенство разложение (8.17) функции $H(y,\tau,\varepsilon)$, получим равенство

$$\varepsilon^2 E^T R(x)\frac{\partial F(y,\tau)}{\partial \tau} - \varepsilon x'(\tau)E^T R(x)\frac{\partial F(y,\tau)}{\partial y} - \varepsilon^2 x'(\tau)E^T \frac{\partial h(y,\tau)}{\partial y} =$$

$$= -\varepsilon E^T V(x)R(x)\frac{\partial F(y,\tau)}{\partial y} - \varepsilon^2 E^T V(x)\frac{\partial h(y,\tau)}{\partial y} -$$

$$-\varepsilon^2 E^T V'(x)R(x)\frac{\partial\left(yF(y,\tau)\right)}{\partial y} + \frac{\varepsilon^2}{2}E^T D(x)R(x)\frac{\partial^2 F(y,\tau)}{\partial y^2} + o(\varepsilon^2),$$

которое, в силу условия нормировки (8.10) и дифференциального уравнения (8.16), перепишем следующим образом:

$$\varepsilon^2 \frac{\partial F(y,\tau)}{\partial \tau} - \varepsilon^2 x'(\tau) E^T \frac{\partial h(y,\tau)}{\partial y} = -\varepsilon^2 E^T V(x) \frac{\partial h(y,\tau)}{\partial y} -$$

$$-\varepsilon^2 E^T V'(x) R(x) \frac{\partial \left(yF(y,\tau) \right)}{\partial y} + \frac{\varepsilon^2}{2} E^T D(x) R(x) \frac{\partial^2 F(y,\tau)}{\partial y^2} + o(\varepsilon^2).$$

Выполнив несложные преобразования в этом равенстве, получим

$$\frac{\partial F(y,\tau)}{\partial \tau} = -E^T \left(V(x) - x'(\tau) I \right) \frac{\partial h(y,\tau)}{\partial y} - E^T V'(x) R(x) \frac{\partial \left(yF(y,\tau) \right)}{\partial y} +$$

$$+\frac{1}{2} E^T D(x) R(x) \frac{\partial^2 F(y,\tau)}{\partial y^2}. \qquad (8.24)$$

Рассмотрим первое слагаемое в правой части этого равенства. Подставляя в него представление (8.19) вектора $h(y,\tau)$, получим

$$E^T \left(V(x) - x'(\tau) I \right) \frac{\partial h(y,\tau)}{\partial y} = E^T \left(V(x) - x'(\tau) I \right) h^{(1)}(x) \frac{\partial^2 F(y,\tau)}{\partial y^2} +$$

$$+E^T \left(V(x) - x'(\tau) I \right) R'(x) \frac{\partial \left(yF(y,\tau) \right)}{\partial y}. \qquad (8.25)$$

Здесь

$$E^T \left(V(x) - x'(\tau) I \right) R'(x) = E^T V(x) R'(x) - E^T x'(\tau) R'(x) =$$

$$= E^T V(x) R'(x) - x'(\tau) \left(E^T R(x) \right)'.$$

Используя равенство (8.21), запишем

$$E^T \left(V(x) - x'(\tau) I \right) h^{(1)}(x) = E^T \left(V(x) - x'(\tau) I \right) \tilde{K}^{-1}(x) \left(V(x) - x'(\tau) I \right) R(x).$$

Подставляя эти выражения в (8.25), получим

$$E^T \left(V(x) - x'(\tau) I \right) \frac{\partial h(y,\tau)}{\partial y} = E^T V(x) R'(x) \frac{\partial \left(yF(y,\tau) \right)}{\partial y} +$$

$$+E^T \left(V(x) - x'(\tau) I \right) \tilde{K}^{-1} \left(V(x) - x'(\tau) I \right) R(x) \frac{\partial^2 F(y,\tau)}{\partial y^2}.$$

Подставляя это выражение в (8.24), получим равенство

$$\frac{\partial F(y,\tau)}{\partial \tau} = -E^T V(x) R'(x) \frac{\partial\left(yF(y,\tau)\right)}{\partial y} - E^T V'(x) R(x) \frac{\partial\left(yF(y,\tau)\right)}{\partial y} -$$

$$-E^T\left(V(x) - x'(\tau)I\right)\tilde{K}^{-1}\left(V(x) - x'(\tau)I\right)R(x)\frac{\partial^2 F(y,\tau)}{\partial y^2} +$$

$$+\frac{1}{2}E^T D(x) R(x) \frac{\partial^2 F(y,\tau)}{\partial y^2},$$

которое очевидно совпадает с уравнением (8.22).

Теорема доказана.

В силу (8.13) коэффициент диффузии случайного процесса $y(\tau)$ можно переписать в виде

$$B^2(x) = E^T\left\{D(x) - 2\left(V(x) - E^T V(x) R(x)\right)\tilde{K}^{-1}\left(V(x) - E^T V(x) R(x)\right)\right\}R(x),$$

а для диффузионного процесса $y(\tau)$ соответствующее стохастическое дифференциальное уравнение имеет вид

$$dy(\tau) = \left(E^T V(x) R(x)\right)' y(\tau) d\tau + B(x) dw(\tau),$$

где $w(\tau)$ – стандартный винеровский процесс.

Теперь рассмотрим для достаточно малых значений параметра ε случайный процесс

$$z(\tau) = x(\tau) + \varepsilon y(\tau), \quad z = \varepsilon^2 i(\tau/\varepsilon^2) \tag{8.26}$$

и докажем следующую теорему:

Теорема 5. С точностью до $o(\varepsilon)$ случайный процесс $z(\tau)$ является решением стохастического дифференциального уравнения

$$dz(\tau) = E^T V(z) R(z) d\tau + \varepsilon B(z) dw(\tau), \tag{8.27}$$

то есть $z(\tau)$ является однородным диффузионным процессом с коэффициентами переноса $E^T V(z) R(z)$ и диффузии $\varepsilon^2 B^2(z)$.

Доказательство

Дифференцируя равенство (8.26), получим $dz(\tau) = x'(\tau)d\tau + \varepsilon dy(\tau)$. В силу (8.13) и (8.25) его правую часть перепишем в виде

$$x'(\tau)d\tau + \varepsilon dy(\tau) = E^T V(x) R(x) d\tau + \varepsilon\left(E^T V(x) R(x)\right)' y d\tau + \varepsilon B(x) dw(\tau) =$$

$$=\left\{ E^T V(x)R(x) + \varepsilon y\left(E^T V(x)R(x)\right)' \right\} d\tau + \varepsilon B(x)dw(\tau) =$$

$$= E^T V(x+\varepsilon y)R(x+\varepsilon y)d\tau + \varepsilon B(x+\varepsilon y)dw(\tau) + o(\varepsilon) =$$

$$= E^T V(z)R(z)d\tau + \varepsilon B(z)dw(\tau) + o(\varepsilon),$$

следовательно, процесс $z(\tau)$ с точностью до $o(\varepsilon)$ удовлетворяет стохастическому дифференциальному уравнению (8.27).

Теорема доказана.

Таким образом, для достаточно малых значений параметра ε случайный процесс $\varepsilon^2 i(\tau/\varepsilon^2)$ можно аппроксимировать однородным диффузионным процессом $z(\tau)$, удовлетворяющим стохастическому дифференциальному уравнению (8.27).

§ 8.3. Исследование сетей случайного доступа с динамическими протоколами

Рассмотрим компьютерную сеть связи с оповещением о конфликте, управляемую динамическим протоколом случайного множественного доступа. Структура математической модели такой сети изображена на рис. 8.6.

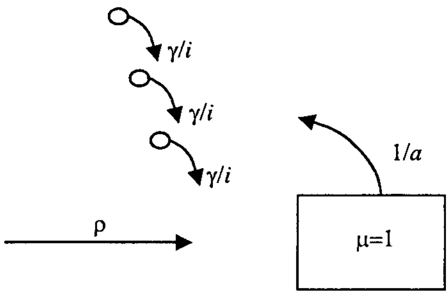


Рис. 8.6

Здесь рассматривается однолинейная СМО, на вход которой поступает простейший поток заявок с параметром ρ. Заявка, заставшая прибор свободным, начинает обслуживаться. Продолжительность обслуживания случайная, имеет экспоненциальное распределение с параметром $\mu = 1$. Если прибор занят, то обслуживаемая и поступившая заявки вступают в конфликт и переходят в источник повторных вызовов. От момента возникновения конфликта в канале реализуется сигнал оповещения о конфликте, продолжительность которого случайная и также имеет экспоненциальное распределение с параметром $1/a$. По завершении этапа оповещения канал вновь становится свободным. Заявки, поступившие во время оповещения, также переходят в ИПВ. Все заявки в

ИПВ осуществляют случайную задержку, после которой повторно обращаются к обслуживающему прибору.

Будем полагать, что для каждой заявки в ИПВ вероятность окончания случайной задержки за бесконечно малый промежуток времени длительности Δt равна $\frac{\gamma}{i}\Delta t + o(\Delta t)$, где i – число заявок в ИПВ, а γ – некоторый положительный параметр, характеризующий продолжительность случайной задержки в ИПВ.

Определим состояние СМО вектором (k,i), где k – состояние канала. Очевидно, процесс $\{k(t),i(t)\}$ изменения состояний СМО во времени является двумерным марковским процессом с дискретным множеством состояний. Его распределение вероятностей

$$P(k(t)=k, i(t)=i) = P_k(i,t)$$

удовлетворяет следующим равенствам:

$$P_0(i,t+\Delta t) = [1-(\rho+\gamma)\Delta t]P_0(i,t) + \Delta t P_1(i,t) + \frac{1}{a}\Delta t P_2(i,t) + o(\Delta t),$$

$$P_1(i,t+\Delta t) = [1-(\rho+\gamma+1)\Delta t]P_1(i,t) + \rho\Delta t P_0(i,t) + \gamma\Delta t P_0(i+1,t) + o(\Delta t),$$

$$P_2(i,t+\Delta t) = \left[1-\left(\rho+\frac{1}{a}\right)\Delta t\right]P_2(i,t) + \rho\Delta t P_2(i-1,t) + \gamma\Delta t P_1(i-1,t) +$$
$$+\rho\Delta t P_1(i-2,t) + o(\Delta t).$$

Отсюда, выполнив несложные преобразования, для стационарного распределения $P_k(i,t) = P_k(i)$ получаем систему уравнений

$$-(\rho+\gamma)P_0(i) + P_1(i) + \frac{1}{a}P_2(i) = 0,$$

$$-(\rho+\gamma+1)P_1(i) + \rho P_0(i) + \gamma P_0(i+1) = 0,$$

$$-\left(\rho+\frac{1}{a}\right)P_2(i) + \rho P_2(i-1) + \gamma P_1(i-1) + \rho P_1(i-2) = 0. \qquad (8.28)$$

Обозначив вектор $P(i) = \{P_0(i), P_1(i), P_2(i)\}^T$, систему (8.28) перепишем в виде

$$A_0(\rho)P(i) + A_1 P(i+1) + B_1(\rho)P(i-1) + B_2 P(i-2) = 0, \qquad (8.29)$$

где постоянные относительно i, не зависящие от величины ρ, матрицы A_0, A_1, B_1, B_2 определяются системой (8.28). Нетрудно выписать явный вид этих матриц.

Отметим, что совершенно аналогично можно выписать уравнение (8.29) для марковских моделей и других сетей случайного доступа, например сети Ethernet.

Систему (8.29) будем решать методом асимптотического анализа в условиях большой загрузки, полагая, что

$$\rho = s - \varepsilon,$$

где s – величина пропускной способности сети, ее значение определим ниже, а ε – малый положительный параметр, рассматриваемый в асимптотических условиях $\varepsilon \to 0$.

В системе (8.29) выполним замену $i\varepsilon = x$, $\frac{1}{\varepsilon}P(i) = \pi(x,\varepsilon)$ и получим

$$A_0(\rho)\pi(x,\varepsilon) + A_1\pi(x+\varepsilon,\varepsilon) + B_1(\rho)\pi(x-\varepsilon,\varepsilon) + B_2\pi(x-2\varepsilon,\varepsilon) = 0. \quad (8.30)$$

Уравнение (8.30) будем решать в четыре этапа.

Этап 1. Положив в (8.30) $\varepsilon \to 0$, $\rho \to s$ и обозначив $\pi(x,0) = \pi(x)$, уравнение (8.30) перепишем в виде

$$K(s)\pi(x) = 0, \quad (8.31)$$

где $K(s) = A_0(s) + A_1 + B_1(s) + B_2(s)$.

Уравнение (8.31) является однородной системой линейных алгебраических уравнений относительно векторной функции $\pi(x)$ с определителем матрицы $K(s)$, равным нулю, так как выполняется равенство

$$E^T K(s) = 0, \quad (8.32)$$

где E – единичный вектор.

Решение $\pi(x)$ системы (8.31) найдем в виде

$$\pi(x) = R(x)H(x), \quad (8.33)$$

где $H(x)$ – скалярная функция, а вектор $R(x)$ определяется системой

$$R(s)K(s) = 0 \quad (8.34)$$

и условием нормировки

$$E^T R(s) \equiv 1. \quad (8.35)$$

Отметим, что вектор $R(s)$ имеет смысл асимптотического распределения вероятностей состояний канала, то есть стационарного распределения вероятностей значений процесса $k(\tau/\varepsilon)$ при $\varepsilon \to 0$.

Этап 2. В уравнении (8.30) разложим функции в ряд по приращениям аргумента x с точностью до $o(\varepsilon)$, получим

$$K(\rho)\pi(x,\varepsilon) - \varepsilon V(s)\frac{\partial \pi(x,\varepsilon)}{\partial x} = o(\varepsilon), \qquad (8.36)$$

где матрица $V(s)$ имеет вид

$$V(s) = B_1(s) + 2B_2(s) - A_1. \qquad (8.37)$$

Представим матрицу $K(\rho)$ в виде

$$K(\rho) = K(s) - \varepsilon\frac{\partial K(s)}{\partial s}, \qquad (8.38)$$

подставим в (8.36), получим

$$K(s)\pi(x,\varepsilon) = \varepsilon\frac{\partial K(s)}{\partial s}\pi(x,\varepsilon) + \varepsilon V(s)\frac{\partial \pi(x,\varepsilon)}{\partial x} + o(\varepsilon). \qquad (8.39)$$

Сложив все уравнения системы (8.39) и учитывая свойство (8.32) матрицы $K(s)$, получим

$$\varepsilon E^T V(s)\frac{\partial \pi(x,\varepsilon)}{\partial x} = o(\varepsilon).$$

Поделив первую и правую часть этого равенства на ε и полагая $\varepsilon \to 0$, получим

$$E^T V(s)\frac{\partial \pi(x)}{\partial x} = 0.$$

Подставив сюда вид (8.33) для вектора $\pi(x)$, запишем равенство

$$E^T V(s)R(s)H'(x) = 0.$$

Так как производная $H'(x)$ плотности распределения вероятностей не может тождественно равняться нулю, то имеет место равенство

$$E^T V(s)R(s) = 0, \qquad (8.40)$$

которое определяет значение s пропускной способности рассматриваемой СМО.

Таким образом, значение s является решением уравнения (8.40).

Этап 3. Решение $\pi(x,\varepsilon)$ системы (8.36) найдем в виде

$$\pi(x,\varepsilon) = R(s)H(x) + \varepsilon h(x) + o(\varepsilon). \tag{8.41}$$

Подставляя (8.38) и (8.41) в (8.36), получим

$$\left(K(s) - \varepsilon\frac{\partial K(s)}{\partial s}\right)\left(R(s)H(x) + \varepsilon h(x)\right) - \varepsilon V(s)R(s)H'(x) = o(\varepsilon).$$

Выполнив в этом равенстве несложные преобразования и учитывая равенство (8.34), получим систему

$$K(s)h(x) = \frac{\partial K(s)}{\partial s}R(s)H(x) + V(s)R(s)H'(x), \tag{8.42}$$

которая является неоднородной системой линейных алгебраических уравнений относительно компонент вектора $h(x)$, в которой ранги собственной матрицы $K(s)$ и расширенной матрицы системы совпадают, поэтому система (8.42) имеет решение.

Решение $h(x)$ системы (8.42) будем искать в виде

$$h(x) = h^{(2)}(s)H(x) + h^{(1)}(s)H'(x).$$

В результате получим, что $h^{(1)}(s)$ удовлетворяет системе

$$K(s)h^{(1)}(x) = V(s)R(s), \tag{8.43}$$

а $h^{(2)}(s)$ – следующей системе:

$$K(s)h^{(2)}(x) = \frac{\partial K(s)}{\partial s}R(s).$$

Отсюда имеем

$$\frac{\partial K(s)}{\partial s}R(s) - K(s)h^{(2)}(x) = 0.$$

Очевидно, положив $h^{(2)}(x) = -\frac{\partial R(s)}{\partial s}$, получим

$$\frac{\partial K(s)}{\partial s}R(s) + K(s)\frac{\partial R(s)}{\partial s} = \frac{\partial}{\partial s}\left(K(s)R(s)\right) = 0.$$

Следовательно, $h(x)$ имеет вид

$$h(x) = -\frac{\partial R(s)}{\partial s}H(x) + h^{(1)}(s)H'(x), \tag{8.44}$$

где $h^{(1)}(s)$ определяется системой (8.43).

Этап 4. В уравнении (8.30) разложим функции в ряд по приращениям аргумента x с точностью до $o(\varepsilon^2)$, получим

$$K(\rho)\pi(x,\varepsilon)-\varepsilon V(\rho)\frac{\partial\pi(x,\varepsilon)}{\partial x}+\frac{\varepsilon^2}{2}D(s)\frac{\partial^2\pi(x,\varepsilon)}{\partial x^2}=o(\varepsilon^2), \qquad (8.45)$$

где матрица $D(x)$ имеет вид

$$D(x)=B_1(s)+4B_2(s)+A_1. \qquad (8.46)$$

Сложив все уравнения системы (8.45) и учитывая свойство (8.32) матрицы $K(\rho)$, получим

$$-\varepsilon E^T V(\rho)\frac{\partial\pi(x,\varepsilon)}{\partial x}+\frac{\varepsilon^2}{2}E^T D(s)\frac{\partial^2\pi(x,\varepsilon)}{\partial x^2}=o(\varepsilon^2).$$

В это равенство подставим $\pi(x,\varepsilon)$ в виде (8.41), а матрицу $V(\rho)$ в виде

$$V(\rho)=V(s)-\varepsilon\frac{\partial V(s)}{\partial s}.$$

В результате будем иметь

$$-\varepsilon E^T\left(V(s)-\varepsilon\frac{\partial V(s)}{\partial s}\right)\frac{\partial}{\partial x}\big(R(s)H(x)+\varepsilon h(x)\big)+\frac{\varepsilon^2}{2}E^T D(s)R(s)H''(x)=o(\varepsilon^2).$$

Выполнив в этом равенстве несложные преобразования, перепишем его в виде

$$-\varepsilon E^T\left\{V(s)R(s)H'(x)-\varepsilon\frac{\partial V(s)}{\partial s}R(s)H'(x)+\varepsilon V(s)h'(x)\right\}+$$

$$+\frac{\varepsilon^2}{2}E^T D(s)R(s)H''(x)=o(\varepsilon^2). \qquad (8.47)$$

Учитывая равенство (8.40), перепишем (8.47) следующим образом:

$$E^T\frac{\partial V(s)}{\partial s}R(s)H'(x)-E^T V(s)h'(x)+\frac{1}{2}E^T D(s)R(s)H''(x)=0.$$

Подставляя сюда (8.44), получим равенство

$$E^T\frac{\partial V(s)}{\partial s}R(s)H'(x)+E^T V(s)\frac{\partial R(s)}{\partial s}H'(x)-E^T V(s)h^{(1)}(s)H''(x)+$$

$$+\frac{1}{2}E^T D(s)R(s)H''(x)=0,$$

которое перепишем в виде

$$\frac{\partial}{\partial s}\left\{E^T V(s)R(s)\right\}H'(x)+E^T\left\{\frac{1}{2}D(s)R(s)-V(s)h^{(1)}(s)\right\}H''(x)=0\ ,$$

то есть в виде линейного однородного обыкновенного дифференциального уравнения второго порядка с постоянными коэффициентами относительно неизвестной функции $H(x)$. Его решение, которое удовлетворяет условию нормировки, имеет вид

$$H(x)=\nu e^{-\nu x}\ , \tag{8.48}$$

где

$$\nu=\frac{\partial\left\{E^T V(s)R(s)\right\}}{\partial s}\Bigg/E^T\left\{\frac{1}{2}D(s)R(s)-V(s)h^{(1)}(s)\right\}. \tag{8.49}$$

Здесь $h^{(1)}(s)$ является решением системы (8.43).

§ 8.4. Исследование сетей случайного доступа с адаптивными протоколами

Рассмотрим компьютерную сеть связи, управляемую адаптивным протоколом случайного множественного доступа с оповещением о конфликте. Структура математической модели такой сети изображена на рис. 8.7.

Здесь T – состояние адаптера в текущий момент времени, которое меняется следующим образом:

$$T(t+\Delta t)=\begin{cases}T(t)+\alpha\Delta t, & \text{если } k(t)=0,\\ T(t), & \text{если } k(t)=1,\\ T(t)+\beta\Delta t, & \text{если } k(t)=2,\end{cases}$$

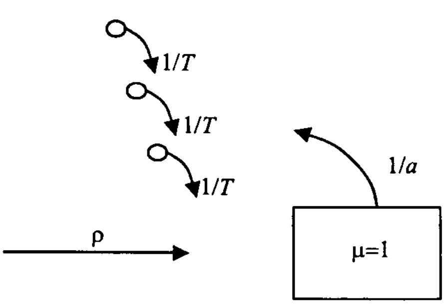


Рис. 8.7

где k – состояние прибора; $\alpha>0$, $\beta>0$ – параметры, характеризующие структуру адаптера.

Будем полагать, что для каждой заявки в ИПВ вероятность окончания случайной задержки за бесконечно малый промежуток времени длительности Δt равна $\Delta t/T+o(\Delta t)$, где T – текущее состояние адаптера.

Трехмерный случайный процесс $\{k(t), i(t), T(t)\}$ является марковским, поэтому распределение вероятностей

$$P_k(i,T,t) = P\{k(t)=k, i(t)=i, T \le T(t) < T+dT\}/dT$$

удовлетворяет следующим равенствам:

$$P_0(i,T-\alpha\Delta t, t+\Delta t) = \left[1-\left(\rho+\frac{i}{T}\right)\Delta t\right]P_0(i,T,t) + \Delta t P_1(i,T,t) + $$
$$+\frac{1}{a}\Delta t P_2(i,T,t) + o(\Delta t),$$

$$P_1(i,T,t+\Delta t) = \left[1-\left(\rho+\frac{i}{T}+1\right)\Delta t\right]P_1(i,T,t) + \rho\Delta t P_0(i,T,t) +$$
$$+\frac{i+1}{T}\Delta t P_0(i+1,T,t) + o(\Delta t),$$

$$P_2(i,T+\beta\Delta t, t+\Delta t) = \left[1-\left(\rho+\frac{1}{a}\right)\Delta t\right]P_2(i,T,t) + \rho\Delta t P_2(i-1,T,t) +$$
$$+\frac{i-1}{T}\Delta t P_1(i-1,T,t) + \rho\Delta t P_1(i-2,T,t) + o(\Delta t).$$

Отсюда, выполнив несложные преобразования, получим

$$\frac{\partial P_0(i,T,t)}{\partial t} = \alpha\frac{\partial P_0(i,T,t)}{\partial T} - \left(\rho+\frac{i}{T}\right)P_0(i,T,t) + P_1(i,T,t) + \frac{1}{a}P_2(i,T,t),$$

$$\frac{\partial P_1(i,T,t)}{\partial t} = -\left(\rho+\frac{i}{T}+1\right)P_1(i,T,t) + \rho P_0(i,T,t) + \frac{i+1}{T}P_0(i+1,T,t),$$

$$\frac{\partial P_2(i,T,t)}{\partial t} = -\beta\frac{\partial P_2(i,T,t)}{\partial T} - \left(\rho+\frac{1}{a}\right)P_2(i,T,t) + \rho P_2(i-1,T,t) +$$
$$+\frac{i-1}{T}P_1(i-1,T,t) + \rho P_1(i-2,T,t). \qquad (8.50)$$

Обозначив вектор

$$P(i,T,t) = \{P_0(i,T,t), P_1(i,T,t), P_2(i,T,t)\}^T,$$

систему (8.50) перепишем в виде уравнения

$$\frac{\partial P(i,T,t)}{\partial t}+A\frac{\partial P(i,T,t)}{\partial T}=A_0\left(\rho,\frac{i}{T}\right)P(i,T,t)+A_1\left(\frac{i+1}{T}\right)P(i+1,T,t)+$$

$$+B_1\left(\rho,\frac{i-1}{T}\right)P(i-1,T,t)+B_2(\rho)P(i-2,T,t), \qquad (8.51)$$

где матрицы A, A_0, A_1, B_1, B_2 однозначно определяются системой (8.51).

Уравнение (8.51) будем решать методом асимптотического анализа в условиях большой загрузки, полагая, что

$$\rho=s-\varepsilon,$$

где s – величина пропускной способности (ее значение определим ниже), а ε – малый положительный параметр, формирующий асимптотическое условие $\varepsilon\to 0$.

В уравнении (8.51), выполнив замены

$$t\varepsilon=\tau, \quad i\varepsilon=x, \quad T\varepsilon=y, \quad \frac{1}{\varepsilon^2}P(i,T,t)=\pi(x,y,\tau,\varepsilon),$$

получим

$$\varepsilon\frac{\partial\pi(x,y,\tau,\varepsilon)}{\partial\tau}+\varepsilon A\frac{\partial\pi(x,y,\tau,\varepsilon)}{\partial y}=A_0\left(\rho,\frac{x}{y}\right)\pi(x,y,\tau,\varepsilon)+$$

$$+A_1\left(\frac{x+\varepsilon}{y}\right)\pi(x+\varepsilon,y,\tau,\varepsilon)+B_1\left(\rho,\frac{x-\varepsilon}{y}\right)\pi(x-\varepsilon,y,\tau,\varepsilon)+ \qquad (8.52)$$

$$+B_2(\rho)\pi(x-2\varepsilon,y,\tau,\varepsilon).$$

Это уравнение будем решать в два этапа.

Этап 1. Положив в (8.52) $\varepsilon\to 0$, $\rho\to s$ и обозначив $\pi(x,y,\tau,0)=$ $=\pi(x,y,\tau)$, уравнение (8.52) запишем в виде

$$K\left(s,\frac{x}{y}\right)\pi(x,y,\tau)=0, \qquad (8.53)$$

где

$$K\left(s,\frac{x}{y}\right)=A_0\left(s,\frac{x}{y}\right)+A_1\left(\frac{x}{y}\right)+B_1\left(s,\frac{x}{y}\right)+B_2(s). \qquad (8.54)$$

Очевидно, это уравнение и его решение обладает всеми свойствами уравнения (8.31) из предыдущего раздела, поэтому

$$\pi(x,y,\tau)=R\left(s,\frac{x}{y}\right)H(x,y,\tau), \tag{8.55}$$

где $H(x,y,\tau)$ – скалярная функция, а вектор $R\left(s,\frac{x}{y}\right)$ определяется системой

$$K\left(s,\frac{x}{y}\right)R\left(s,\frac{x}{y}\right)=0 \tag{8.56}$$

и условием нормировки

$$E^T R\left(s,\frac{x}{y}\right)=1. \tag{8.57}$$

Этап 2. В уравнении (8.52) все функции разложим в ряд по приращению аргумента x с точностью до $o(\varepsilon)$. Тогда получаем

$$\varepsilon\frac{\partial\pi(x,y,\tau,\varepsilon)}{\partial\tau}+\varepsilon A\frac{\partial\pi(x,y,\tau,\varepsilon)}{\partial y}=K\left(\rho,\frac{x}{y}\right)\pi(x,y,\tau,\varepsilon)-$$

$$-\varepsilon\frac{\partial}{\partial x}\left\{V\left(s,\frac{x}{y}\right)\pi(x,y,\tau,\varepsilon)\right\}+o(\varepsilon), \tag{8.58}$$

где

$$V\left(s,\frac{x}{y}\right)=B_1\left(s,\frac{x}{y}\right)+2B_2(s)-A_1\left(\frac{x}{y}\right).$$

Учитывая свойство

$$E^T K\left(\rho,\frac{x}{y}\right)=0$$

матрицы K, из уравнения (8.58) получим

$$E^T\frac{\partial\pi(x,y,\tau)}{\partial\tau}=-\frac{\partial}{\partial x}\left\{E^T V\left(s,\frac{x}{y}\right)\pi(x,y,\tau)\right\}-\frac{\partial}{\partial y}\left\{E^T A\ \ \pi(x,y,\tau)\right\}.$$

Подставляя сюда выражение (8.55) для вектора $\pi(x,y,\tau)$, получим уравнение относительно скалярной функции $H(x,y,\tau)$ следующего вида:

$$\frac{\partial H(x,y,\tau)}{\partial \tau} = -\frac{\partial}{\partial x}\left\{\left[E^T V\left(s,\frac{x}{y}\right)R\left(s,\frac{x}{y}\right)\right]H(x,y,\tau)\right\} -$$

$$-\frac{\partial}{\partial y}\left\{\left[E^T A \quad R\left(s,\frac{x}{y}\right)\right]H(x,y,\tau)\right\},$$

которое является вырожденным уравнением Фоккера – Планка для плотности $H(x,y,\tau)$ распределения вероятностей значений некоторого двумерного диффузионного процесса $\{x(\tau),y(\tau)\}$, коэффициенты переноса которого имеют вид $E^T V\left(s,\frac{x}{y}\right)R\left(s,\frac{x}{y}\right)$; $E^T AR\left(s,\frac{x}{y}\right)$. Коэффициенты диффузии в этом случае равны нулю, поэтому $\{x(\tau),y(\tau)\}$ являются детерминированными функциями, вид которых определяется системой обыкновенных дифференциальных уравнений

$$\begin{cases} x'(\tau) = E^T V\left(s,\dfrac{x}{y}\right)R\left(s,\dfrac{x}{y}\right), \\ y'(\tau) = E^T A \quad R\left(s,\dfrac{x}{y}\right). \end{cases} \tag{8.59}$$

Если эта система дифференциальных уравнений имеет точку покоя, ее координаты (x,y) удовлетворяют следующей системе уравнений:

$$\begin{cases} E^T V\left(s,\dfrac{x}{y}\right)R\left(s,\dfrac{x}{y}\right) = 0, \\ E^T A \quad R\left(s,\dfrac{x}{y}\right) = 0. \end{cases} \tag{8.60}$$

Очевидно, что полученная система (8.60) определяет величину s пропускной способности сети и постоянное значение $\gamma = \frac{x}{y}$ отношения $\frac{x}{y}$.

Следовательно, множество точек покоя образуют прямую, определяемую уравнением

$$x = \gamma y\ ,$$

а математическая модель сети с адаптивным протоколом в условиях большой загрузки эквивалентна модели сети с динамическим протоколом, значение параметра γ которого определяется системой (8.60).

Дальнейшее исследование сети с адаптивным протоколом реализуется методами, изложенными в предыдущем разделе для сетей с динамическими протоколами и найденным значением параметра γ .

Литература

1. *Гельфонд А.О.* Исчисление конечных разностей. – М.: ОНТИ, 1936.
2. *Скитович В.П.* Элементы теории массового обслуживания. – Л.: Изд-во Лен. ун-та, 1976.
3. *Климов Г.П.* Стохастические системы массового обслуживания. – М.: Наука, 1966.
4. *Саати Т.Л.* Элементы теории массового обслуживания и ее приложения. – М.: Сов. радио, 1971.
5. *Кофман А., Крюон Р.* Массовое обслуживание. Теория и приложения. – М.: Мир, 1965.
6. *Гнеденко Б.В., Коваленко И.Н.* Введение в теорию массового обслуживания. – М.: Наука, 1966.
7. *Хинчин А.Я.* Работы по математической теории массового обслуживания. – М.: Наука, 1963.
8. *Ивченко Г.И., Каштанов В.А., Коваленко И.Н.* Теория массового обслуживания. – М.: Высшая школа, 1982.
9. *Кениг Д., Рыков В., Штойян Д.* Теория массового обслуживания (основной курс: марковские модели, методы марковизации): Учебное пособие по математике для студентов специальности 0647 – «Прикладная математика». – М., 1979.
10. *Кениг Д., Штойян Д.* Методы теории массового обслуживания. – М.: Радио и связь, 1981.
11. *Буримов А.Д., Малинковский Ю.В., Матальщкий М.А.* Теория массового обслуживания: Учебное пособие по спецкурсу. – Гродно, 1984.
12. *Бронштейн О.И., Духовный И.М.* Модели приоритетного обслуживания в информационно-вычислительных сетях. – М.: Наука, 1976.
13. *Мова В.В., Пономаренко Л.А., Калиновский А.Н.* Организация приоритетного обслуживания в АСУ. – Киев: Техніка, 1977.
14. *Джейсуол Н.* Очереди с приоритетами. – М.: Мир, 1973.
15. *Назаров А.А.* Управляемые системы массового обслуживания и их оптимизация. – Томск: Изд-во Том. ун-та, 1984.
16. *Назаров А.А.* Асимптотический анализ марковизируемых систем. – Томск: Изд-во Том. ун-та, 1991.
17. *Кузнецов Д.Ю., Назаров А.А.* Адаптивные сети случайного доступа. – Томск: Дельтаплан, 2002.

Для заметок

Анатолий Андреевич Назаров
Александр Федорович Терпугов

ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ

Учебное пособие

Редактор *Н.И. Шидловская*
Верстка *Л.В. Пермякова*

Изд. лиц. ИД № 04000 от 12.02.01. Подписано к печати 22.09.2010.
Формат 60×84$^1/_{16}$. Бумага офсетная. Печать офсетная. Гарнитура «Times».
Усл. печ. л. 13,25. Уч.-изд. л. 14,84. Тираж 500 экз.

ООО «Издательство научно-технической литературы»
634050, Томск, пл. Ново-Соборная, 1, тел. (382-2) 53-33-35

Отпечатано в типографии ЗАО «М-Принт», г. Томск, ул. Пролетарская, 38/1