Министерство образования Республики Беларусь
Учреждение образования «Витебский государственный
университет имени П.М. Машерова»
Кафедра прикладной математики и механики

Л.В. Командина

ИССЛЕДОВАНИЕ ОПЕРАЦИЙ. ЭЛЕМЕНТЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

Методические рекомендации

Витебск
ВГУ имени П.М. Машерова
2015

УДК 519.8(075.8)
ББК 22.183я73
К63

Печатается по решению научно-методического совета учреждения образования «Витебский государственный университет имени П.М. Машерова». Протокол № 1 от 23.10.2015 г.

Автор: доцент кафедры прикладной математики и механики ВГУ имени П.М. Машерова, кандидат физико-математических наук **Л.В. Командина**

Р е ц е н з е н т:
доцент кафедры геометрии и математического анализа
ВГУ имени П.М. Машерова, кандидат физико-математических наук,
доцент *С.А. Шлапаков*

Командина, Л.В.

К63 Исследование операций. Элементы теории массового обслуживания : методические рекомендации / Л.В. Командина. – Витебск : ВГУ имени П.М. Машерова, 2015. – 46 с.

В методических рекомендациях приведены основные понятия теории массового обслуживания: система массового обслуживания, входной поток, поток обслуживаний, канал обслуживания, марковский процесс, процессы гибели и размножения. Рассмотрены следующие виды систем массового обслуживания: с отказами, многоканальная с конечной очередью, одноканальная с бесконечной очередью, многоканальная с бесконечной очередью.

Предназначается для студентов специальностей «Прикладная математика», «Прикладная информатика», «Программное обеспечение информационных технологий», «Компьютерная безопасность» (дисциплина «Исследование операций»).

УДК 519.8(075.8)
ББК 22.183я73

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Предлагаемые методические рекомендации содержат материал раздела «Элементы теории массового обслуживания» курса «Исследование операций», который соответствует образовательному стандарту по специальностям «Прикладная математика», «Прикладная информатика», «Программное обеспечение информационных технологий», «Компьютерная безопасность». При изложении теоретического материала предполагается, что студенты владеют основными понятиями теории вероятности: случайная величина, случайный процесс, марковский процесс, математическое ожидание, дисперсия, закон распределения, плотность распределения и т.д. Рассматриваются основные простейшие модели систем массового обслуживания. Приводятся типовые примеры.

§ 1. ЗАДАЧИ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

Теория массового обслуживания или теория очередей – прикладная математическая дисциплина, истоки которой в начале двадцатого столетия, когда датский инженер А. Эрланг исследовал вопросы организации работы телефонной сети. Нужны были методы, которые позволили бы просчитать эффективность обслуживания потребителей в зависимости от применяемых устройств. Основные же результаты этой теории относятся ко второй половине столетия. Объект исследования – некая вероятностная модель реального процесса обслуживания, в котором с одной стороны в случайные моменты времени возникает необходимость в выполнении некоторых работ, а с другой стороны идет процесс удовлетворения этой необходимости, вообще говоря, опять-таки в случайные моменты времени. В теории массового обслуживания их принято называть **системами массового обслуживания** (СМО). Примерами таких систем могут служить: телефонные станции, ремонтные мастерские, билетные кассы, совокупность станков, нуждающихся в ремонте, справочные бюро, магазины и т.п.

Каждая система массового обслуживания состоит из какого-то числа обслуживающих единиц, которые мы будем называть **каналами** обслуживания. В качестве каналов могут фигурировать линии связи, рабочие точки, приборы, железнодорожные пути, лифты, автомашины и т.д. По количеству каналов системы массового обслуживания делятся на *одноканальные* и *многоканальные*.

Каждая СМО предназначена для обслуживания какого-то **потока заявок** (или требований), поступающих в СМО. Обслуживание поступившей заявки продолжается некоторое время, после чего канал освобождается и готов к обслуживанию следующей заявки. Случайный характер, как потока заявок, так и времени обслуживания приводит к тому, что в какие-то промежутки времени на входе СМО скапливается излишне большое число заявок: они либо образуют очередь, либо покидают СМО необслуженными; в другие же периоды СМО будет работать с недогрузкой или вообще простаивать. В связи с этим различают *системы с отказами* (или потерями) и *системы с очередью*. В случае очередей важен порядок обслуживания, например, различают такие: первым пришел – первым обслужился, последним пришел – первым обслужился, или очередь с приоритетами.

По характеру, возникновения заявок в системе различают *замкнутые* и *разомкнутые* СМО. В первом случае источник заявок находится в самой системе, а во втором – вне системы. Примером замкнутой системы является станок, состоящий из нескольких узлов, которые время от времени выходят из строя, после чего их требуется отремонтировать. Пример разомкнутой системы – автозаправочная станция.

Каждая СМО, в зависимости от числа каналов и их производительности, а также от характера потока заявок, обладает некоторой **пропускной способностью**, позволяющей ей более или менее успешно справляться с потоком заявок.

Предмет теории массового обслуживания – установление зависимости между характером потока заявок, числом каналов, их производительностью, правилами работы СМО с одной стороны и успешностью (эффективностью) обслуживания с другой стороны.

В качестве характеристик эффективности обслуживания, в зависимости от условий задачи и целей исследования, могут применяться различные величины и функции, например:

- среднее количество заявок, которое может обслужить СМО в единицу времени;
- средний процент заявок, получающих отказ и покидающих СМО необслуженными;
- вероятность того, что поступившая заявка немедленно будет принята к обслуживанию;
- среднее время ожидания в очереди;
- среднее количество заявок, находящихся в очереди;
- средний доход, приносимый СМО в единицу времени и т.д.

Чтобы дать рекомендации по рациональной организации работы системы и предъявить разумные требования к ней, необходимо изучить случайный процесс, протекающий в системе, описать его математически. Этим и занимается теория массового обслуживания.

Заметим, что за последние годы область применения математических методов теории массового обслуживания непрерывно расширяется и все больше выходит за пределы задач, связанных с «обслуживающими организациями» в буквальном смысле слова. Как своеобразные СМО могут рассматриваться электронные цифровые вычислительные машины, системы сбора и обработки информации, автоматизированные производственные цехи, поточные линии, транспортные системы, системы противовоздушной обороны и т.п. Близкими к задачам теории массового обслуживания являются многие задачи, возникающие при анализе надежности технических устройств.

Математический анализ работы СМО очень облегчается, если случайный процесс, протекающий в системе, является ***марковским***. Тогда удается сравнительно просто описать работу СМО с помощью аппарата обыкновенных дифференциальных уравнений, а в предельном случае – линейных алгебраических уравнений, и выразить в явном виде основные характеристики эффективности обслуживания через параметры СМО и потока заявок.

Если случайный процесс, протекающий в системе, не является марковским, то его математическое описание становится более сложным и требует более громоздкого аппарата, получение явных аналитических формул удается редко. Однако, аппарат «марковской» теории массового обслуживания может пригодиться и в этом случае – с его помощью характеристики эффективности СМО могут быть оценены приближенно. Следует заметить, что чем сложнее СМО, чем больше в ней каналов обслуживания, тем точнее оказываются приближенные формулы, полученные с помощью марковской теории. Следует также заметить, что в ряде случаев для принятия обоснованных решений по управлению СМО вовсе и не требуется точного знания всех ее характеристик – зачастую достаточно и приближенного, ориентировочного.

§ 2. ПОТОК. ВИДЫ ПОТОКОВ

Потоком заявок называется последовательность однородных событий, следующих один за другим в какие-то случайные моменты времени.

Примеры: поток вызовов на телефонной станции, поток железнодорожных составов на сортировочной станции, поток частиц, попадающих на счетчик Гейгера, поток поломок узлов станка, поток покупателей в магазине и т.п.

Важной характеристикой потока является его интенсивность – среднее число заявок в единицу времени. Обозначается она обычно греческими λ или μ. Интенсивность может быть постоянной, а может быть функцией времени. Например, поток машин через конкретный перекресток в течение суток имеет переменную интенсивность, зависящую от времени суток, а интенсивность потока машин там же с 10-00 до 11-00 можно считать постоянной.

Рассмотрим некоторые виды потоков с точки зрения моментов поступления в систему.

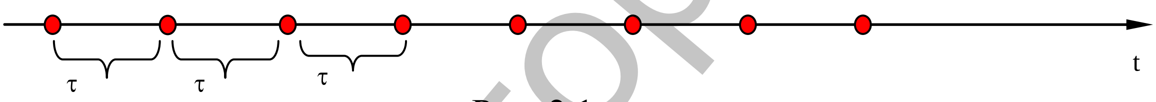


Рис. 2.1.

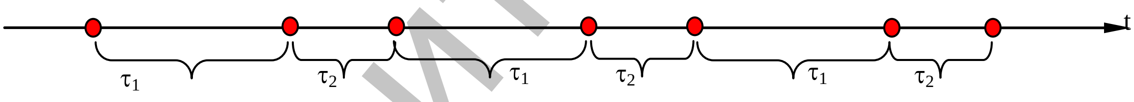


Рис. 2.2.

Поток заявок называется ***регулярным***, если заявки поступают в систему через одно и то же время τ (рис. 2.1) или когда существует периодичность в поступлении заявок (рис. 2.2). На практике они встречаются редко.
характеристики не зависят от времени. Прежде всего, речь идет о вероятности поступления одного и того числа заявок в единицу времени, таким образом, *интенсивность стационарного потока заявок есть величина постоянная*. Фактически на один участок длиной в Δt единиц времени может попасть больше заявок, а на другой участок такой же длины – меньше, но среднее число, приходящее на единицу времени постоянно и не зависит от времени, т.е. не зависит от того, где находится этот промежуток Δt на оси времени.

Поток заявок называется потоком ***без последействия***, если для любых двух непересекающихся интервалов τ_1 и τ_2 времени количество заявок, поступающих на один из них, не зависит от того, сколько заявок поступает на другой участок (см. рис. 2.3). Другими словами, заявки

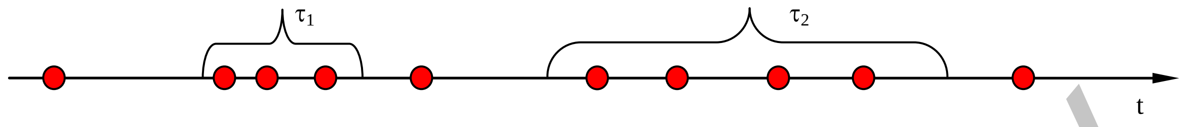


Рис. 2.3.

появляются в те или иные моменты времени независимо друг от друга, каждая по своей причине. Пример: поток пассажиров, входящих в метро, является потоком без последействия. Поток покупателей, рассчитавшихся за покупку, не является потоком без последействия, т.к. интервал между отдельными покупателями не может быть меньше минимального времени обслуживания каждого покупателя.

Поток заявок называется ***ординарным***, если вероятность поступления более одной заявки за малый промежуток времени τ есть величина более высокого порядка малости, чем промежуток τ. Это означает, что заявки в систему поступают поодиночке. А значит, вероятностью попадания двух и более заявок на малый участок τ можно пренебречь. Пример: поток больных к зубному врачу, поток составов на подходе к станции являются обычно ординарными; но поток вагонов на подходе к станции, поток посетителей музея чаще всего не являются ординарными.

Поток заявок называется ***простейшим***, если он является стационарным, ординарным и не имеет последействия. Название простейший связано с тем, что потоки такого вида имеют наиболее простое математическое описание. Кстати, регулярный поток не является простейшим, так как обладает последействием.

Доказано (А.Я. Хинчин), что при наложении (суперпозиции) достаточно большого числа стационарных и ординарных потоков, сравнимых между собой по интенсивности, получается суммарный поток, близкий к простейшему.

Поток заявок описывается в виде некоторого закона распределения, который показывает распределение вероятностей для длительностей промежутков времени между последовательными моментами поступления заявок.

Доказано, что если поток является простейшим, то указанный закон является законом Пуассона. Рассмотрим на оси времени некоторый участок длины Δt , начинающийся в момент t_0 . В курсе теории вероятностей доказывается, что вероятность попадания на этот участок ровно m заявок выражается формулой

$$P_m = \frac{a^m}{m!} e^{-a}, \quad m = 0,\ 1,\ 2,\ \ldots \tag{2.1}$$

где a – среднее число заявок, приходящееся на участок Δt . Для стационарного потока заявок, а значит, и для простейшего, значение a равно

$$a = \lambda \Delta t.$$

Найдем закон распределения интервалов времени T между поступлением двух соседних заявок

$$F(t) = P(T<t),$$

т.е. вероятность того, что величина T примет значение меньшее, чем t . Отложим от начала интервала – точка t_0, отрезок t, и найдем вероятность

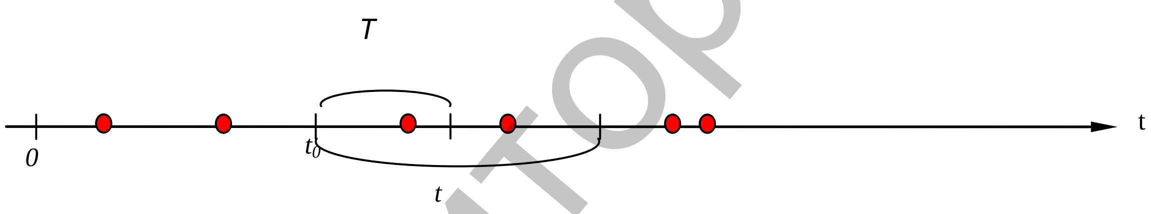


Рис. 2.4.

того, что интервал T будет меньше t . Для этого нужно, чтобы на участок, примыкающий к t_0 , попала хотя бы одна заявка (рис.2.4). Вероятность этого события вычислим через вероятность противоположного события – на участок t не попадает ни одной заявки, которую найдем по формуле (2.1), положив $m = 0$, и учитывая, что $a = \lambda t$:

$$P_0 = \frac{a^0}{0!} e^{-a} = e^{-\lambda t}.$$

Тогда

$$F(t) = 1 - e^{-\lambda t}, \quad t > 0. \tag{2.2}$$

Чтобы найти плотность распределения, продифференцируем (2.2) и получим:

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0. \tag{2.3}$$

Таким образом, имеем показательный закон распределения с параметром λ. Отметим числовые характеристики:

1. математическое ожидание

$$m_T = \int_0^{\infty} t \cdot f(t)\, dt = \lambda \int_0^{\infty} t \cdot e^{-\lambda t} dt = (\text{по частям}) = \frac{1}{\lambda}.$$

2. дисперсию

$$D_T = \int_0^{\infty} t^2 \cdot f(t)\, dt - m_t^2 = \lambda \int_0^{\infty} t^2 \cdot e^{-\lambda t} dt - \frac{1}{\lambda^2} = (\text{по частям}) = \frac{1}{\lambda^2}.$$

3. среднее квадратичное отклонение

$$\sigma_T = \sqrt{D_t} = \frac{1}{\lambda}.$$

Как оказалось $m_T = \sigma_T$.

В теории вероятностей в качестве «меры случайности» неотрицательной случайной величины рассматривают так называемый коэффициент вариации:

$$v_T = \sigma_T / m_T\ .$$

Для простейшего потока заявок коэффициент вариации интервалов между заявками равен единице. Для регулярного потока заявок, у которого интервал между заявками не случаен, т.е. $\sigma_T = 0$, а значит, и коэффициент вариации равен нулю. Для большинства потоков встречающихся на практике коэффициент вариации принадлежит интервалу (0, 1) и служит как бы мерой регулярности потока: чем ближе значение коэффициент вариации к нулю, тем «регулярнее» поток. Простейший поток – наименее регулярный.

Поток заявок называется ***рекуррентным*** (или *«потоком Пальма»*), если он является стационарным, ординарным, а интервалы времени между заявками есть независимые случайные величины с одинаковым *произвольным* распределением. Ясно, что простейший поток – частный случай рекуррентного потока.

В теории массового обслуживания рассматривают еще так называемые потоки Эрланга, которые получаются «прореживанием» простейшего потока. Если в простейшем потоке сохраняется каждая k-ая заявка, а остальные не учитываются, то имеем поток Эрланга k-го порядка. Простейший поток есть поток Эрланга 1-го порядка. Доказано, что коэффициент вариации для потока Эрланга k-го порядка равен $v_T^n = \frac{1}{\sqrt{n}}$., т.е. при увеличении числа k поток Эрланга приближается к регулярному потоку.

Замечание 1. Выше разговор о потоке заявок без уточнения его места в системе массового обслуживания. На самом деле должно различать входной поток, поступающий в систему (его интенсивность обычно обозначают через λ) и выходящий поток после обслуживания. Сам выходящий поток нас мало интересует, если только он не является в свою очередь входным потоком другой системы. Но его характеристики определяются механизмом обслуживания в системе. Основной характеристикой функционирования канала является время обслуживания заявки, этот показатель характеризует пропускную способность канала. Время обслуживания является случайной величиной. Считают, что продолжительность обслуживания разных заявок некоторым каналом есть независимые случайные величины, имеющие один закон распределения. Наиболее часто полагают этот закон распределения показательным.

Обозначим через μ интенсивность обслуживания заявок (среднее число заявок, обслуживаемых в единицу времени). Если μ не зависит от времени, то имеем:

функцию распределения $F(t) = P(T<t)= 1 - e^{-\mu t}$

плотность распределения $g(t) = \mu e^{-\mu t}$

Основное свойство показательного закона распределения времени обслуживания состоит в том, что вероятность завершения обслуживания заявки в любой последующий момент не зависит от времени, которое уже было затрачено на обслуживание.

Замечание 2. В литературе различные системы массового обслуживания принято обозначать с помощью так называемых символов Кенделя. Символ включает три позиции. Первая характеризует входной поток, вторая – механизм обслуживания, третья определяет число каналов. Например, буква M обозначает показательный закон распределения, D – регулярный (детерминированный) характер распределения, G – произвольный характер распределения. Пример: M|D|3.

§ 3. ПОНЯТИЕ МАРКОВСКОГО ПРОЦЕССА

Пусть имеется некоторая физическая система S , которая с течением времени меняет свое состояние: переходит из одного состояния в другое, причем заранее неизвестно каким образом, другими словами, случайно. В этом случае говорят, что в системе S протекает случайный процесс. Например, пусть S – техническое устройство, состоящее из нескольких узлов, каждый из которых время от времени выходит из строя и требует или замены или ремонта. Понятно, что процесс, протекающий в этой системе, является случайным.

Случайный процесс, протекающий в системе, ***называется марковским***, если для любого момента времени t_0 вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент t_0 и не зависят от того, когда и как система пришла в это состояние.

Другими словами, пусть система S в момент t_0 находится в состоянии S_0 , и до момента t_0 известно все о состояниях системы. Если в системе происходит марковский процесс, то при $t>t_0$ о вероятностных характеристиках процесса можно говорить только на основании состояния S_0 , забыв о том, что было до момента t_0 . Образно говоря, будущее зависит от прошлого через настоящее.

Пример 1. Пусть S – счетчик Гейгера, регистрирующий частицы из космоса. Состояние характеризуется количеством этих частиц в данный момент. Вероятность того, что в момент $t>t_0$ счетчик покажет то или иное число частиц зависит от состояния в момент t_0 и не зависит от того, в какие именно моменты времени зарегистрированные частицы приходили.

На практике часто встречаются случайные процессы, которые не являются марковскими, но в каком-то приближении могут быть рассмотрены как марковские. Вообще говоря, любой случайный процесс можно рассматривать как марковский, если прошлое «зашить» в настоящее через какие-то параметры. Но здесь есть опасность: прошлое может быть богатым и число параметров велико, и тогда приходим к «проклятию размерности».

Мы в дальнейшем иметь дело с марковскими процессами с *дискретными состояниями* и *непрерывным временем*. Процесс называется процессом с дискретными состояниями, если его состояния S_1, S_2, S_3, … можно заранее перечислить и перенумеровать, а переход системы из состояния в состояние происходит практически мгновенно, «скачком». Процесс называется процессом с непрерывным временем, если моменты возможных переходов из состояния в состояние не фиксированы заранее, случайны, этот переход может произойти в любой момент времени.

Пример 1. Пусть S – техническое устройство с двумя узлами, каждый из которых в случайный момент времени может отказать, после чего мгновенно начинается его ремонт. Ремонт может длиться не фиксированный промежуток времени, а, вообще говоря, случайный. Перечислим возможные состояния системы:

S_0 – оба узла исправны,

S_1 – первый узел ремонтируется, второй узел исправен,

S_2, – второй узел ремонтируется, первый узел исправен,

S_3 – оба узла ремонтируются.

Переходы из состояния в состояние происходят мгновенно в случайные моменты времени. При анализе случайных процессов используют геометрическую интерпретацию, так называемый *граф состояний*. Вершины графа – это состояния, дуги графа – возможные переходы из состояния в состояние. Построим граф для рассмотренного выше примера (см. рис. 3.1).

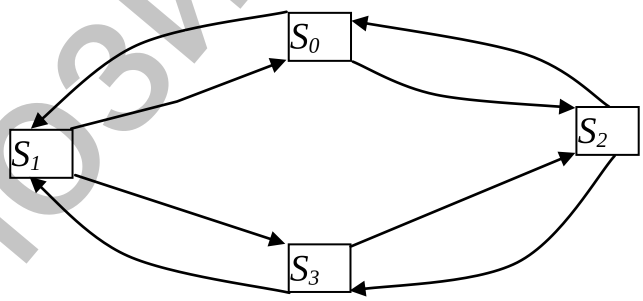


Рис. 3.1.

Дуга из S_2 в S_0 означает момент окончания ремонта второго узла, когда первый был исправен. Дуга из S_0 в S_1 означает момент выхода из строя первого узла, второй при этом исправен. Нет прямой дуги из S_0 в S_3 , хотя в жизни это реально (короткое замыкание и оба узла вышли из строя). Предполагается, что узлы выходят из строя независимо друг от друга, вероятностью строго одновременного отказа в работе обоих узлов пренебрегают, как очень малой величиной.

Для описания марковского процесса с дискретными состояниями и непрерывным временем строится достаточно простая математическая модель.

§ 4. УРАВНЕНИЯ КОЛМОГОРОВА. ПРЕДЕЛЬНЫЕ ВЕРОЯТНОСТИ

Имеем систему S, все переходы которой из состояния в состояние происходят под действием некоторого потока заявок. Если поток заявок является простейшим, то процесс, происходящий в системе, будет марковским. Это и естественно, так как простейший поток не обладает последействием: в нем «будущее» не зависит от «прошлого».

Пусть система S находится в каком-то состоянии S_i, из которого есть непосредственный переход в состояние S_j (на графе состояний имеется дуга из S_i в S_j), на систему, пока она находится в состоянии S_i, действует простейший поток заявок, как только появится первая заявка, происходит «перескок» системы из S_i в S_j . Для наглядности удобно на графе состояний у каждой дуги проставлять интенсивность того потока заявок, который переводит систему по данной стрелке. Обозначим через λ_{ij} интенсивность потока, переводящего из состояния S_i в состояние S_j . Такой граф называется размеченным.

Разметим граф в примере 1 предыдущего параграфа, в предположении, что ремонтом каждого узла занят отдельный специалист. Обозначим через λ_1 интенсивность, с которой выходит из строя первый узел. Этот параметр

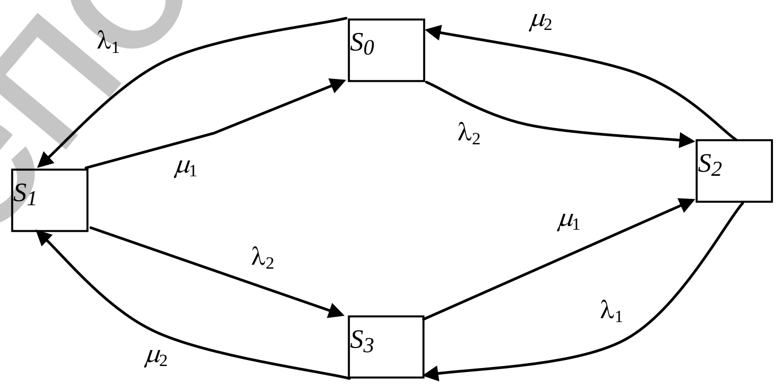


Рис. 4.1.

соответствует переходам из S_0 в S_1 и из S_2 в S_3. Интенсивность λ_2, с которой выходит из строя второй узел, соответствует переходам из S_0 в S_1 и из S_2 в S_3. Какой поток переводит систему состояния из S_1 в S_0? Очевидно поток окончаний ремонтов первого узла. Его интенсивность μ_1 есть величина обратная среднему времени ремонта первого узла. Аналогично вычисляется интенсивность μ_2. Размеченный граф представлен на рис. 4.1

Имея в своем распоряжении размеченный граф состояний системы, легко построить математическую модель данного процесса.

Пусть рассматривается система S, имеющая n возможных состояний $S_1, S_2, \ldots, S_n$. Вероятность $p_i(t)$ того, что в момент t система будет находиться в состоянии S_i, назовем вероятностью i-го состояния системы. Очевидно, что для любого момента сумма всех вероятностей состояний равна единице:

$$\sum_{i=1}^{n} p_i(t) = 1. \qquad (4.1)$$

По размеченному графу состояний можно найти все вероятности состояний $p_i(t)$ как функции времени. Для этого составляются и решаются так называемые уравнения Колмогорова – особого вида дифференциальные уравнения, в которых неизвестными функциями являются вероятности состояний.

Как составить эти уравнения покажем на конкретном примере (рис. 4.1). Система имеет четыре состояния S_0, S_1, S_2, S_3. Рассмотрим $p_0(t)$ – это вероятность того, что система в момент времени t находится в состоянии S_0. Придадим t малое приращение Δt и найдем $p_0(t+\Delta t)$ – вероятность того, что и в момент $t+\Delta t$ система будет в состоянии S_0. Как это может произойти? Очевидно, двумя способами: либо 1) в момент t система уже была в состоянии S_0, а за время Δt не вышла из него; либо 2) в момент t система была в состоянии S_1 или в состоянии S_2, а за время Δt перешла в состояние S_0.

Найдем вероятность первого варианта. Вероятность того, что система в момент t была в состоянии S_0, равна $p_0(t)$. Эту вероятность нужно умножить на вероятность того, что, находившись в момент t в состоянии S_0, система за время Δt не перейдет из него ни в состояние S_2, ни в состояние S_3. Суммарный поток, выводящий систему из состояния S_0, тоже будет простейшим, с интенсивностью $\lambda_1 + \lambda_2$ (при наложении двух простейших потоков получается опять простейший поток, так как свойства стационарности, ординарности и отсутствия последействия сохраняются). Значит, вероятность того, что за время Δt система выйдет из состояния S_0, равна $(\lambda_1 + \lambda_2)\Delta t$; а вероятность того, что не выйдет: $1-(\lambda_1 + \lambda_2)\Delta t$. Отсюда вероятность первого варианта равна $p_0(t)(1-(\lambda_1 + \lambda_2)\Delta t)$.

Найдем вероятность второго варианта. Она равна вероятности того, что в момент t система была в состоянии S_1, а за время Δt перейдет в состояние S_0, т.е. она равна $p_1(t)\mu_1\Delta t)$ или была в состоянии S_2, а за время Δt перейдет в состояние S_0, т.е. она равна $p_2(t)\mu_2\Delta t)$.

Складывая вероятности обоих вариантов (по правилу сложения вероятностей), получим:

$$p_0(t+\Delta t) = p_0(t)(1-(\lambda_1 + \lambda_2)\Delta t) + p_1(t)\mu_1\Delta t) + p_2(t)\mu_2\Delta t).$$

Раскроем скобки, перенесем $p_0(t)$ в левую часть и разделим на Δt :

$$\frac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = \mu_1 p_1(t) + \mu_2 p_2(t) - (\lambda_1 + \lambda_2)\, p_0(t).$$

Устремим Δt к нулю. Слева в пределе получим производную функции $p_0(t)$. Таким образом, имеем дифференциальное уравнение:

$$\frac{dp_0(t)}{dt} = \mu_1 p_1(t) + \mu_2 p_2(t) - (\lambda_{12} + \lambda_2)\, p_0(t).$$

или

$$\frac{dp_0}{dt} = \mu_1 p_1 + \mu_2 p_2 - (\lambda_1 + \lambda_2)\, p_0. \tag{4.2}$$

Рассуждая аналогично для всех остальных состояний, напишем еще три дифференциальных уравнения. Присоединяя к ним уравнение (4.2), получим систему дифференциальных уравнений для вероятностей состояний:

$$
\begin{cases}
\dfrac{dp_0}{dt} = \mu_1 p_1 + \mu_2 p_2 - (\lambda_1 + \lambda_2)\, p_0, \\[2mm]
\dfrac{dp_1}{dt} = \lambda_1 p_0 + \mu_2 p_3 - (\lambda_2 + \mu_1)\, p_1\,, \\[2mm]
\dfrac{dp_2}{dt} = \lambda_2 p_0 + \mu_1 p_3 - (\lambda_1 + \mu_2)\, p_2, \\[2mm]
\dfrac{dp_3}{dt} = \lambda_2 p_1 + \lambda_1 p_2 - (\mu_1 + \mu_2)\, p_3\,.
\end{cases}
\tag{4.3}
$$

Получили систему четырех линейных дифференциальных уравнений с четырьмя неизвестными функциями p_0, p_1, p_2, p_3. Заметим, что одно из них (любое) можно отбросить, пользуясь тем, что $p_0 + p_1 + p_2 + p_3 = 1$:

Сформулируем общее правило составления уравнений Колмогорова. В левой части каждого из них стоит производная вероятности какого-то (например, *i*-го) состояния. В правой части – сумма произведений вероятностей всех состояний, *из которых идут дуги в данное состояние*, на интенсивности соответствующих потоков, минус суммарная интенсивность всех потоков, *выводящих систему из данного состояния*, умноженная на вероятность данного (*i*-го) состояния.

Пользуясь этим правилом, запишем уравнения Колмогорова для системы S, размеченный граф состояний которой дан на рис. 4.2.

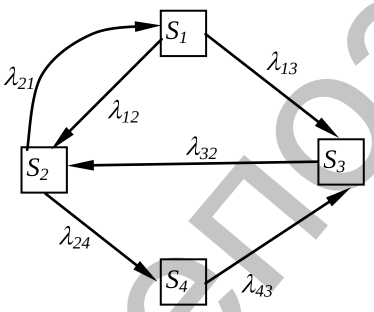


Рис. 4.2

$$
\begin{cases}
\dfrac{dp_1}{dt} = \lambda_{21}\, p_2 - (\lambda_{12} + \lambda_{13})\, p_1\,, \\[2mm]
\dfrac{dp_2}{dt} = \lambda_{12}\, p_1 + \lambda_{32}\, p_3 - (\lambda_{21} + \lambda_{24})\, p_2\,, \\[2mm]
\dfrac{dp_3}{dt} = \lambda_{13}\, p_1 + \lambda_{43}\, p_4 - \lambda_{32}\, p_3\,, \\[2mm]
\dfrac{dp_4}{dt} = \lambda_{24}\, p_2 - \lambda_{43}\, p_4\,.
\end{cases}
$$

Чтобы решить уравнения Колмогорова и найти вероятности состояний, прежде всего надо задать начальные условия. Если мы точно знаем начальное состояние системы S_i, то в начальный момент (при $t = 0$) имеем $p_i(0) = 1$, а все остальные начальные вероятности равны нулю. Так, например, уравнения (4.3) естественно решать при начальных условиях

$p_0(0) = 1$, $p_1(0) = p_2(0) = p_3(0) = 0$ (в начальный момент оба узла исправны).

Как решать подобные уравнения? Вообще говоря, линейные дифференциальные уравнения с постоянными коэффициентами можно решать аналитически, но это удобно только когда число уравнений не превосходит двух (иногда – трех). Если число уравнений больше, обычно решают численно.

Таким образом, уравнения Колмогорова дают возможность найти все вероятности состояний как функции времени.

Поставим теперь вопрос: что будет происходить с вероятностями состояний при $t \to \infty$? Будут ли $p_1(t), p_2(t), \ldots$ стремиться к каким-то пределам? Если эти пределы существуют и не зависят от начального состояния, то они ***называются предельными (финальными) вероятностями состояний***. В теории случайных процессов доказывается, что если число n состояний системы конечно и из каждого из них можно (за конечное число шагов) перейти в любое другое, то предельные вероятности существуют:

$$\lim_{t \to \infty} p_i(t) = p_i,\ i = 1,\ 2,\ \ldots n. \qquad (4.4)$$

Предельные вероятности мы будем обозначать теми же буквами p_1, p_2, … , что и сами вероятности состояний, но разумея под ними уже не переменные величины (функции времени), а постоянные числа. Очевидно, они тоже дают в сумме единицу:

$$\sum_{i=1}^{n} p_i = 1. \qquad (4.5)$$

Как понимать эти предельные вероятности? При $t \to \infty$ в системе S устанавливается предельный стационарный режим, в ходе которого система случайным образом меняет свои состояния, но их вероятности уже не зависят от времени. Предельную вероятность состояния S_i можно истолковать как *среднее относительное время пребывания системы в этом состоянии*. Например, если система имеет три состояния S_1, S_2, S_3 и их предельные вероятности равны 0,2, 0,3 и 0,5, это значит, что в предельном, стационарном режиме система в среднем две десятых времени проводит в состоянии S_1 , три десятых – в состоянии S_2 и половину времени в состоянии S_3 .

Как же вычислить предельные вероятности? Очень просто. Если вероятности p_1, p_2, … постоянны, то их производные равны нулю. Значит, чтобы найти предельные вероятности, нужно все левые части в уравнениях Колмогорова положить равными нулю решить полученную систему уже не дифференциальных, а линейных алгебраических уравнений. Эту систему линейных алгебраических уравнений легко написать прямо по графу. Перенесем отрицательный член каждого уравнения из правой части в левую. Получим систему уравнений, где в левой части каждого уравнения стоит предельная вероятность данного состояния, умноженная на суммарную интенсивность всех потоков, ведущих из данного состояния, а в правой части – сумма произведений интенсивностей всех потоков, входящих в данное состояние, на вероятности тех состояний, из которых эти потоки исходят.

Пользуясь этим правилом, запишем линейные алгебраические уравнения для предельных вероятностей состояний системы, граф которой дан на рис. 4.1:

$$\begin{cases} (\lambda_1 + \lambda_2)p_0 = \mu_1 p_1 + \mu_2 p_2, \\ (\lambda_2 + \mu_1)\, p_1 = \lambda_1 p_0 + \mu_2 p_3, \\ (\lambda_1 + \mu_2)\, p_2 = \lambda_2 p_0 + \mu_1 p_3, \\ (\mu_1 + \mu_2)p_3 = \lambda_2 p_1 + \lambda_1 p_2. \end{cases} \tag{4.6}$$

Эту систему четырех уравнений с четырьмя неизвестными казалось бы, вполне можно решить. Но уравнения (4.6) однородны (не имеют свободного члена) и, значит, определяют неизвестные только с точностью до произвольного множителя. К счастью, мы можем воспользоваться так называемым нормировочным условием:

$$p_0 + p_1 + p_2 + p_3 = 1 \tag{4.7}$$

и с его помощью решить систему. При этом одно (любое) из уравнений можно отбросить (оно вытекает как следствие остальных).

Пусть даны значения интенсивностей $\lambda_1 = 1,\ \lambda_2 = 2,\ \mu_1 = 2,\ \mu_2 = 3$, решим систему (4.6). Пожертвуем четвертым уравнением, добавив вместо него нормировочное условие (4.7). Уравнения примут вид:

$$\begin{cases} 3p_0 = 2\,p_1 + 3\,p_2, \\ 4\,p_1 = p_0 + 3\,p_3, \\ 4\,p_2 = 2p_0 + 2p_3, \\ p_0 + p_1 + p_2 + p_3 = 1. \end{cases} \tag{4.8}$$

Решив систему (4.8), получим:

$$p_0 = 6/15 = 0{,}40,\ p_1 = 3/15 = 0{,}2,\ p_2 – 4/15 \approx 0{,}27,\ p_3 = 2/15 \approx 0{,}13,$$

т.е. в предельном, стационарном режиме система S в среднем 40% времени будет находиться в состоянии S_0 (оба узла исправны), 20% – в состоянии S_1 (первый узел ремонтируется, второй работает), 27% – в состоянии S_2 (второй узел ремонтируется, первый работает), 13% – в состоянии S_3 (оба узла ремонтируются). Знание этих предельных вероятностей может помочь оценить среднюю эффективность работы системы и загрузку ремонтных бригад. Предположим, что система S в состоянии S_0 приносит в единицу времени доход 8 ден.ед., в состоянии S_1 – доход равен 3ден.ед., в состоянии S_2 – доход равен 5ден.ед., а в состоянии S_3 вообще дохода не приносит. Тогда в предельном, стационарном средний доход в единицу времени будет $0{,}40 \cdot 8 + 0{,}20 \cdot 3 + 0{,}27 \cdot 5 + 0 = 5{,}15$.

Теперь оценим загрузку ремонтных органов (рабочих), занятых ремонтом узлов 1 и 2. Узел 1 ремонтируется долю времени, равную $p_1 + p_3 = 0{,}20 + 0{,}13 = 0{,}33$. Узел 2 ремонтируется долю времени, равную $p_2 + p_3 = 0{,}27 + 0{,}13 = 0{,}40$.

Здесь может возникнуть вопрос об оптимизации решения. Допустим, что мы можем уменьшить среднее время ремонта того или иного узла (может быть и того, и другого), но это нам обойдется в какую-то сумму.

§ 5. ПРОЦЕССЫ ГИБЕЛИ И РАЗМНОЖЕНИЯ. ФОРМУЛА ЛИТТЛА

5.1. Схема гибели и размножения

Итак, имея в распоряжении размеченный граф состояний, можно легко записать уравнения Колмогорова для вероятностей состояний, а также написать и решить алгебраические уравнения для финальных (предельных) вероятностей. Для некоторых случаев удается получить решения в явном виде. В частности, это удается сделать, если граф состояний системы представляет собой, так называемую схему «гибели и размножения».

Граф состояний для схемы гибели и размножения имеет вид, показанный на рис. 5.1. Особенность этого графа в том, что все состояния системы можно вытянуть в одну цепочку, в которой каждое из средних состояний (S_1, S_2, …, S_{n-1}) связано прямой и обратной дугой с каждым из соседних состояний – правым и левым, а крайние состояния – S_0, S_n – только с одним соседним состоянием. Термин схема «гибели и размножения» ведет начало от биологических задач, где подобной схемой описывается изменение численности популяции.

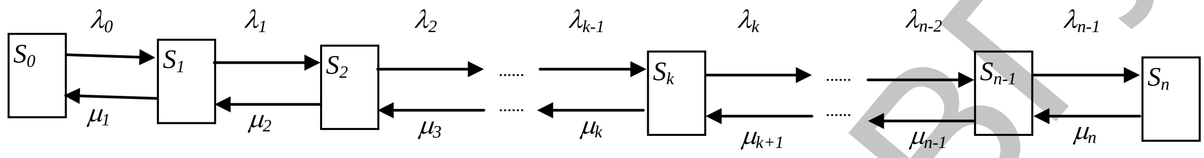


Рис. 5.1.

Схема гибели и размножения очень часто встречается в разных задачах практики, в частности – в теории массового обслуживания, поэтому полезно найти для нее финальные вероятности состояний.

Предположим, что все потоки заявок, переводящие систему по стрелкам графа, – простейшие (для краткости будем называть и систему S и протекающий в ней процесс – простейшими).

Пользуясь графом на рис. 5.1 и правилом, полученным в предыдущем параграфе, составим и решим алгебраические уравнения для предельных вероятностей состояний, (их существование вытекает из того, что из каждого состояния можно перейти в каждое другое, и число состояний конечно). Для первого состояния S_0 имеем:

$$\lambda_0 p_0=\mu_1 p_1. \tag{5.1}$$

Для второго состояния S_1:

$$(\lambda_1 + \mu_1)p_1=\lambda_0 p_0 + \mu_2 p_2.$$

Преобразуем последнее равенство, учитывая (5.1):

$$\lambda_1 p_1 + \mu_1 p_1=\lambda_0 p_0 + \mu_2 p_2 \Rightarrow \lambda_1 p_1=\mu_2 p_2.$$

Рассуждая аналогично, получим уравнения для любого состояния:

$$\lambda_k p_k=\mu_{k+1} p_{k+1},\ k = 0,\ 1,\ \dots ,\ n-1.$$

Таким образом, предельные вероятности удовлетворяют системе уравнений:

$$\begin{cases} \lambda_0 p_0 = \mu_1 p_1, \\ \lambda_1 p_1 = \mu_2 p_2, \\ \ldots\ldots\ldots\ldots\ldots \\ \lambda_{n-1} p_{n-1} = \mu_n p_n, \\ p_0 + p_1 + p_2 + p_1 = 1. \end{cases}$$

Выразим из первого уравнения p_1 через p_0 и подставим во второе. Из полученного уравнения выразим p_2 через p_0 и т.д. Получим, что

$$\begin{cases} p_1 = \dfrac{\lambda_0}{\mu_1} p_0, \\ p_2 = \dfrac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0, \\ \ldots\ldots\ldots\ldots \\ p_n = \dfrac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} p_0, \\ p_0 + p_1 + p_2 + p_1 = 1. \end{cases} \qquad (5.2)$$

Итак, все вероятности выражены через p_0. Подставим их в последнее уравнение системы

$$p_0 \left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \ldots + \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} \right) = 1$$

Тогда

$$p_0 = \left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \ldots + \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} \right)^{-1} \qquad (5.3)$$

По формулам (5.2) находим вероятности остальных состояний.

5.2. Формула Литтла

Обозначим через $L_{сист}$ среднее число заявок, находящихся в системе массового обслуживания (т.е. обслуживаемых или стоящих в очереди), а через $W_{сист}$ – среднее время пребывания заявки в системе. Формула Литтла и связывает эти две характеристики системы, когда в ней установился стационарный режим.

Рассмотрим любую СМО (одноканальную, многоканальную, марковскую, немарковскую, с ограниченной или неограниченной очередью и т.п.) и связанные с ней два потока заявок: поток заявок, прибывающих в СМО, и поток заявок покидающих СМО. Если в системе установился предельный, стационарный режим, то среднее число заявок, прибывающих в СМО за единицу времени, равно среднему числу заявок, покидающих СМО: оба потока имеют одну и ту же интенсивность λ.

Обозначим через $X(t)$ число заявок, прибывших в СМО за промежу-

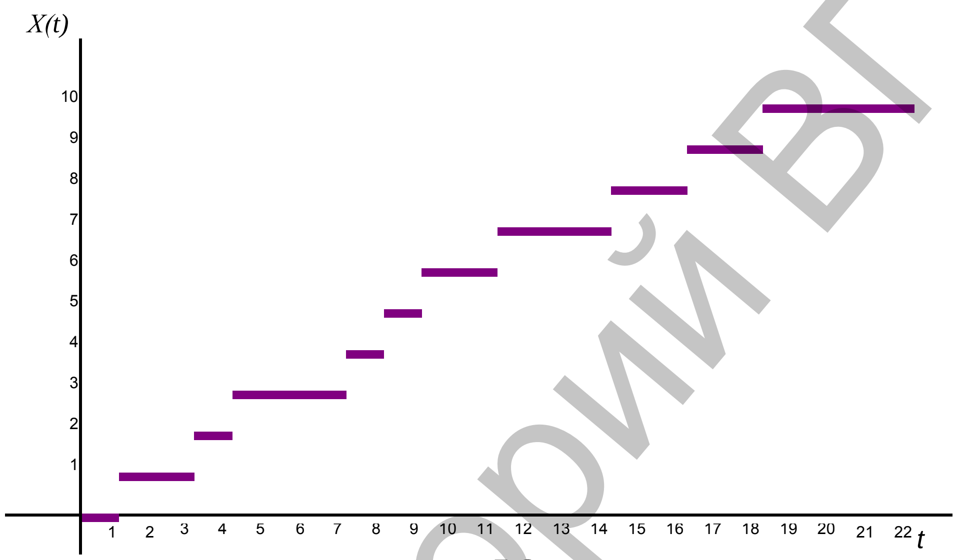


Рис. 5.2.

ток времени $(0;t)$, а через $Y(t)$ – число заявок, покинувших систему за промежуток времени $(0;t)$. Заявки поступают на обслуживание в случайные моменты времени, поэтому $X(t)$ – случайная функция, которая принимает

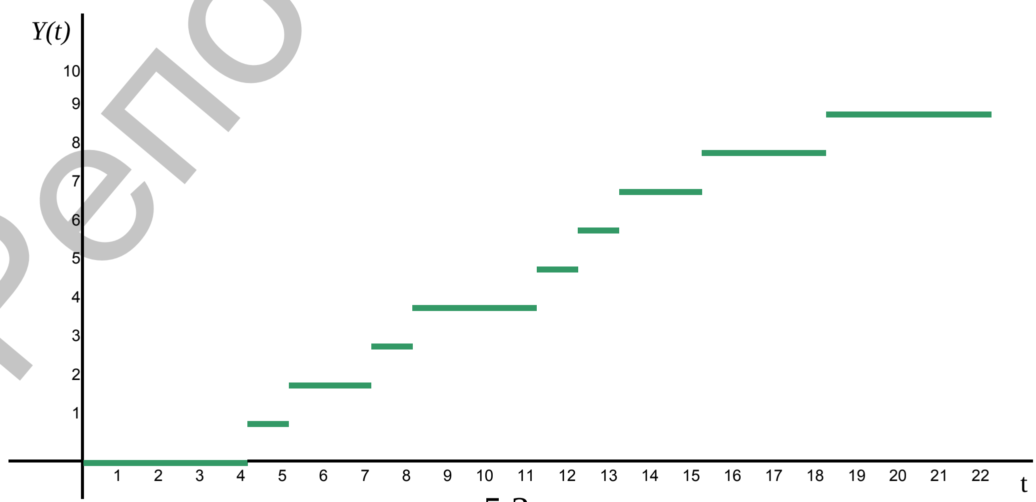


Рис. 5.3.

только целые неотрицательные значения и является неубывающей функцией при возрастании t (одна из возможных реализаций показана на рис. 5.2).

В силу того, что время обслуживания также случайно, то заявки покидают систему также в случайные моменты времени. Значит, относительно функции $Y(t)$ можно сказать то же, что и о функции $X(t)$ (одна из возможных реализаций функции $Y(t)$показана на рис. 5.3). Функции изменяют свои значения скачком (увеличиваются на единицу), $X(t)$ – в моменты приходов заявок, а $Y(t)$ – в моменты уходов заявок.

Очевидно, что для любого момента t разность $Z(t)=X(t)-Y(t)$есть не что иное, как число заявок, находящихся в СМО в момент времени t. График функции $Z(t)$, соответствующий графикам на рис. 5.2 и рис. 5.3,

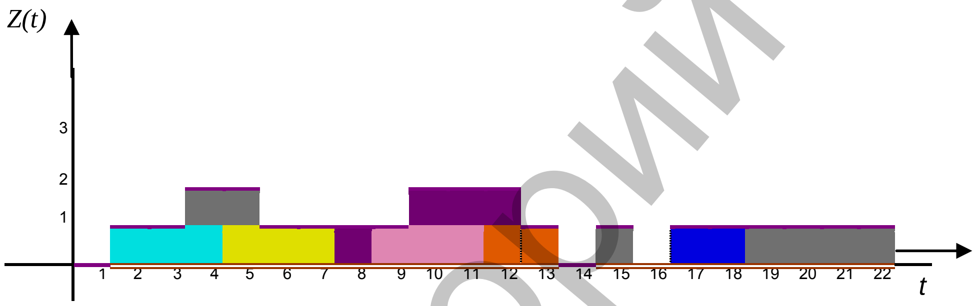


Рис. 5.4.

приведен на рис. 5.4.

На рис. 5.5 оба графика представлены одновременно: верхний – $X(t)$, нижний – $Y(t)$. Когда линии сливаются, в системе нет заявок. Например,

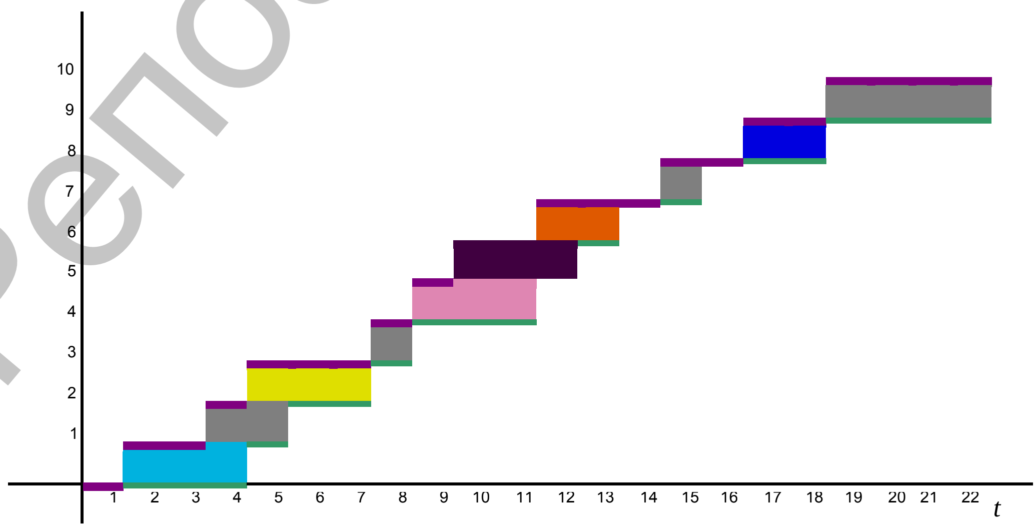


Рис. 5.5.

на промежутках [0; 1), [13; 14), [15; 16) в системе нет заявок, в момент

$t = 2$ или $t = 6$ в системе одна заявка, в момент $t = 10$ или на интервале [*4; 5*) системе находятся две заявки и т.д.

Рассмотрим очень большой промежуток времени T и вычислим для него среднее число заявок, находящихся в системе. Оно будет равно интегралу от функции $Z(t)$ на этом промежутке, деленному на длину интервала:

$$L_{сист} = \frac{1}{T}\int_0^T Z(t)dt \qquad (5.4)$$

Замечание-пояснение. $Z(t)$ – значение функции в точке, интеграл дает сумму значений функции на отрезке. $Z(t)$ – кусочно-непрерывная функция (имеет конечное число разрывов первого рода), такие функции интегрируемые.

С геометрической точки зрения интеграл (5.4) есть не что иное, как площадь фигуры, заштрихованной на рис. 5.4. На рис. 5.5, где совмещены графики функций $X(t)$ и $Y(t)$, та же площадь заштрихована по-другому. Фигура состоит из прямоугольников, каждый из которых имеет высоту, равную единице. Обозначим через t_1 время пребывания в системе первой из поступивших заявок, через t_2 – второй из поступивших заявок и т.д. Тогда основание каждого из прямоугольников равно времени пребывания в системе соответствующей заявки (первой, второй и т.д.). Правда, под конец промежутка T некоторые прямоугольники войдут не полностью в заштрихованную фигуру, но при большом промежутке T эти мелочи не будут играть роли. Таким образом, получили, что

$$\int_0^T Z(t)dt = \sum_i t_i ,$$

где сумма считается для всех заявок, пришедших в систему за время T. Среднее число заявок, пришедших в систему за время T, равно $T\lambda$. Если мы разделим сумму всех времен на среднее число заявок, то получим среднее время пребывания заявки в СМО:

$$W_{сист} = \frac{1}{T\lambda}\sum_i t_i , \qquad (5.5)$$

Из (5.4) и (5.5) получаем

$$L_{сист} = \frac{1}{T}\sum_i t_i \;\Rightarrow\; L_{сист} = \frac{1}{T}W_{сист}\cdot T\lambda \;\Rightarrow\; L_{сист} = W_{сист}\cdot\lambda ,$$

откуда

$$W_{сист} = \frac{1}{\lambda} \cdot L_{сист} \tag{5.6}$$

Это и есть формула Литтла: для любой СМО, при любом характере потока заявок, при любом распределении времени обслуживания заявок, при любой дисциплине обслуживания: ***среднее время пребывания заявки в системе равно среднему числу заявок в системе, деленному на интенсивность потока заявок***.

Точно так же выводится вторая формула Литтла, связывающая среднее время пребывания заявки в очереди и среднее число заявок в очереди:

$$W_{оч} = \frac{1}{\lambda} \cdot L_{оч}. \tag{5.7}$$

§ 6. СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С ОТКАЗАМИ

Рассмотрим СМО, которая была исследована по времени одной из первых и которая носит название «Задача Эрланга». Связана она была с работой телефонной станции.

Постановка задачи. Пусть имеется n каналов, на обработку которым поступает простейший поток заявок с интенсивностью равной λ. Интенсивность потока обслуживания заявок равна μ. Если заявка поступает в систему, и имеются свободные каналы, то заявка принимается на обслуживание. В случае, когда все каналы заняты, заявка получает отказ в обслуживании и покидает систему. Такая система массового обслуживания называется СМО с отказами (или с потерями). Требуется найти предельные вероятности состояний системы, а также характеристики ее эффективности:

$P_{отк}$ – вероятность отказа (вероятность того, что заявка покинет систему, не получив обслуживания);

Q – относительную пропускную способность (среднюю долю пришедших заявок, обслуживаемых системой или вероятность того, что заявка принимается к обслуживанию);

A – абсолютную пропускную способность (среднее число заявок, обслуживаемых в единицу времени);

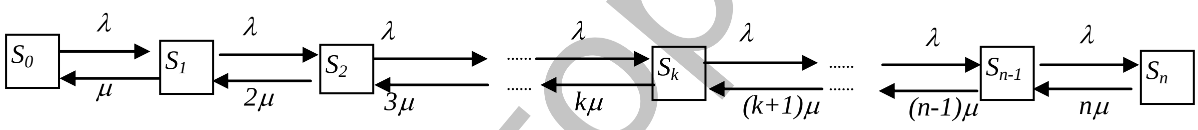
$\bar{k}$ – среднее число занятых каналов.

<u>Решение задачи</u>. По числу заявок в системе можно выделить следующие состояния:

- S_0 – в системе нет заявок (все каналы свободны),
- S_1 – в системе одна заявка (один канал занят, свободны *(n-1)* каналов),
- …………,
- S_n – в системе *n* заявок (все *n* каналов заняты).

Граф состояний соответствует схеме гибели и размножения и имеет вид, представленный на рис. 6.1.

Разметим этот граф. Интенсивность всех дуг вида (S_k, S_{k+1}), $k=\overline{0, n-1}$, соответствует интенсивности входного потока заявок и равна λ. (Из любого левого состояния S_k система переходит в правое соседнее состояние S_{k+1} под действие входного потока). Рассмотрим дугу (S_1, S_0). Система из со-



стояния S_1, когда работает один канал, переходит в состояние S_0, интенсивность обслуживания равна μ. Рассмотрим дугу (S_2, S_1). Когда система находится в состоянии S_2, работают два канала, т.е. интенсивность обслуживания равна 2μ. Аналогично, дуге (S_n, S_{n-1}) приписываем интенсивность, равную $n\mu$, так как работают *n* каналов, каждый с интенсивностью μ.

После разметки графа воспользуемся формулами (5.3) для нахождения предельной вероятности p_0:

$$p_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda\cdot\lambda}{\mu\cdot 2\mu} + \ldots + \frac{\lambda\cdot\lambda\cdot\ldots\cdot\lambda}{\mu\cdot 2\mu\cdot\ldots\cdot n\mu}\right)^{-1} = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2} + \ldots + \frac{\lambda^n}{n!\mu^n}\right)^{-1}$$

Введем обозначение

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

Величину ρ называют *приведенной интенсивностью* потока заявок. Она дает среднее число заявок, поступающих в систему за среднее время обслуживания одной заявки. Тогда для p_0 получаем

$$p_0 = \left(1 + \rho + \frac{\rho^2}{2!} + \ldots + \frac{\rho^n}{n!}\right)^{-1} \qquad (6.1)$$

Далее, используя формулы (5.2), получим предельные вероятности остальных состояний:

$$p_1 = \rho p_0 \ , \quad p_2 = \frac{\rho^2}{2!} \cdot p_0 \ , \quad \ldots \quad , p_n = \frac{\rho^n}{n!} \cdot p_0 \qquad (6.2)$$

Формулы (6.1) и (6.2) называются формулами Эрланга.

Далее найдем характеристики эффективности системы. Прежде всего найдем $P_{\text{отк}}$. Чтобы заявка покинула систему, не получив обслуживания, должны быть заняты все каналы. Другими словами

$$P_{\text{отк}} = p_n = \frac{\rho^n}{n!} \cdot p_0. \qquad (6.3)$$

Далее логично найти вероятность того, что заявка принимается к обслуживанию, т.е. относительную пропускную способность:

$$Q = 1 - P_{\text{отк}} = 1 - \frac{\rho^n}{n!} \cdot p_0. \qquad (6.4)$$

Зная относительную пропускную способность системы и интенсивность входного потока заявок, получим абсолютную пропускную способность – среднее число заявок, обслуживаемых в единицу времени:

$$A = \lambda Q = \lambda \cdot \left(1 - \frac{\rho^n}{n!} \cdot p_0\right). \qquad (6.5)$$

Немаловажный показатель эффективности системы – среднее число занятых каналов. Эту величину можно найти как математическое ожидание дискретной случайной величины. Возможные значения ее: 0, 1, 2, … , n. Известны вероятности этих значений: p_0 , p_1, p_2, … , p_n, поэтому

$$\bar{k} = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + \ldots + n \cdot p_n$$

Подставив значения вероятностей из (6.2), получим:

$$\bar{k} = 0 \cdot p_0 + 1 \cdot \rho p_0 + 2 \cdot \frac{\rho^2}{2!} \cdot p_0 + \ldots + n \cdot \frac{\rho^n}{n!} \cdot p_0 =$$

$$= \rho p_0 + \frac{\rho^2}{1!} \cdot p_0 + \ldots + \frac{\rho^n}{(n-1)!} \cdot p_0 =$$

$$= \rho p_0 \left(1 + \frac{\rho}{1!} + \ldots + \frac{\rho^{n-1}}{(n-1)!}\right)$$

Из (6.1) следует, что

$$p_0 \left(1 + \frac{\rho}{1!} + \ldots + \frac{\rho^{n-1}}{(n-1)!}\right) = 1 - \frac{\rho^n}{n!} \cdot p_0.$$

Тогда

$$\bar{k} = \rho\left(1 - \frac{\rho^n}{n!} \cdot p_0\right). \qquad (6.6)$$

Замечание. Формулу (6.6) можно получить и другим путем. Величина A – абсолютная пропускная способность – есть ни что иное как интенсивность потока заявок, *обслуженных* системой. Каждый занятый канал в среднем обслуживает μ заявок, поэтому среднее число занятых каналов равно

$$\bar{k} = \frac{A}{\mu} .$$

Отсюда с учетом (6.5) получаем формулу (6.6).

Пример 6.1. Рассмотрим систему телекоммуникаций, которая имеет одну телефонную линию. Если вызов приходит в момент, когда телефон занят, то имеем отказ. Пусть $\lambda = 0{,}8$ вызовов в минуту; средняя продолжительность разговора равна $\bar{t}_{обс} = 1{,}5$ мин. Будем считать поток заявок простейшим. Ставится задача: определить параметры СМО и сравнить фактическую абсолютную пропускную способность с той, которая имела бы место, если бы каждый разговор длился ровно 1,5 минуты и, вызовы шли один за другим.

Решение. Найдем интенсивность потока обслуживаний:

$$\mu = \frac{1}{\bar{t}_{обс}} = \frac{1}{1{,}5} = \frac{2}{3} \approx 0{,}667 ,$$

и приведенную интенсивность потока заявок

$$\rho = \frac{\lambda}{\mu} = \frac{0{,}8}{0{,}667} = \frac{6}{5} = 1{,}2 .$$

По формулам (6.1) и (6.2), где $n = 1$, имеем

$$p_0 = (1 + \rho)^{-1} = (1 + 1{,}2)^{-1} = \frac{5}{11} \approx 0{,}455 ;$$

$$p_1 = \rho\, p_0 = 1{,}2 \cdot \frac{5}{11} = \frac{6}{11} \approx 0{,}545.$$

Далее по формулам (6.3), (6.4) и (6.5) найдем

$P_{отк} = p_1 \approx 0{,}545;$

$Q = 1 - P_{отк} \approx 0{,}455;$

$A = \lambda Q = 0{,}8 \cdot 0{,}455 \approx 0{,}364.$

Вывод. СМО неудачная: больше половины заявок получают отказ в обслуживании.

Если бы каждый разговор длился ровно 1,5 минуты и, вызовы шли один за другим, то абсолютная пропускная способность была равна

$$A = \frac{1}{1,5} \approx 0,667 .$$

Сравнение этого значения с фактической абсолютной пропускной способностью системы говорит о том, что случайный характер заявок почти вдвое уменьшает пропускную способность.

Пример 6.2. Система рассматривается трехканальная (n = 3). Интенсивность входного потока равна λ = 0,8 вызовов в минуту; средняя продолжительность разговора равна $\overline{t}_{обс}$ = 1,5 мин. Определить характеристики эффективности СМО.

Решение. Как и в предыдущем примере найдем интенсивность потока обслуживаний: $\mu \approx 0,667$, и приведенную интенсивность потока заявок $\rho = 1,2$. Далее по формулам (6.1) и (6.2), где n = 3, имеем $p_0 \approx$ 0,312; $p_1 \approx$ 0,3; $p_2 \approx$ 0,224; $p_3 \approx$ 0,08. По формулам (6.3), (6.4) и (6.5)

$$P_{отк} = p_3 \approx 0,08;$$
$$Q = 1 - P_{отк} \approx 0,92;$$
$$A = \lambda Q = 0,8 \cdot 0,92 \approx 0,73.$$

Для данной системы необходимо определить средне число занятых каналов согласно (6.6):

$$\overline{k} = \;= 1,2 \cdot \left(1 - \frac{1,2^3}{3!} \cdot 0,312\right) \approx 1,092.$$

Вывод. По сравнению с одноканальной системой (см. пример 1) эффективность трехканальной системы с точки зрения обслуживания возросла, т.к. практически все заявки принимаются к обслуживанию, однако, эффективность использования приборов крайне низка (фактически постоянно простаивают два прибора).

§ 7. СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С ОЧЕРЕДЬЮ

Здесь рассмотрим системы массового обслуживания, которые для заявки допускают возможность стать в очередь в ожидании обслуживания. Подобного рода ситуации на практике встречаются очень часто. Различают СМО с конечными очередями и с неограниченными очередями.

7.1. СМО с конечной очередью

Пусть имеется n каналов, на обработку которым поступает простейший поток заявок с интенсивностью λ. Интенсивность потока обслуживания заявок равна μ. Если заявка поступает в систему, и имеются свободные каналы, то заявка принимается на обслуживание. В случае, когда все каналы заняты, заявка становится в очередь, в которой m мест. Если при поступлении заявки в систему все каналы заняты и в очереди нет свободных мест, то заявка получает отказ в обслуживании и покидает систему. Требуется найти предельные вероятности состояний системы, а также характеристики ее эффективности:

$P_{\text{отк}}$ – вероятность отказа (вероятность того, что заявка покинет систему, не получив обслуживания);

Q – относительную пропускную способность (вероятность того, что заявка принимается к обслуживанию);

A – абсолютную пропускную способность (среднее число заявок, обслуживаемых в единицу времени);

$\bar{k}$ – среднее число занятых каналов;

$L_{сист}$ – среднее число заявок в системе;

$W_{сист}$ – среднее время пребывания заявки в системе;

$L_{оч}$ – среднее число заявок в очереди (или среднюю длину очереди);

$W_{оч}$ – среднее время ожидания заявки в очереди.

Решение задачи. По числу заявок в системе можно выделить следующие состояния:

- S_0 – в системе нет заявок (все каналы свободны, и очередь пуста),
- S_1 – в системе одна заявка (один канал занят обслуживанием, $n\text{-}1$ каналов свободных, очередь пуста),

– ………..,

– S_n – в системе n заявок (все n каналов заняты, очередь пуста),

– S_{n+1} – в системе число заявок равно $n+1$ (все каналы заняты, в очереди одна заявка),

– ………..,

– S_{n+m} – в системе число заявок равно $n+m$ (все каналы заняты, все m мест в очереди заняты).

Граф состояний соответствует схеме гибели и размножения (см. § 5) и имеет вид, представленный на рис. 7.1. Разметим этот граф.

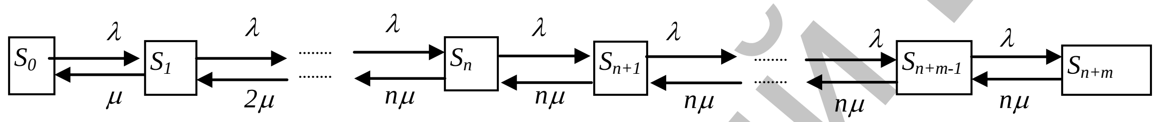


Рис. 7.1.

Интенсивность всех дуг вида (S_k, S_{k+1}), $k=\overline{0,n+m-1}$, соответствует интенсивности входного потока заявок и равна λ. (Из любого левого состояния S_k система переходит в правое соседнее состояние S_{k+1} под действие входного потока). Рассмотрим дугу (S_1, S_0). Система из состоянияS_1, когда работает один канал, переходит в состояние S_0, интенсивность обслуживания равна μ. Рассмотрим дугу (S_2, S_1). Когда система находится в состоянии S_2, работают два канала, т.е. интенсивность обслуживания равна 2μ. Аналогично, дуге (S_n, S_{n-1}) приписываем интенсивность, равную $n\mu$, так как работают n каналов, каждый с интенсивностью μ. Все дуги вида (S_k, S_{k-1}), $k=\overline{n+1,n+m}$, имеют интенсивность равную $n\mu$, так как опять-таки работают все n каналов. Таким образом, имеем

$$\lambda_0=\lambda_1=\ldots=\lambda_{n+m-1}=\lambda,$$

$$\mu_1=\mu,\ \mu_2=2\mu,\ \ldots,\ \mu_n=\mu_{n+1}=\ldots\ \mu_{n+m}=n\mu.$$

Число состояний системы конечно и из каждого из них можно (за конечное число шагов) перейти в любое другое, поэтому предельные вероятности существуют. После разметки графа воспользуемся формулами (5.3) для нахождения предельной вероятности p_0:

$$p_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2!\mu^2} + \ldots + \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^{n+1}}{n!\,n\mu^{n+1}} + \ldots + \frac{\lambda^{n+m}}{n!\,n^m\mu^{n+m}}\right)^{-1}$$

Как и прежде введем величину ρ – приведенную интенсивность потока заявок. Тогда для p_0 получаем

$$p_0 = \left(1 + \rho + \frac{\rho^2}{2!} + \ldots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n!n} + \ldots + \frac{\rho^{n+m}}{n!n^m}\right)^{-1} =$$

$$= \left(1 + \rho + \frac{\rho^2}{2!} + \ldots + \frac{\rho^{n-1}}{(n-1)!} + \frac{\rho^n}{n!}\left(1 + \frac{\rho}{n} + \ldots + \frac{\rho^m}{n^m}\right)\right)^{-1}.$$

Выражение во внутренних скобках есть сумма членов геометрической прогрессии с первым членом равным 1, знаменателем $\frac{\rho}{n}$, число членов равно $m+1$. Воспользуемся известной формулой и получим

$$p_0 = \left(\sum_{k=0}^{n-1} \frac{\rho^k}{k!} + \frac{\rho^n\left(1 - \left(\frac{\rho}{n}\right)^{m+1}\right)}{n!\left(1 - \frac{\rho}{n}\right)}\right)^{-1}. \qquad (7.1)$$

Далее, используя формулы (5.2), получим предельные вероятности остальных состояний:

$$p_1 = \rho p_0\,,\ p_2 = \frac{\rho^2}{2!}\cdot p_0\,,\ \ldots,\ p_n = \frac{\rho^n}{n!}\cdot p_0\,,\ p_{n+1} = \frac{\rho^{n+1}}{n!n}\cdot p_0\,,\ \ldots,\ p_{n+m} = \frac{\rho^{n+m}}{n!n^m}\cdot p_0. \qquad (7.2)$$

Формулы (7.1) и (7.2) представляют вариант формул Эрланга для системы массового обслуживания с конечной очередью.

Определим характеристики эффективности системы. Прежде всего найдем $P_{\text{отк}}$. Чтобы заявка покинула систему, не получив обслуживания, должны быть заняты все каналы и в очереди не было свободного места. Другими словами

$$P_{\text{отк}} = p_{n+m} = \frac{\rho^{n+m}}{n!\,n^m}\cdot p_0. \qquad (7.3)$$

Далее найдем вероятность того, что заявка принимается к обслуживанию, т.е. относительную пропускную способность:

$$Q = 1 - P_{\text{отк}} = 1 - \frac{\rho^{n+m}}{n!n^m}\cdot p_0. \qquad (7.4)$$

Зная относительную пропускную способность системы и интенсивность входного потока заявок, получим абсолютную пропускную способность – среднее число заявок, обслуживаемых в единицу времени:

$$A = \lambda Q = \lambda\cdot\left(1 - \frac{\rho^{n+m}}{n!\,n^m}\cdot p_0\right). \qquad (7.5)$$

Среднее число занятых каналов равно (см. замечание в § 6)

$$\bar{k} = \frac{A}{\mu} = \rho\left(1 - \frac{\rho^{n+m}}{n!\,n^m}\cdot p_0\right). \qquad (7.6)$$

Найдем $L_{оч}$, используя формулу для вычисления математического ожидания дискретной случайной величины. $L_{оч}$ – дискретная случайная величина, имеющая возможные значения $1, 2,\ldots, m$ с вероятностями $p_{n+1},\ldots, p_{n+m}$ соответственно.

$$L_{оч} = \sum_{k=1}^{m} k\,p_{n+k} = \sum_{k=1}^{m} k\cdot\frac{\rho^{n+k}}{n!n^k}\cdot p_0 =$$

$$= p_0\,\frac{\rho^{n+1}}{n!\,n}\sum_{k=1}^{m} k\,\frac{\rho^{k-1}}{n^{k-1}} = p_0\,\frac{\rho^{n+1}}{n!\,n}\sum_{k=1}^{m} k\left(\frac{\rho}{n}\right)^{k-1}.$$

Вычислим сумму в последнем выражении. Ведем (временно) обозначение

$$\frac{\rho}{n} = g.$$

Тогда

$$\sum_{k=1}^{m} k\left(\frac{\rho}{n}\right)^{k-1} = \sum_{k=1}^{m} k\,g^{k-1} = \sum_{k=1}^{m}\frac{d}{dg}g^{k}.$$

Меняя местами операции дифференцирования и суммирования, получим

$$\frac{d}{dg}\sum_{k=1}^{m} g^{k}.$$

Сумма в последней формуле есть не что иное, как сумма m членов геометрической прогрессии с первым членом g и знаменателем g:

$$\frac{d}{dg}\left(\frac{g\,(1-g^m)}{1-g}\right) = \frac{1-g^m(1+m(1-g))}{(1-g)^2}.$$

Выполним обратную замену и окончательно вычислим среднее число заявок в очереди

$$L_{оч} = p_0\,\frac{\rho^{n+1}\left(1-\left(\frac{\rho}{n}\right)^m\left(1+m\left(1-\frac{\rho}{n}\right)\right)\right)}{n!\,n\left(1-\frac{\rho}{n}\right)^2} = p_0\,\frac{\rho^{n+1}\left(1-\left(\frac{\rho}{n}\right)^m\left(1+m\left(1-\frac{\rho}{n}\right)\right)\right)}{(n-1)!\,(n-\rho)^2}. \qquad (7.7)$$

Среднее число $L_{сист}$ заявок в СМО складывается из двух величин – среднее число заявок в очереди и среднее число заняты каналов:

$$L_{сист} = L_{оч} + \bar{k}. \qquad (7.8)$$

Разделим выражения для $L_{оч}$ и $L_{сист}$ на интенсивность входного потока и получим среднее время ожидания заявки в очереди и среднее время пребывания заявки в системе:

$$W_{оч} = \frac{1}{\lambda}\cdot L_{оч}, \quad W_{сист} = 0\frac{1}{\lambda}\cdot L_{сист}. \qquad (7.9)$$

Замечание. Значение можно получить и другим путем. Величина $W_{оч}$ – среднее время ожидания заявки в очереди – есть дискретная случайная величина. Если в очереди одна заявка, то время ожидания равно $\frac{1}{\mu_{n+1}} = \frac{1}{n\mu}$, вероятность того, что эта одна заявка будет ждать обслуживания равна p_n, т.е. вероятности того, что все каналы заняты. Если в очереди две заявки, то время ожидания равно $\frac{1}{\mu_{n+1}}+\frac{1}{\mu_{n+2}} = \frac{2}{n\mu}$, вероятность того, что эти две заявки будут ждать обслуживания, равна p_{n+1}, т.е. вероятности того, что все каналы заняты и одно место в очереди уже занято. Значит,

$$W_{оч} = \sum_{k=1}^{m} p_{n-1+k}\cdot\frac{k}{n\mu}.$$

Отсюда затем можно получить формулу (7.7).

Пример 7.1. Рассмотрим автозаправочную станцию (АЗС), на которой действуют 2 заправочные колонки. Интенсивность поступления на АЗС машин для заправки равна $\lambda = 2$ машины в минуту; среднее время обслуживания – 2 минуты. Размеры АЗС по площади невелики и лишь 3 машины могут стать в очередь. Задача: определить параметры СМО и проанализировать эффективность работы АЗС.

Решение. Найдем интенсивность потока обслуживаний:

$$\mu = \frac{1}{\bar{t}_{обс}} = \frac{1}{2} = 0{,}500,$$

и приведенную интенсивность потока заявок

$$\rho = \frac{\lambda}{\mu} = \frac{2}{0{,}5} = 4\ .$$

По формулам (7.1), где $n = 2$, $m = 3$, имеем

$$p_0=\left(\sum_{k=0}^{1}\frac{4^k}{k!}+\frac{4^2\left(1-\left(\frac{4}{2}\right)^{3+1}\right)}{2!\left(1-\frac{4}{2}\right)}\right)^{-1}=\left(1+\frac{4}{1!}+\frac{16(1-2^4)}{2(1-2)}\right)^{-1}=\left(1+4+\frac{16(-15)}{2(-1)}\right)^{-1}=0{,}008.$$

Далее, используя формулы (7.2), получим предельные вероятности остальных состояний:

$$\begin{cases} p_1 = 4p_0, \\ p_2 = \dfrac{4^2}{2!} \cdot p_0 = 8 \cdot p_0, \\ p_3 = \dfrac{4^3}{2!\,2} \cdot p_0 = 16 \cdot p_0, \\ p_4 = \dfrac{4^4}{2!\,2^2} \cdot p_0 = 32 \cdot p_0, \\ p_5 = \dfrac{4^5}{2!\,2^3} \cdot p_0 = 64 \cdot p_0, \end{cases} \Rightarrow \begin{cases} p_1 = 0{,}032, \\ p_2 = 0{,}064, \\ p_3 = 0{,}128, \\ p_4 = 0{,}256, \\ p_5 = 0{,}512. \end{cases}$$

Далее

$$P_{\text{отк}} = p_5 = 0{,}512,$$

$$Q = 1 - P_{\text{отк}} = 0{,}488,$$

$$A = \lambda Q = 2 \cdot 0{,}488 = 0{,}976,$$

$$\bar{k} = \frac{A}{\mu} = \frac{0{,}976}{0{,}5} = 1{,}952,$$

$$L_{оч} = 0{,}008 \frac{4^3\left(1 - \left(\frac{4}{2}\right)^3 \left(1 + 3\left(1 - \frac{4}{2}\right)\right)\right)}{(2\text{-}1)!\,(2-4)^2} = 0{,}008 \frac{64\left(1 - 2^3\left(1 + 3(1-2)\right)\right)}{1!\,(-2)^2} =$$

$$= 0{,}008 \frac{64\left(1 - 8\,(1\text{-}3)\right)}{4} = 0{,}008 \cdot 16(1 + 16) = 2{,}176,$$

$$L_{сист} = L_{оч} + \bar{k} = 2{,}176 + 1{,}952 = 4{,}128,$$

$$W_{оч} = \frac{1}{\lambda} \cdot L_{оч} = \frac{1}{2} \cdot 2{,}176 = 1{,}088,$$

$$W_{сист} = \frac{1}{\lambda} \cdot L_{сист} = \frac{1}{2} \cdot 4{,}128 = 2{,}064.$$

Вывод. С точки зрения и потребителей, и владельца система вполне приемлемая. Среднее время пребывания в системе (*2,064 мин*) почти совпадает с длительностью обслуживания (*2 мин*). Загруженность каналов полная, в очереди есть одно свободное место.

7.2. Одноканальная СМО с бесконечной очередью

Такие системы на практике встречаются очень часто (небольшой магазин с одним продавцом, врач в поликлинике, телефон-автомат и т.п.). Итак, пусть имеется СМО, где $n=1$, а на очередь не накладывается никаких ограничений. Будем считать, что в СМО поступает простейший поток заявок с интенсивностью λ. Интенсивность потока обслуживания заявок равна μ. Требуется найти предельные вероятности состояний системы, а также характеристики ее эффективности:

$L_{сист}$ – среднее число заявок в системе;

$W_{сист}$ – среднее время пребывания заявки в системе;

$L_{оч}$ – среднее число заявок в очереди (или среднюю длину очереди);

$W_{оч}$ – среднее время ожидания заявки в очереди;

$P_{зан}$ – вероятность того, что канал занят.

Такие величины, как абсолютную пропускную способность A, относительную пропускную способность Q , то их вычислять не надо, в силу того, что очередь неограниченна (каждая заявка рано или поздно будет обслужена). Поэтому $A = \lambda$, $Q = 1$.

Решение задачи. По числу заявок в системе можно выделить следующие состояния (число состояний бесконечно):

– S_0 – в системе нет заявок (канал свободен, и очередь пуста),

– S_1 – канал занят обслуживанием, очередь пуста,

– S_2 – канал занят обслуживанием, в очереди одна заявка,

– …………

– S_k – канал занят обслуживанием, в очереди k-1 заявок,

– …………

Граф состояний соответствует схеме гибели и размножения

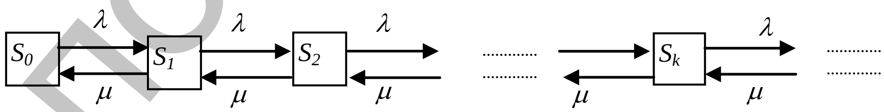


Рис. 7.2.

(см. § 5) и имеет вид, представленный на рис. 7.2. Разметим этот граф. Интенсивность всех дуг вида (S_k, S_{k+1}), $k=\overline{0,\infty}$, соответствует интенсивности входного потока заявок и равна λ. А интенсивность дуг (S_k, S_{k-1}), $k=\overline{1,\infty}$, соответствует интенсивности потока обслуживаний и равна μ. Таким образом, имеем

$$\lambda_0 = \lambda_1 = \ldots = \lambda_k = \ldots = \lambda,$$

$$\mu_1 = \mu_2 = \ldots = \mu_k = \ldots = \mu.$$

Число состояний системы бесконечно, и поэтому, вообще говоря, предельные вероятности не всегда существуют. Если система перегружена, то при $t \to \infty$ очередь может неограниченно возрастать. Доказано, что предельные вероятности существуют при $\rho<1$. Даже при $\rho=1$ (за время обслуживания одной заявки в среднем в систему приходит также одна заявка) СМО справляется с потоком заявок, если входной поток регулярный, и время обслуживания тоже не случайное, а равно интервалу между заявками.

Воспользуемся формулами (5.3). В данном случае число слагаемых в сумме бесконечно. Для предельной вероятности p_0 получаем:

$$p_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{\mu^2} + \ldots + \frac{\lambda^k}{\mu^k} + \ldots\right)^{-1} = \left(1 + \rho + \rho^2 + \ldots + \rho^k + \ldots\right)^{-1}.$$

Выражение в скобках есть сумма членов бесконечной геометрической прогрессии с первым членом равным 1 и знаменателем ρ. Воспользуемся известной формулой и получим

$$p_0 = \left(\frac{1}{1-\rho}\right)^{-1} = 1-\rho. \tag{7.10}$$

Далее, используя формулы (5.2), получим предельные вероятности остальных состояний:

$$\begin{cases} p_1 = \rho p_0, \\ p_2 = \rho^2 \cdot p_0, \\ \ldots\ldots\ldots\ldots \\ p_k = \rho^k \cdot p_0, \\ \ldots\ldots\ldots\ldots \end{cases} \Rightarrow \begin{cases} p_1 = \rho(1-\rho), \\ p_2 = \rho^2 \cdot (1-\rho), \\ \ldots\ldots\ldots\ldots \\ p_k = \rho^k \cdot (1-\rho), \\ \ldots\ldots\ldots\ldots \end{cases} \tag{7.11}$$

Из сравнения значений предельных вероятностей по формулам (7.10) и (7.11) видим, что p_0 – самая большая из всех, т.е. самое вероятное число заявок в системе равно нулю.

Определим характеристики эффективности системы. В данном случае, прежде всего, найдем $L_{сист}$, которая является дискретной случайной величиной, имеющей возможные значения $0, 1, 2, \ldots, k, \ldots$ с вероятностями $p_0, p_1, p_2, \ldots, p_k, \ldots$ соответственно. Тогда

$$L_{сист} = \sum_{k=0}^{\infty} k p_k = \sum_{k=1}^{\infty} k p_k.$$

Подставим значения из (7.11) и преобразуем

$$L_{сист}= \sum_{k=1}^{\infty} k\rho^k \cdot (1-\rho) = \rho\cdot(1-\rho)\cdot \sum_{k=1}^{\infty} k\rho^{k-1}.$$

Далее применим тот же прием, что и при вычислении $L_{оч}$ в п.I.

$$L_{сист}= \rho\cdot(1-\rho)\cdot \sum_{k=1}^{\infty} \frac{d}{d\rho}(\rho^k) = \rho\cdot(1-\rho)\cdot\frac{d}{d\rho}\sum_{k=1}^{\infty}\rho^k = \rho\cdot(1-\rho)\cdot\frac{d}{d\rho}\left(\frac{\rho}{1-\rho}\right).$$

Найдем производную и окончательно получим:

$$L_{сист}= \frac{\rho}{1-\rho}. \qquad (7.12)$$

По формуле Литтла вычислим

$$W_{сист} = \frac{1}{\lambda}\cdot L_{сист} = \frac{\rho}{\lambda(1-\rho)}. \qquad (7.13)$$

Определим $P_{зан}$. Поскольку в системе только один канал (и он либо занят, либо свободен), то вероятность того, что канал занят, равна

$$P_{зан} = 1 - p_0 = 1 - (1-\rho) = \rho. \qquad (7.14)$$

Среднее число $L_{сист}$ заявок в СМО складывается из двух величин – среднее число $L_{оч}$ заявок в очереди и среднее число $L_{обс}$ заявок под обслуживанием. Последнее значение равно либо нулю (канал свободен с вероятностью p_0), либо единице (канал занят с вероятностью $P_{зан}$). Следовательно,

$$L_{обс}= 0\cdot p_0 + 1\cdot P_{зан}= P_{зан} = \rho. \qquad (7.15)$$

Тогда

$$L_{оч} = L_{сист} - L_{обс} = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}. \qquad (7.16)$$

По формуле Литтла вычислим

$$W_{оч} = \frac{1}{\lambda}\cdot L_{оч} = \frac{\rho^2}{\lambda(1-\rho)}. \qquad (7.17)$$

Пример 7.2. (для самостоятельного решения). Рассмотрим железнодорожную сортировочную станцию, на которую поступает простейший поток составов с интенсивностью $\lambda = 2$ состава в час. Обслуживание (расформирование) состава длится случайное (показательное) время со средним значением $\bar{t}_{обс}= 20$ минут. В парке прибытия имеются два пути, на которых могут ожидать обслуживания прибывающие составы. Если оба пути заняты, то составы ожидают обслуживания на внешних путях.

Требуется определить (для предельного, стационарного режима работы станции) следующие параметры: $L_{сист}$– среднее число составов в системе, $W_{сист}$ – среднее время пребывания заявки в системе (на обслуживании, на внутренних путях, на внешних путях), $L_{оч}$ – среднее число составов в очереди на расформирование (как на внешних, так и на внутренних путях), $W_{оч}$– среднее время ожидания состава в очереди, $L_{внеш}$– среднее число составов, ожидающих расформирования на внешних путях, $W_{внеш}$– среднее время ожидания обслуживания на внешних путях. Кроме этого, найти суммарный суточный штраф, который заплатит железнодорожная станция за простои составов на внешних путях, если за один час простоя одного состава штраф составляет *a* ден.ед.

Ответы: $L_{сист} = 2$ состава, $W_{сист} = 1$ час, $L_{оч} = 4/3$ состава, $W_{оч} = 2/3$ часа, $L_{внеш} = 16/27$ состава, $W_{внеш} = 8/27$ часа. Штраф составляет 14,2*a* у.е.

7.3. Многоканальная СМО с бесконечной очередью

Пусть имеется СМО, в которой n каналов обслуживания, на очередь не накладывается никаких ограничений. Будем считать, что в СМО поступает простейший поток заявок с интенсивностью λ. Интенсивность потока обслуживания заявок равна μ. Требуется найти предельные вероятности состояний системы, а также характеристики ее эффективности:

$\bar{k}$ – среднее число занятых каналов;

$L_{сист}$ – среднее число заявок в системе;

$W_{сист}$ – среднее время пребывания заявки в системе;

$L_{оч}$ – среднее число заявок в очереди (или среднюю длину очереди);

$W_{оч}$ – среднее время ожидания заявки в очереди.

Решение задачи. По числу заявок в системе можно выделить следующие состояния (теоретически число состояний бесконечно):

– S_0 – в системе нет заявок (все каналы свободны, и очередь пуста),
– S_1 – в системе одна заявка (один канал занят обслуживанием, $n-1$ каналов свободных, очередь пуста),
– …………
– S_n – в системе n заявок (все n каналов заняты, очередь пуста),
– S_{n+1} – в системе число заявок равно $n+1$ (все каналы заняты, в очереди одна заявка),
– …………

– S_{n+k} – в системе число заявок равно $n+ k$ (все каналы заняты, k мест в очереди заняты),

– …………

Граф состояний соответствует схеме гибели и размножения (см. § 5) и имеет вид, представленный на рис. 7.3. Аналогично п. I , имеем

$$\lambda_0 = \lambda_1 = \ldots = \lambda_n = \ldots = \lambda_k = \ldots = \lambda,$$

$$\mu_1 = \mu, \quad \mu_2 = 2\mu, \quad \ldots, \quad \mu_n = \mu_{n+1} = \ldots = \mu_{n+k} = \ldots = n\mu.$$

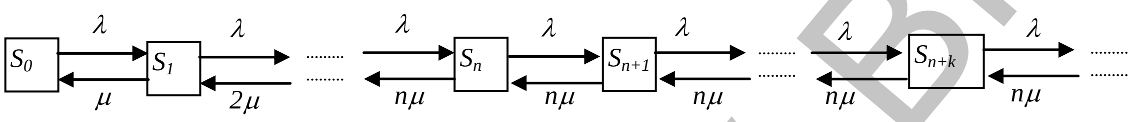


Рис. 7.3

Число состояний системы бесконечно, и поэтому, вообще говоря, предельные вероятности не всегда существуют. Если система перегружена, то при $t \to \infty$ очередь может неограниченно возрастать. Доказано, что предельные вероятности существуют при $\rho/n<1$.

Воспользуемся формулами (5.3). В данном случае число слагаемых в сумме бесконечно. Для предельной вероятности p_0 получаем:

$$p_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2!\mu^2} + \ldots + \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^{n+1}}{n!\, n\mu^{n+1}} + \ldots + \frac{\lambda^{n+k}}{n!\, n^k\mu^{n+k}} + \ldots\right)^{-1} =$$

$$= \left(1 + \rho + \frac{\rho^2}{2!} + \ldots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n!n} + \ldots + \frac{\rho^{n+k}}{n!n^k} + \ldots\right)^{-1} =$$

$$= \left(1 + \rho + \frac{\rho^2}{2!} + \ldots + \frac{\rho^n}{n!} + \frac{\rho^n}{n!}\left(\frac{\rho}{n} + \ldots + \frac{\rho^k}{n^k} + \ldots\right)\right)^{-1}.$$

Выражение во внутренних скобках есть сумма членов бесконечной геометрической прогрессии с первым членом равным $\frac{\rho}{n}$, и знаменателем $\frac{\rho}{n}<1$. Воспользуемся известной формулой и получим

$$p_0 = \left(\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^n\left(\frac{\rho}{n}\right)}{n!\left(1-\frac{\rho}{n}\right)}\right)^{-1} = \left(\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)}\right)^{-1}. \qquad (7.18)$$

Далее, используя формулы (5.2), получим предельные вероятности остальных состояний:

$$p_1=\rho\, p_0,\ \ p_2=\frac{\rho^2\cdot p_0}{2!},\ \dots,\ p_n=\frac{\rho^n\cdot p_0}{n!},\ \ p_{n+1}=\frac{\rho^{n+1}\cdot p_0}{n!n},\ \dots,\ p_{n+k}=\frac{\rho^{n+k}\cdot p_0}{n!\,n^k},\ \dots \qquad (7.19)$$

Поскольку $A=\lambda$, то среднее число занятых каналов равно

$$\bar{k}=\frac{A}{\mu}\ \Rightarrow\ \bar{k}=\frac{\lambda}{\mu}=\rho. \qquad (7.20)$$

Как и п. I, найдем $L_{оч}$, используя формулу для вычисления математического ожидания дискретной случайной величины:

$$L_{оч}=\sum_{k=1}^{\infty}kp_{n+k}=\sum_{k=1}^{\infty}k\cdot\frac{\rho^{n+k}}{n!\,n^k}\cdot p_0=$$

$$=p_0\frac{\rho^{n+1}}{n!\,n}\sum_{k=1}^{\infty}k\frac{\rho^{k-1}}{n^{k-1}}=p_0\frac{\rho^{n+1}}{n!\,n}\sum_{k=1}^{\infty}k\left(\frac{\rho}{n}\right)^{k-1}.$$

Для вычисления суммы в последнем выражении применим тот же прием, что и п. I. Получим:

$$\sum_{k=1}^{\infty}k\left(\frac{\rho}{n}\right)^{k-1}=\frac{1}{\left(1-\frac{\rho}{n}\right)^2}.$$

И окончательно вычислим среднее число заявок в очереди

$$L_{оч}=p_0\frac{\rho^{n+1}}{n!\,n}\frac{1}{\left(1-\frac{\rho}{n}\right)^2}=\frac{\rho^{n+1}p_0}{(n-1)!\,(n-\rho)^2}. \qquad (7.21)$$

Затем найдем среднее число $L_{сист}$ заявок в СМО:

$$L_{сист}=L_{оч}+\bar{k}=\frac{\rho^{n+1}p_0}{(n-1)!\,(n-\rho)^2}+\rho. \qquad (7.22)$$

Разделим выражения для $L_{оч}$ и $L_{сист}$ на интенсивность входного потока и получим среднее время ожидания заявки в очереди и среднее время пребывания заявки в системе:

$$W_{оч}=\frac{1}{\lambda}\cdot L_{оч},\qquad W_{сист}=\frac{1}{\lambda}\cdot L_{сист}. \qquad (7.23)$$

Пример 7.3. (Из книги Вентцель Е.С. Исследование операций. 2001) Железнодорожная касса по продаже билетов с двумя окошками представляет собой двухканальную СМО с неограниченной очередью сразу к двум окошкам (если одно окошко освобождается, ближайший в очереди пассажир его занимает). Касса продает билеты в два пункта: А и В. Интенсивность потока заявок (пассажиров, желающих брать билет) для обоих пунктов одинакова: $\lambda_A=\lambda_B=0,45$ (пассажира в минуту). В сумме они образуют общий поток заявок с интенсивностью $\lambda_A + \lambda_B = 0,9$. Кассир тратит на обслуживание пассажира в среднем 2 минуты. Опыт показывает, что у кассы скапливается очередь, пассажиры жалуются на медленность обслуживания. Поступило рационализаторское предложение: создать две специализированные кассы (по одному окошку в каждой), продающие билеты одна – только в пункт А, другая – только в пункт В. Разумность этого вызывает споры – кое-кто утверждает, что очереди останутся прежними. Требуется проверить полезность предложения расчетом. Так как мы умеем считать характеристики только для простейших СМО, допустим, что все потоки – простейшие.

Итак, рассмотрим два варианта организации продажи билетов – существующий и предлагаемый.

Вариант I (существующий). На двух канальную СМО поступает поток заявок с интенсивностью $\lambda = 0{,}9$; интенсивность потока обслуживаний $\mu = ½= 0{,}5$; приведенная интенсивность $\rho = \lambda/\mu = 1{,}8$. Так как $\rho/n = 1{,}8/2 = 0{,}9 < 1$, то предельные вероятности существуют. Находим по формуле (7.18)

$$p_0 = \left(\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!\,(n-\rho)} \right)^{-1} = \ldots \approx 0{,}0525.$$

Среднее число заявок в очереди находим по формуле (7.21)

$$L_{оч} = \frac{\rho^{n+1} p_0}{(n-1)!\,(n-\rho)^2} = \ldots \approx 7{,}68 \text{ (чел.)}.$$

Среднее время, проводимое заявкой в очереди по первой из формул (7.23)

$$W_{оч} = \frac{1}{\lambda} \cdot L_{оч} = \ldots \approx 8{,}54 \text{ (мин)}.$$

Вариант II (предлагаемый). Надо рассмотреть две одноканальные СМО (Два специализированных окошка). На каждую поступает поток заявок с интенсивностью $\lambda = 0{,}45$; интенсивность потока обслуживаний $\mu = ½= 0{,}5$; приведенная интенсивность $\rho = \lambda/\mu = 0{,}9 < 1$. Следовательно, предельные вероятности существуют. Находим по формуле (7.16) среднюю длину очереди

$$L_{оч} = \frac{\rho^2}{1-\rho} = \ldots \approx 8{,}1.$$

Сравним результаты. Длина очереди не только не уменьшилась, а увеличилась! Время ожидания в очереди по формуле Литтла имеем $W_{оч}=\ldots\approx 18$ (мин). Вот так рационализация: и средняя длина очереди, и среднее время ожидания в ней увеличились.

Почему так произошло? … потому, что в первом варианте меньше средняя доля времени, которую простаивает каждый из двух кассиров: если он не занят обслуживанием пассажира, покупающего билет в пункт А , он может обслужить того, кто покупает билет в пункт В., и наоборот. Во втором варианте такой взаимозаменяемости нет, незанятый кассир сидит, сложа руки.

В данном примере обе СМО работают на пределе своих возможностей. Если чуть-чуть увеличить время обслуживания (т.е. уменьшить интенсивность обслуживания), то СМО перестанут справляться с потоком. «Лишние» простои кассира – это в каком-то смысле уменьшение интенсивности обслуживания.

Правда, здесь есть вопрос: если касса выдают билет только в один пункт, то время обслуживания должно уменьшиться, а мы считали в обоих вариантах его одинаковым. Итак, на сколько должно уменьшиться время обслуживания, чтобы рационализаторское предложение стало выгодным? Это уже задача оптимизации.

ЛИТЕРАТУРА

1. Вагнер, Г. Основы исследования операций / Г. Вагнер. – М.: Мир, 1973. – Т. 3.
2. Вентцель, Е.С. Исследование операций / Е.С. Вентцель. – М.: Просвещение, 1966.
3. Высшая математика для экономистов: в 3 т.: учебник / И.И. Гайшун, С.А. Минюк, Л.И. Шевченко, И.И. Белько, К.К. Кузьмич. – Минск: БГЭУ, 2005. – Т. 2: Теория вероятностей в экономике. Методы оптимизации и экономические модели.
4. Волков, И.К. Исследование операций: учебник для вузов / И.К. Волков, Е.А. Загоруйко. – 2-е изд. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2002.
5. Кузнецов, А.В. Высшая математика: Математическое программирование / А.В. Кузнецов, В.А. Сакович, Н.И. Холод. – Минск: Выш. шк., 2001.
6. Краснопрошин, В.В. Исследование операций: учеб. пособие / В.В. Краснопрошин, Н.А. Лепешинский. – Минск: БГУ, 2013. – 191 с. – (Классическое университетское издание).

Учебное издание

КОМАНДИНА Лариса Владимировна

**ИССЛЕДОВАНИЕ ОПЕРАЦИЙ.
ЭЛЕМЕНТЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ**

Методические рекомендации

Технический редактор *Г.В. Разбоева*
Компьютерный дизайн *Т.Е. Сафранкова*

Подписано в печать .2015. Формат 60x84 $^{1}/_{16}$. Бумага офсетная.
Усл. печ. л. 2,73. Уч.-изд. л. 3,00. Тираж экз. Заказ .

Издатель и полиграфическое исполнение – учреждение образования
«Витебский государственный университет имени П.М. Машерова».

Свидетельство о государственной регистрации в качестве издателя,
изготовителя, распространителя печатных изданий
№ 1/255 от 31.03.2014 г.

Отпечатано на ризографе учреждения образования
«Витебский государственный университет имени П.М. Машерова».
210038, г. Витебск, Московский проспект, 33.