

ГЛАВА 3

ВВЕДЕНИЕ В ТЕОРИЮ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

3.1. Классификация систем массового обслуживания

Исторически теория систем массового обслуживания берет начало от работ, в которых рассматривалась проблема перегрузки телефонных линий, весьма обострившаяся в начале 20-го столетия, в частности работ датского ученого А. Эрланга, относящихся к 1908 – 1922 гг. Термин *теория массового обслуживания* был введен российским математиком А.Я. Хинчиным. В зарубежной литературе обычно используется название *исследование операций* (Operation Research).

Современное общество и современные технологии не могут обойтись без разветвленной сети систем массового обслуживания. Несмотря на их разнообразие, все эти системы имеют общие черты и общие проблемы, наиболее важной из которых является проблема очередей. Источником подобных проблем является случайный характер происходящих в них явлений. Например, случаен поток вызовов, поступающих на телефонную станцию, случайна длительность телефонного разговора. Случайности принципиально нельзя устранить, но их можно должным образом учесть для рациональной организации систем массового обслуживания.

Любая система массового обслуживания предназначена для выполнения некоторого потока заявок. В роли заявки может выступать появление клиента в банке или билетной кассе, возникновение неисправности в устройстве, появление частицы в счетчике Гейгера-Мюллера и т.д. Обслуживающие единицы (например, работающие кассы в зале или операторы в банке) называются каналами обслуживания.

Одна из возможных классификаций систем массового обслуживания (СМО) показана на рис.3.1. В этой классификации достаточно важным является их разделение на системы с отказами и системы с очередью. В системе с отказами, если все каналы обслуживания заняты, заявка получает отказ (например, короткие гудки в телефонной трубке) и выходит из игры. В системе с очередью при невозможности обслуживания заявка встает в очередь и ожидает, пока не освободится какой-либо канал обслуживания (например, в трубке раздается «ждите ответа»). Очереди тоже бывают разными. В частности, они могут быть неограниченными (заявка рано или поздно будет обслужена, это

только вопрос времени, но отказа не будет) или ограниченными, в которых после появления в очереди определенного числа заявок следующая заявка получает отказ. Очереди с относительным приоритетом предполагают обслуживание приоритетной заявки вне очереди *после завершения обслуживания текущей заявки*, а очереди с абсолютным приоритетом даже *прерывают обслуживание текущей заявки* (вероятно, вы сами можете привести примеры из жизни, замечательно обобщенные фразой Дж. Оруэлла «Все животные равны, но некоторые животные *равнее* других»).

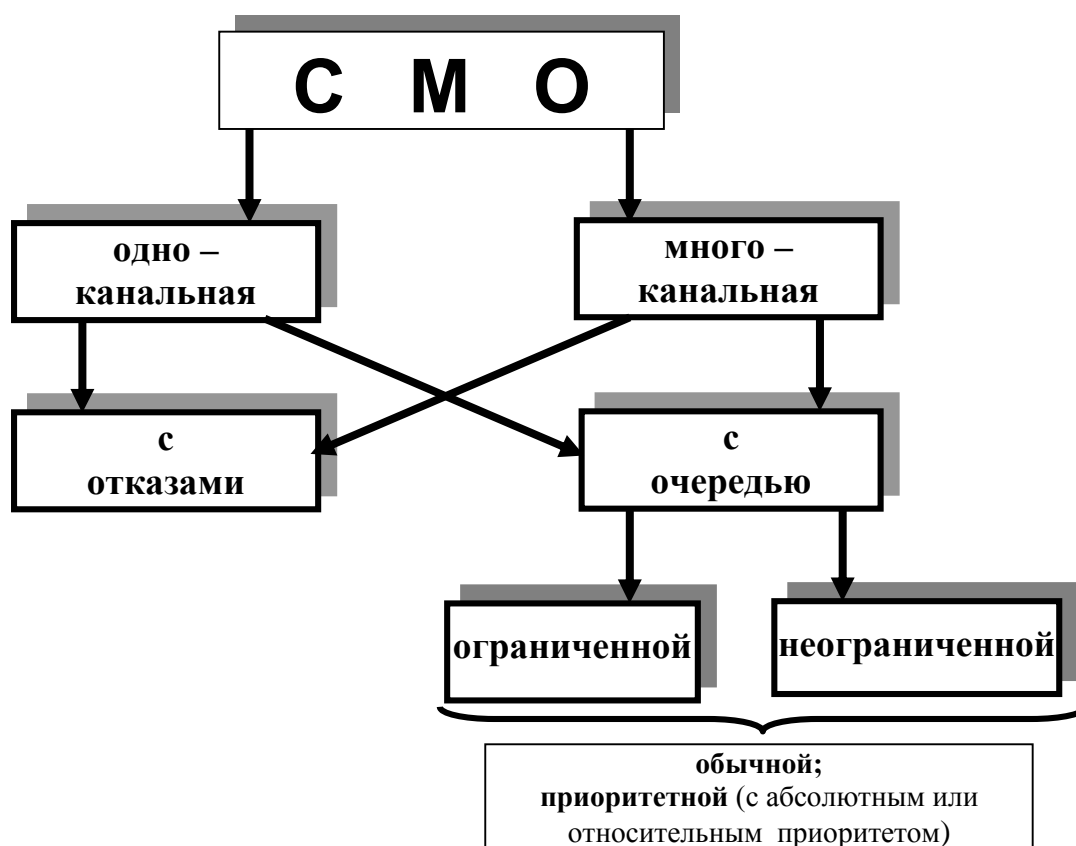


Рис.3.1. Классификация систем массового обслуживания

Системы массового обслуживания всегда являются многокритериальными: они характеризуются *набором* показателей эффективности. Это среднее число заявок, которые система обслуживает в единицу времени, среднее число занятых каналов обслуживания, среднее число заявок, находящихся в очереди, и среднее время стояния в очереди, доля заявок, получающих отказ, и доля заявок, начинающихся обслуживаться немедленно, и др. Моделирование таких систем методами теории цепей Маркова позволяет прогнозировать эти показатели и выявлять пути их улучшения.

3.2. Основные понятия и простейшие модели

Напомним, что случайным процессом называют процесс перехода системы из одних состояний в другие, протекающие случайным образом. Мы рассматриваем процессы с дискретными случайными состояниями, когда система характеризуется набором этих состояний, то есть дискретным пространством состояний.

Рассмотрим основные понятия на примере работы некоторого устройства (рис.3.2), которое может находиться в двух состояниях: работает (состояние 1) и ремонтируется (состояние 2).

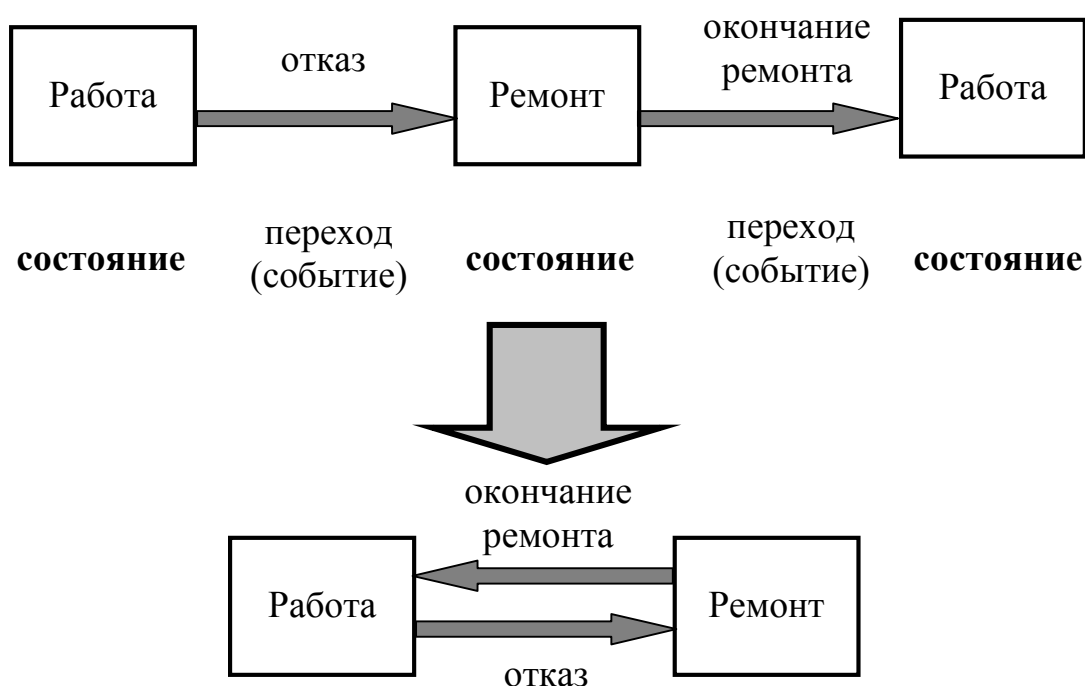


Рис.3.2. Состояния и переходы в эксплуатации устройства

Переход из состояния 1 в состояние 2 связан с событием *отказ* (устройство вышло из строя). Будем считать, что после этого оно мгновенно направляется в ремонт, то есть переходит в состояние 2. После окончания ремонта оно мгновенно возвращается в работу (переход из состояния 2 в состояние 1 с событием *окончание ремонта*). Поскольку состояния «работает» до и после ремонта идентичны, то граф состояний можно представить так, как это показано на рис.3.2 внизу с двумя противоположными стрелками переходов. Очевидно, что оба события происходят в случайные моменты времени, которые для одного конкретного устройства спрогнозировать невозможно.

Предположим, что однородные события следуют один за другим в случайные моменты времени. В этом случае можно говорить о потоке событий. Это может быть поток отказов устройства, поток заказов такси по телефону, поток включений бытовых электроприборов и т.п. Пусть, например, диспетчер таксомоторного парка фиксирует моменты поступления телефонных заказов между моментами времени t_1 и t_2 (рис.3.3). Обозначив эти моменты в виде точек на оси времени, мы получим показанную на рис.3.3 картину, приведенную для 4-х последовательных дней. Каждая из картин условно изображает одну реализацию потока заказов для выбранного промежутка времени. Моменты заказов в каждой реализации случайны, поток событий проявляет статистическую устойчивость: полное количество событий в данном промежутке времени слабо меняется день ото дня (8, 7, 9, 7 соответственно).

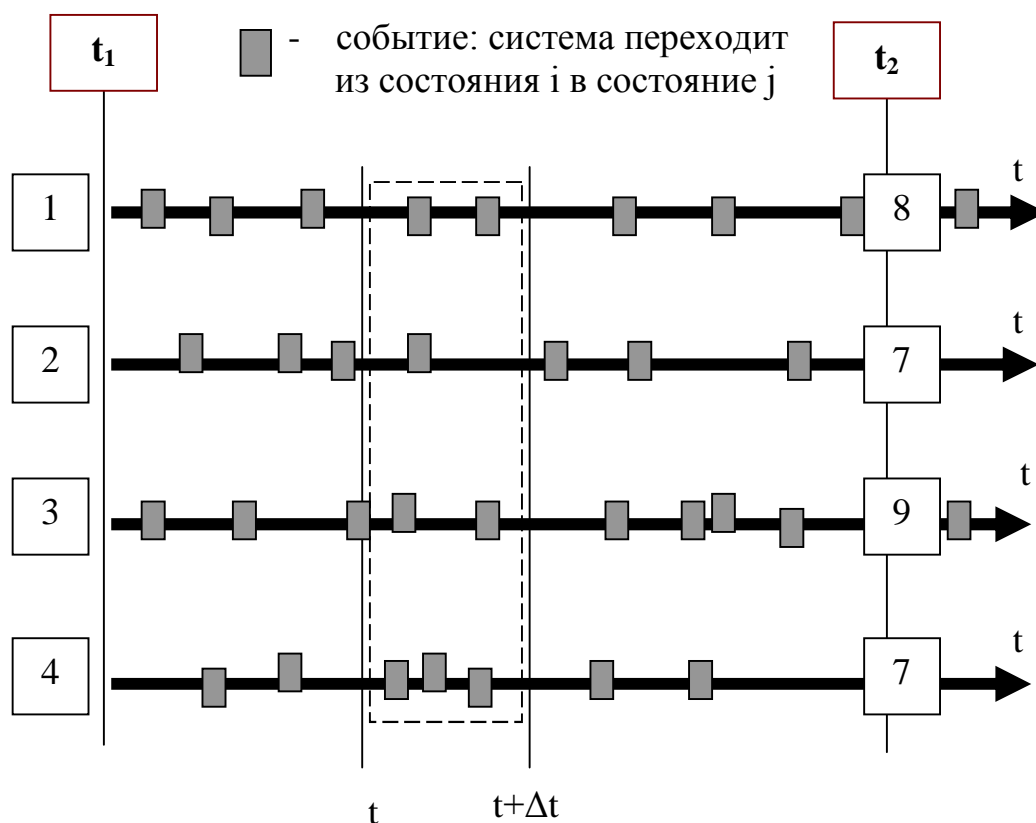


Рис.3.3. К определению потока событий и его интенсивности

При появлении события система переходит из состояния i в состояние j . Нет смысла говорить о вероятности перехода в конкретный момент времени: она всегда равна нулю в силу непрерывности времени. Будем говорить о вероятности перехода $P_{ij}(t; \Delta t)$ не в момент времени t , а на промежутке времени от t до $t+\Delta t$ (рис.3.3). Очевидно, что эта

вероятность зависит от величины Δt : чем больше Δt , тем больше вероятность того, что событие наступит, причем в первом приближении просматривается прямая пропорциональность. Если разделить $P_{ij}(t; \Delta t)$ на Δt и рассмотреть предел этого отношения при $\Delta t \rightarrow 0$, то мы приходим к понятию *плотности вероятности перехода* или *интенсивности потока событий*

$$\lambda_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(t; \Delta t)}{\Delta t}. \quad (3.1)$$

В общем случае величина λ_{ij} зависит от времени (достаточно вспомнить о так называемых часах пик). Однако в дальнейшем мы будем рассматривать для простоты только те случаи, когда λ_{ij} можно считать постоянной, то есть *стационарные случайные процессы*.

В рассмотренном на рис.3.2 примере необходимо оперировать двумя плотностями потоков событий: плотностью потока отказов ν и плотностью потоков окончания ремонта μ . Их размерность – [1/время]. Очевидно, что величина $1/\nu$, имеющая размерность времени, характеризует время чередования отказов (продолжительность безотказной работы), а величина $1/\mu$ – время чередования выходов из ремонта (продолжительность ремонта). Также очевидно, что вероятность перехода из одного состояния в другое зависит от Δt и определяется как $\nu \Delta t$. Однако в дальнейшем мы будем рассматривать процессы, происходящие в некоторую единицу времени, и считать вероятность перехода как ν или μ .

Чтобы не возвращаться к предельно простому примеру, изображенному на рис.3.4, начнем моделирование с более сложной системы, которая состоит из двух устройств 1 и 2, производящих одну и ту же продукцию. Пусть первое *работающее* устройство приносит доход D_1 в единицу времени, а второе – D_2 . Устройства в процессе работы могут выйти из строя (отказаться). Граф (пространство состояний) этой системы показан на рис.3.2. Она может находиться в состоянии 1 (его вероятность обозначена как S_1 и заранее неизвестна, как и вероятности остальных состояний): оба устройства работают; в состоянии 2: второе устройство работает, а первое ремонтируется; в состоянии 3: первое работает, а второе ремонтируется; в состоянии 4: оба устройства в ремонте (не работают). Возможные переходы между этими состояниями показаны стрелками и связаны с событиями отказа каждого устройства и окончания ремонта каждого устройства. Интенсивно-

сти потоков этих событий показаны буквами около соответствующих стрелок. Здесь не рассматривается переход из состояния 1 в 4, так как вероятность одномоментного отказа двух устройств формально равна нулю, а практически просто очень мала.

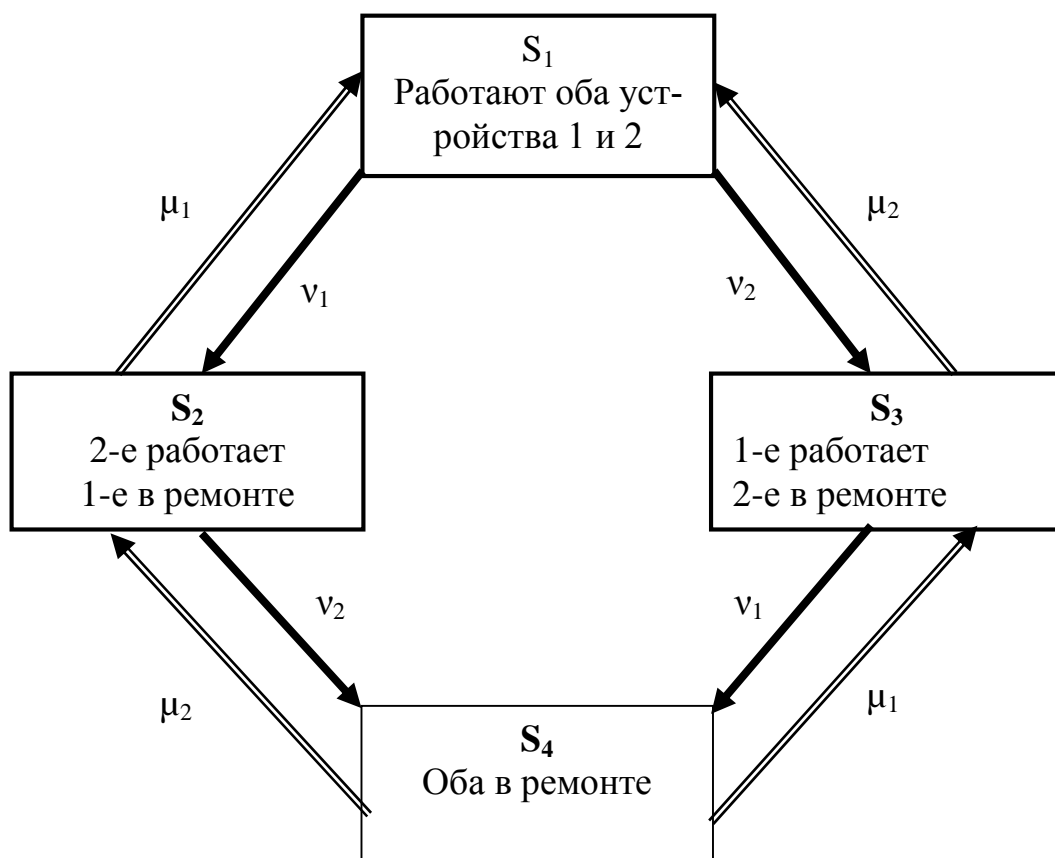


Рис.3.4. Пространство состояний и возможные переходы в системе с двумя устройствами

Несмотря на то, что на графе состояний события не вытянулись в линию, как в предыдущей главе, очевидно, что они образуют цепь, и, используя приобретенный опыт, нетрудно построить матрицу переходных вероятностей для этой цепи.

$$\mathbf{P} = \begin{bmatrix} 1 - v_1 - v_2 & \mu_1 & \mu_2 & 0 \\ v_1 & 1 - \mu_1 - v_2 & 0 & \mu_2 \\ v_2 & 0 & 1 - \mu_2 - v_1 & \mu_1 \\ 0 & v_2 & v_1 & 1 - \mu_1 - \mu_2 \end{bmatrix}. \quad (3.2)$$

Будем интересоваться только установившимся (асимптотическим) состоянием системы \mathbf{S}^∞ , которое, как было показано в разд.2.2, не

зависит от начального вектора состояния и определяется как собственный вектор матрицы \mathbf{P} , соответствующий ее собственному числу, равному единице. Найдя этот собственный вектор для заданных v и μ , получим вероятности всех состояний S_1^∞ , S_2^∞ , S_3^∞ и S_4^∞ . Зная эти вероятности, легко рассчитать доход от эксплуатации обоих устройств с учетом их ремонтов:

$$D = (D_1 + D_2)S_1^\infty + D_2S_2^\infty + D_1S_3^\infty. \quad (3.3)$$

Рассмотрим пример расчета состояний этой системы. Пусть $v_1=0,05$; $v_2=0,1$; $\mu_1=0,1$; $\mu_2=0,15$. Подстановка этих значений в матрицу (3.2) и определение ее собственного вектора для собственного числа, равного единице дает следующие значения вероятностей состояний: $S_1^\infty=0,4$; $S_2^\infty=0,2$; $S_3^\infty=0,2667$ и $S_4^\infty=0,1333$. В частности, оба устройства работают одновременно с вероятностью 0,4, а оба не работают с вероятностью 0,1333. Допустим, что первое работающее устройство приносит в единицу времени доход $D_1=5$ условных единиц, а второе – доход $D_2=10$ единиц. Расчет по формуле (3.3) дает, что реальный доход с учетом отказа устройств составляет 9,33 единицы.

Вы можете самостоятельно выполнить эти расчеты и оценить влияние параметров v и μ на вероятности состояний системы по приведенной в приложении программе MCh8. Советуем непосредственными расчетами убедиться, что на результат влияет только соотношение разных v и μ , а не их абсолютные значения. В частности, если все принятые в примере расчета v и μ разделить на одно и то же число, распределение вероятностей не изменится. Естественно, эти параметры можно и умножить на некоторое число, но при этом необходимо помнить, что если хотя бы один элемент главной диагонали матрицы \mathbf{P} станет меньше нуля, нарушится вычислительная устойчивость.

Пусть теперь мы располагаем некоторыми инвестициями в улучшение ремонта, позволяющими увеличить одно из μ в два раза. Какое из них целесообразно увеличить? Если в расчет ввести $\mu_1=0,2$ (модернизируем ремонт первого устройства), то расчетный доход возрастет до 10 условных единиц, если ввести $\mu_2=0,3$ (модернизируем ремонт второго устройства), то – до 10,83 единиц. Следовательно, рационально вложить деньги в модернизацию ремонта второго устройства.

3.3. Многоканальная СМО с отказами

Перейдем к рассмотрению моделирования собственно систем массового обслуживания. В качестве первого примера проанализируем многоканальную СМО с отказами, начав с трехканальной. Ее граф состояний показан на рис.3.5, где сами состояния прокомментированы внутри ячеек цепи. Их число на единицу больше числа каналов.

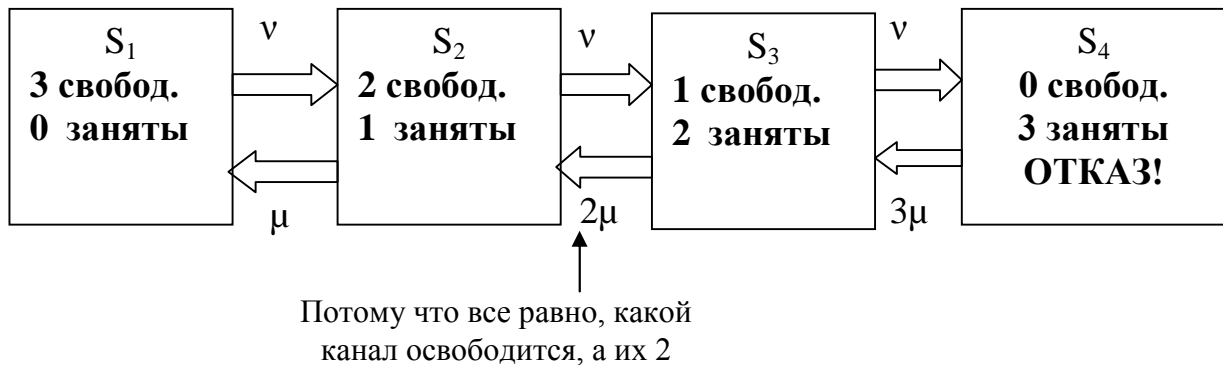


Рис.3.5. Схема трехканальной СМО с отказами

В этой цепи v означает интенсивность потока заявок на обслуживание ($1/v$ – время чередования заявок), а μ – интенсивность потока обслуживания ($1/\mu$ – время обслуживания одной заявки).

Матрица переходных вероятностей для этой цепи имеет вид

$$\mathbf{P} = \begin{bmatrix} 1-v & \mu & 0 & 0 \\ v & 1-v-\mu & 2\mu & 0 \\ 0 & v & 1-v-2\mu & 3\mu \\ 0 & 0 & v & 1-v-3\mu \end{bmatrix}, \quad (3.4)$$

который легко распространить на произвольное число каналов N . Вектор установившихся вероятностей состояний S^∞ есть собственный вектор этой матрицы, соответствующий собственному числу, равному единице.

Пусть все вероятности S_1^∞ , S_2^∞ , S_3^∞ и S_4^∞ определены. Рассмотрим рассчитываемые по ним основные характеристики эффективности работы системы. Прежде всего, величина $S_R = S_4$ (S_{N+1} в общем случае) представляет собой вероятность отказа в обслуживании – одну из важнейших характеристик системы, которую желательно иметь как

можно меньше. Соответственно, $S_S=1-S_R$ представляет собой вероятность того, что поступившая заявка будет принята к обслуживанию. Эту вероятность желательно иметь как можно больше.

Произведение $S_S v$ дает интенсивность потока обслуженных заявок. Так как каждый занятый канал обслуживает в единицу времени μ заявок, то, разделив интенсивность потока обслуженных заявок на μ , получим среднее число занятых каналов $S_S v / \mu$ или $(1-S_{N+1})v/\mu$. Соответственно доля занятых каналов получается делением числа занятых каналов на их общее число N .

С помощью приведенной в приложении программы MCh9 выполним некоторые расчеты многоканальной СМО с отказами.

Пусть $v=0,15$ 1/мин (среднее время чередования заявок 6,67 мин) и $\mu=0,05$ 1/мин (среднее время обслуживания одной заявки 20 мин). На рис.3.6 показано распределение вероятностей состояния для различного числа каналов, где на графиках темным цветом выделена вероятность отказа.

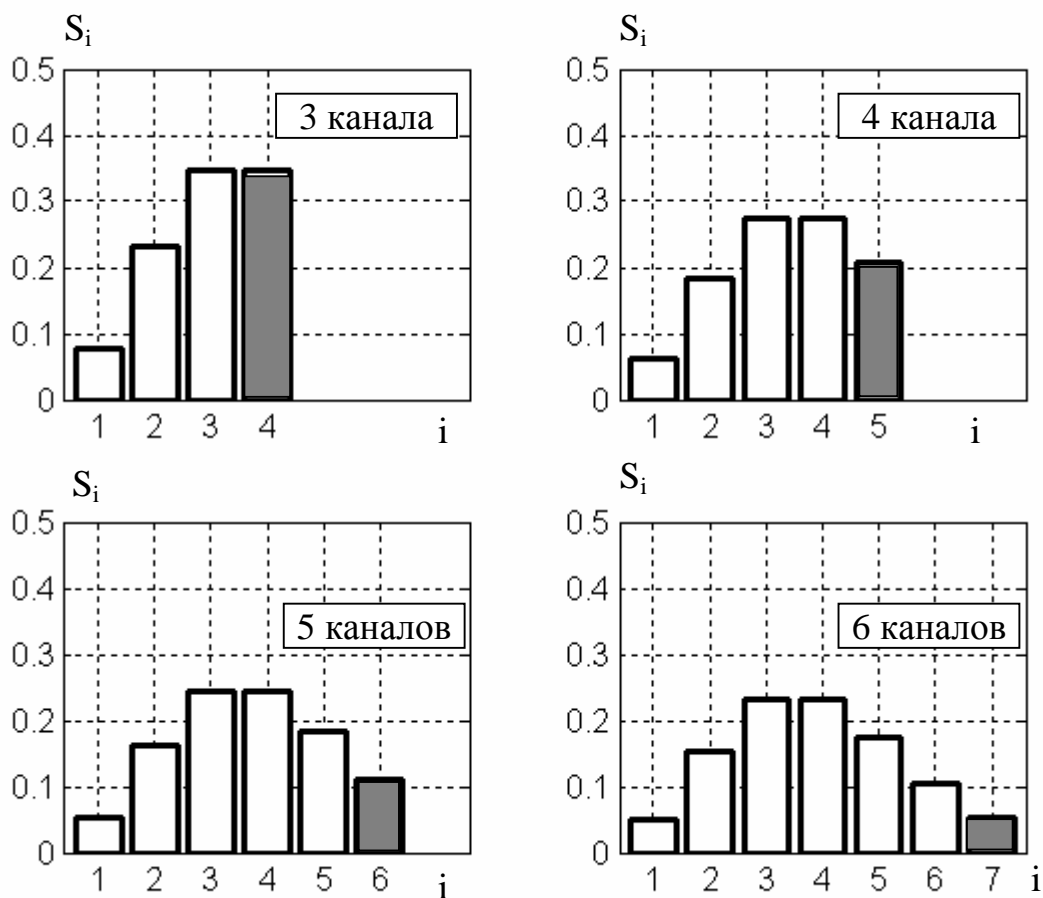


Рис.3.6. Распределение установившихся вероятностей состояния при различном числе каналов ($v=0,15$; $\mu=0,05$)

Зависимость вероятности отказа от числа каналов показана на рис.3.7, а основные характеристики СМО приведены в табл.3.1.

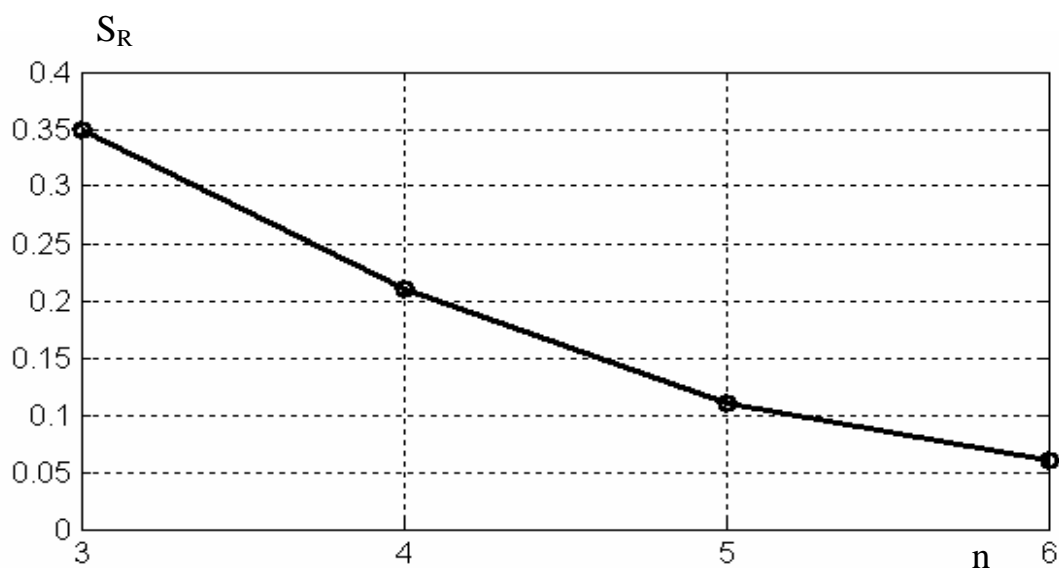


Рис.3.7. Зависимость вероятности отказа от числа каналов ($v=0,15$; $\mu=0,05$)

Таблица 3.1

Основные характеристики многоканальной СМО с отказами при различном числе каналов

Число каналов	3	4	6
Вероятность отказа	0,35	0,21	0,06
Вероятность обслуживания	0,65	0,79	0,94
Доля занятых каналов	0,65	0,60	0,47

Формально увеличение вероятности обслуживания достигается увеличением числа каналов, но при этом снижается доля занятых каналов. Например, при весьма высокой вероятности обслуживания 0,94 при шести каналах более половины каналов в среднем оказываются незанятыми. Здесь необходим более детальный анализ с учетом стоимости каждого канала. Вероятность обслуживания растет и при увеличении μ (уменьшении времени обслуживания; точнее, при уменьшении отношения v/μ , но величина v считается заданной), однако и это требует повышения технической оснащенности каналов и/или повышения квалификации операторов, что также требует дополнительных инвестиций. Тем не менее описанная модель может служить базой для технико-экономических оценок.

3.4. Одноканальная СМО с очередью

Рассмотрим одноканальную систему массового обслуживания с ограниченной очередью. Ограничение осуществляется по числу заявок, стоящих в очереди, то есть по числу m мест в очереди. Примером такой системы может служить автозаправочная станция, имеющая одну колонку (один канал обслуживания) и площадку, на которой одновременно могут находиться не более m машин. Если все места на площадке заняты, очередной автомобиль, подъезжающий к станции, вынужден проехать мимо.

Граф состояний такой системы с очередью на 3 места показан на рис.3.8. В такой системе число возможных состояний равно $m+2$, то есть 5 в рассматриваемом примере. Характер состояний описан внутри ячеек на рисунке. Как и ранее, здесь ν – интенсивность потока поступающих заявок (подъезжающих автомобилей), а μ – интенсивность потока обслуживания. Обратные им величины представляют среднее время между появлениями автомобилей и среднее время заправки автомобиля, соответственно.

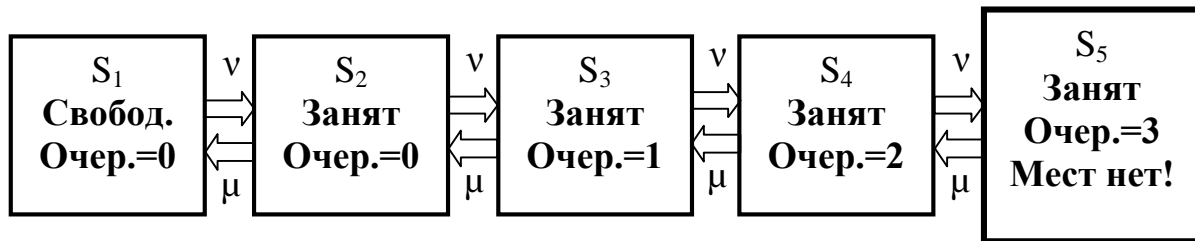


Рис.3.8. Схема одноканальной СМО с ограниченной очередью

Матрица переходных вероятностей такой системы имеет вид

$$\mathbf{P} = \begin{bmatrix} 1-\nu & \mu & 0 & 0 & 0 \\ \nu & 1-\nu-\mu & \mu & 0 & 0 \\ 0 & \nu & 1-\nu-\mu & \mu & 0 \\ 0 & 0 & \nu & 1-\nu-\mu & \mu \\ 0 & 0 & 0 & \nu & 1-\mu \end{bmatrix}, \quad (3.5)$$

который напоминает вид переходной матрицы для конвективной диффузии (2.14) при $\nu_f=0$.

Пусть решение как собственный вектор матрицы \mathbf{P} для собственного числа *единица* получено и мы знаем вероятности всех состояний S_1^∞ , S_2^∞ , S_3^∞ , S_4^∞ и S_5^∞ . Рассчитаем некоторые статистические характеристики образовавшейся в этой системе очереди.

Число заявок, стоящих в очереди, связано с состояниями 3, 4, ... m. С вероятностью S_3^∞ в очереди стоит одна заявка, с вероятностью S_4^∞ - две и т.д. Очевидно, что среднее число заявок в очереди определится следующим образом:

$$N_L = \sum_{k=1}^m k S_{k+2}^\infty. \quad (3.6)$$

При плотности потока заявок ν среднее время стояния заявки в очереди составит

$$T_L = \frac{N_L}{\nu} = \frac{\sum_{k=1}^m k S_{k+2}^\infty}{\nu}. \quad (3.7)$$

Как и ранее, распределение вероятностей состояния зависит только от отношения ν/μ , но величина T_L уже зависит от плотности потока заявок.

На рис.3.9 показаны результаты расчета состояний и некоторые характеристики одноканальной СМО с ограниченной очередью.

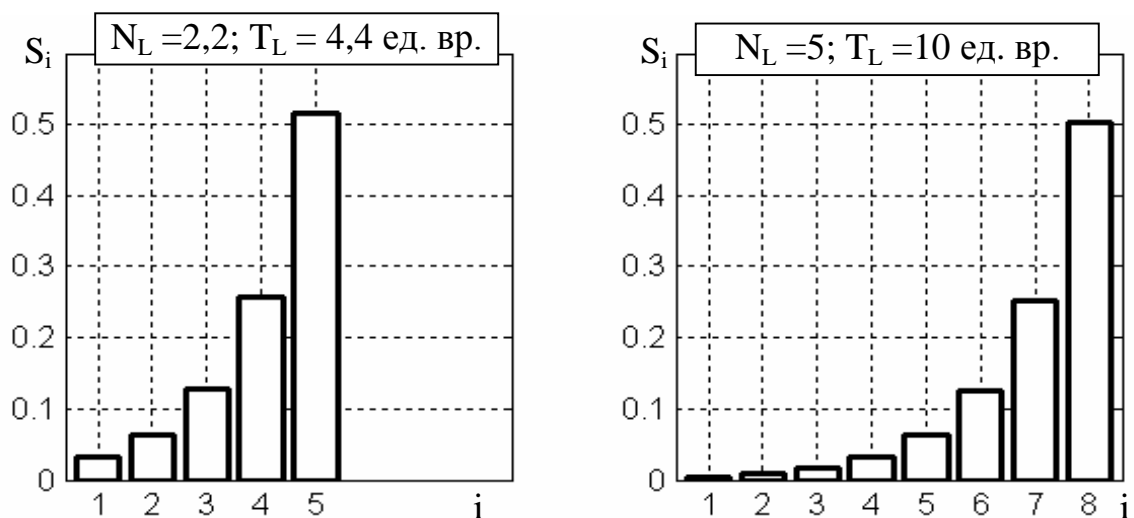


Рис.3.9. Распределение установившихся вероятностей состояния и характеристики очереди для ограниченной очереди на 3 (слева) и 6 (справа) мест

Расчеты выполнены для $\nu=0,5$ 1/мин (время чередования заявок – 2 мин) и $\mu=0,25$ 1/мин (время обслуживания – 4 мин), то есть для отношения $\nu/\mu=2$. Слева показаны результаты для очереди в 3 места. Вероятность отказа в обслуживании равна 0,516 (вероятность обслуживания – 0,484), среднее число заявок в очереди – 2,2, а среднее время стояния в очереди (ожидания обслуживания) – 4,4 мин. Что произойдет, если удвоить число мест в очереди? Оказывается, что вероятность отказа снизится всего до 0,502, то есть почти не изменится. Зато число заявок в очереди возрастет до 5, а среднее время обслуживания – до 10 мин. Поэтому увеличение мест в очереди совсем неэффективно.

На рис.3.10 показано, как влияет на вероятность отказа изменение отношения ν/μ , уменьшение которого при заданном ν соответствует уменьшению продолжительности обслуживания заявки. Здесь зависимость гораздо более сильная, и уменьшение этого отношения является эффективным средством улучшения обслуживания.

Компьютерное моделирование процесса в одноканальной СМО с очередью вы можете выполнить с помощью приведенной в приложении программы MCh10.

Можно отметить, что для многих приведенных выше примеров существуют аналитические решения, позволяющие рассчитывать основные характеристики систем по готовым формулам. Здесь эти решения не приводятся умышленно. Описанный компьютерный алгоритм расчета стационарного распределения вероятностей состояний в системах является универсальным для всех задач, в том числе и для тех, у которых аналитических решений нет. При построении моделей СМО необходимо лишь изобразить граф возможных состояний системы со стрелками допустимых переходов между состояниями и построить соответствующую ему матрицу переходных вероятностей. Всю остальную работу компьютер берет на себя, и вам лишь необходимо проанализировать полученные результаты и исследовать влияние на них параметров системы.

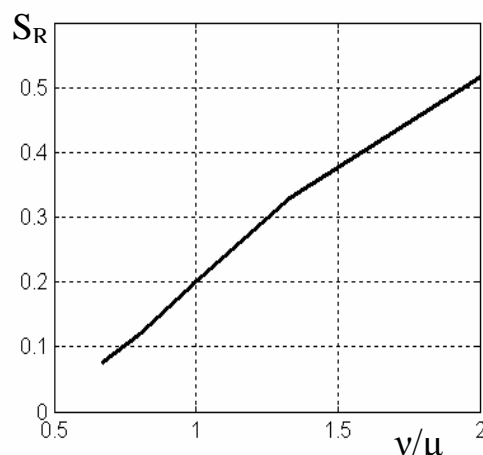


Рис.3.10. Зависимость вероятности отказа в СМО с очередью в 3 места от ν/μ

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. **Тарасов, Л.В.** Мир, построенный на вероятности/ Л.В. Тарасов. – М.: Просвещение, 1984. – 191 с.
2. **Андреев, В.Н.** Эти замечательные цепи/ В.Н. Андреев, А.Я. Иоффе. – М.: Знание, 1987. – 176 с.
3. **Баруча-Рид, А. Т.** Элементы теории марковских процессов и их приложения/ А.Т. Баруча-Рид. – М.: Наука, 1969. – 512 с.
4. **Хант, Дж. А.** Марковские процессы и потенциалы/ Дж.А. Хант. – М.: Изд-во ин. лит-ры, 1962. – 283 с.
5. **Winston, W.L.** Operation Research: Applications and Algorithms. The 3rd Edition, Duxbury Press, Wadsworth Publishing Company, 1993, 1024 p.
6. **Баранцева, Е.А.** Смешивание сыпучих материалов: моделирование, оптимизация, расчет/ Е.А. Баранцева, В.Е. Мизонов, Ю.В. Хохлова; Федеральное агентство по образованию, ГОУВПО «Ивановский государственный энергетический университет им. В.И. Ленина». – Иваново, 2008. – 116 с.
7. **Зайцев, В.А.** Применение теории цепей Маркова к моделированию, расчету и оптимизации процессов тепломассопереноса в промышленных аппаратах/ В.А. Зайцев, В.Е. Мизонов, В.Ю. Волынский// Федеральное агентство по образованию, ГОУВПО «Ивановский государственный химико-технологический университет». – Иваново, 2008. – 268 с.
8. **Мизонов, В.Е.** Моделирование и оптимизация теплового состояния в секционированных объемах с внутренними источниками теплоты/ В.Е. Мизонов, Н.Н. Елин, Е.А. Баранцева; Федеральное агентство по образованию, ГОУВПО «Ивановский государственный энергетический университет». – Иваново, 2010. – 128 с.
9. **Мизонов, В.Е.** Моделирование, расчет и оптимизация тепломассообменных процессов в текстильной промышленности/ В.Е. Мизонов, В.А. Зайцев, В.Ю. Волынский, С.П. Бобков; Федеральное агентство по образованию, ГОУВПО «Ивановский государственный энергетический университет». – Иваново, 2010. – 204 с.
10. **Болотов, И.А.** Математическое моделирование процессов тепло- и массопереноса во вращающихся телах/ И.А. Болотов, В.Е. Мизонов, В.А. Зайцев, П.В. Жуков; Федеральное агентство по образованию, ГОУВПО «Ивановский государственный энергетический университет», «Ивановский государственный химико-технологический университет». – Иваново, 2010. – 56 с.