

Раздел 4. АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

«Всякое уравнение длиной более двух дюймов, скорее всего, неверно!» (Автор неизвестен)

4.1. Одноканальные СМО с однородным потоком заявок

Рассмотрим одноканальную СМО с однородным потоком заявок при следующих предположениях (рис.4.1):

1) СМО содержит *один обслуживающий прибор*, в котором в каждый момент времени может обслуживаться только одна заявка;

2) перед прибором имеется накопитель **Н** *неограниченной ёмкости*, что означает отсутствие отказов поступающим заявкам при их постановке в очередь **О**, то есть любая поступающая заявка всегда найдет в накопителе место для ожидания не зависимо от того, сколько заявок уже находится в очереди;

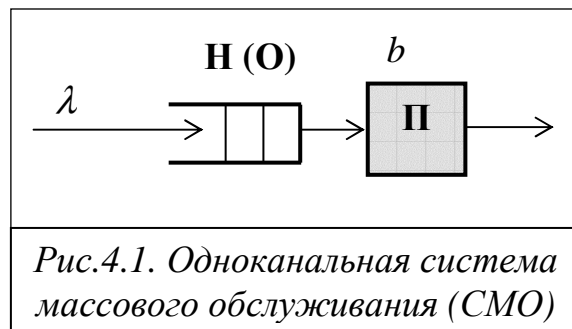


Рис.4.1. Одноканальная система массового обслуживания (СМО)

3) заявки поступают в СМО с интенсивностью λ ;

4) средняя длительность обслуживания одной заявки в приборе равна b , причем длительности обслуживания разных заявок не зависят друг от друга;

5) обслуживающий прибор не простаивает, если в системе (накопителе) имеется хотя бы одна заявка, причем после завершения обслуживания очередной заявки мгновенно из накопителя выбирается следующая заявка;

6) заявки из накопителя выбираются в соответствии с беспriorитетной дисциплиной обслуживания в порядке поступления (ОПП) по правилу «первым пришел – первым обслужен» (FIFO – First In First Out).

7) в системе существует стационарный режим, предполагающий отсутствие перегрузок, то есть нагрузка и, следовательно, загрузка системы меньше 1: $\rho = \lambda b < 1$.

В качестве расчётной характеристики обслуживания заявок в СМО будем использовать среднее время ожидания заявок. Значения остальных характеристик функционирования СМО легко могут быть рассчитаны с использованием приведенных в разделе 3 фундаментальных соотношений (3.13) – (3.15).

Рассмотрим четыре модели СМО с однородным потоком заявок: экспоненциальную СМО М/М/1 и три неэкспоненциальные СМО типа М/Г/1, Г/М/1, Г/Г/1.

4.1.1. Характеристики экспоненциальной СМО М/М/1

Пусть заявки, поступающие в одноканальную СМО, образуют *простейший* поток с интенсивностью λ , а длительность τ_b обслуживания заявок распределена по *экспоненциальному* закону со средним значением b , причём $\rho = \lambda b < 1$, то есть система работает в установившемся режиме. Такая СМО с однородным потоком заявок называется *экспоненциальной*.

С использованием метода средних значений [2] или марковской модели (см. п. 5.4.4) можно получить следующие выражения для расчета средних значений:

- времени ожидания заявок

$$w = \frac{\rho b}{1 - \rho}; \quad (4.1)$$

- времени пребывания заявок

$$u = w + b = \frac{b}{1 - \rho}; \quad (4.2)$$

- длины очереди заявок

$$l = \lambda w = \frac{\rho^2}{1 - \rho};$$

- числа заявок в системе (в очереди и на обслуживании)

$$m = \lambda u = \frac{\rho}{1 - \rho}.$$

Из последнего выражения вытекает, что среднее число заявок в системе $m = l + \rho$, где второе слагаемое ρ определяет среднее число заявок, находящихся на обслуживании в приборе. Кроме того, сравнивая выражения (4.1) и (4.2) получим, что $u = \rho w$.

4.1.2. Характеристики неэкспоненциальной СМО М/G/1

Пусть заявки, поступающие в одноканальную СМО, образуют простейший поток с интенсивностью λ , а длительность τ_b обслуживания заявок распределена по произвольному закону $B(\tau)$ со средним значением b и коэффициентом вариации ν_b .

С использованием метода средних значений можно показать, что среднее время ожидания заявок определяется по формуле Поллачека-Хинчина [2]:

$$w = \frac{\lambda b^2 (1 + \nu_b^2)}{2(1 - \rho)}, \quad (4.3)$$

где $\rho = \lambda b < 1$ – загрузка системы.

Среднее время пребывания заявок в системе:

$$u = w + b = \frac{\lambda b^2 (1 + \nu_b^2)}{2(1 - \rho)} + b.$$

Следует отметить интересную особенность представленных выражений, а именно: средние значения характеристик обслуживания заявок зависят только от двух первых моментов длительности обслуживания заявок и не зависят от моментов более высокого порядка. Другими словами, для того чтобы рассчитать средние характеристики обслуживания, не обязательно знать закон распределения длительности обслуживания заявок – достаточно знать только два первых момента распределения. Можно показать, что для расчета вторых моментов характеристик обслуживания заявок достаточно задать три первых момента длительности обслуживания и т.д. Резюмируя, можно утверждать, что *для СМО с простейшим потоком заявок для расчёта первых k моментов характеристик обслуживания необходимо задать $(k + 1)$ моментов длительности обслуживания заявок.*

4.1.3. Характеристики неэкспоненциальной СМО G/M/1

Пусть в одноканальную СМО с интенсивностью λ поступает случайный поток заявок произвольного вида, задаваемый функцией распределения интервалов между заявками $A(\tau)$, а длительность τ_b обслуживания заявок распределена по экспоненциальному закону $B(\tau)$ со средним значением b (интенсивностью $\mu = 1/b$).

СМО G/M/1 является симметричной по отношению к СМО M/G/1, рассмотренной в предыдущем пункте. Однако получение конечного результата в виде аналитического выражения для расчёта среднего времени ожидания, по аналогии с предыдущей моделью, в общем случае, оказывается невозможным. Это обусловлено тем, что среднее время ожидания, впрочем, как и другие числовые моменты, зависит не только от двух первых моментов интервалов между поступающими заявками, но и от моментов более высокого порядка, т.е. от закона распределения интервалов между заявками.

Среднее время ожидания заявок в очереди может быть рассчитано следующим образом [9]:

$$w = \zeta b / (1 - \zeta), \quad (4.4)$$

где ζ – единственный в области $0 < \zeta < 1$ корень уравнения

$$\zeta = A^*(\mu - \mu\zeta). \quad (4.5)$$

Здесь $A^*(s)$ – преобразование Лапласа плотности распределения $a(\tau)$ интервалов между поступающими в систему заявками:

$$A^*(s) = \int_0^{\infty} e^{-s\tau} a(\tau) d\tau \quad (s \geq 0).$$

Основная сложность при исследовании СМО G/M/1 заключается в том, что уравнение (4.5) для нахождения ζ , в общем случае, является трансцендентным, и невозможно получить выражение для ζ в явном виде. Однако в каждом конкретном случае корень уравнения (4.5) может быть найден с использованием численных методов.

Как сказано выше, средние значения характеристик обслуживания заявок зависят не только от двух первых моментов интервалов между поступающими заявками, но и от моментов более высокого порядка, причем степень влияния соответствующих моментов убывает с увеличением порядка моментов. Другими словами, влияние моментов 4-го порядка менее существенно, чем моментов 3-го порядка и т.д.

Пример 4.1. Проиллюстрируем применение описанного метода расчета к рассмотренной выше СМО M/M/1 с простейшим потоком заявок.

В простейшем потоке интервалы времени между последовательными заявками распределены по экспоненциальному закону, преобразование Лапласа которого имеет вид: $A(s) = \frac{\lambda}{\lambda + s}$.

Тогда уравнение (4.5) примет вид:

$$\zeta = \frac{\lambda}{\lambda + \mu - \mu \zeta}.$$

После некоторых преобразований получим квадратное уравнение:

$$\mu \zeta^2 - (\lambda + \mu) \zeta + \lambda = 0.$$

Разделив левую и правую часть этого уравнения на μ , получим:

$$\zeta^2 - (1 + \rho) \zeta + \rho = 0.$$

Из двух корней $\zeta_1 = 1$ и $\zeta_2 = \rho$ последнего уравнения условию $0 < \zeta < 1$ удовлетворяет только второй корень. Подставляя его в выражение (4.4) окончательно получим выражение для среднего времени ожидания, совпадающее с известным для СМО M/M/1 выражением (4.1).

4.1.4. Характеристики СМО общего вида G/G/1

Наиболее общим случаем одноканальных систем массового обслуживания являются СМО типа G/G/1, в которую поступает произвольный поток заявок общего вида с функцией распределения интервалов между заявками $A(\tau)$. Длительность обслуживания заявок в приборе распределена по произвольному закону $B(\tau)$.

Расчет таких систем требует задания конкретных законов распределений, что не позволяет получить аналитическое решение в общем виде. Аналитическое решение возможно только для некоторых частных распределений, связанных, например, с экспоненциальным распределением. Для большинства законов распределений интервалов между поступающими в систему заявками и длительностей их

обслуживания в приборе невозможно получить точное решение в аналитической форме.

В то же время, на практике при исследовании реальных систем редко бывают известны законы распределений указанных величин. Обычно при описании процессов поступления заявок в систему и их обслуживания в приборе ограничиваются несколькими моментами соответствующих распределений, чаще всего – двумя первыми моментами, задаваемыми в виде математического ожидания и среднеквадратического отклонения или коэффициента вариации искомой случайной величины. Однако при этом оказывается невозможным получение точного результата. Это обусловлено тем, что в случае произвольного (отличного от простейшего) потока заявок, поступающих в систему, характеристики функционирования СМО, в частности среднее время ожидания, зависят не только от двух первых моментов, но и от моментов более высокого порядка – третьего, четвёртого и т.д. Причём эта зависимость тем меньше, чем выше порядок числового момента. Таким образом, все результаты, полученные в аналитической форме при задании интервалов между поступающими в систему заявками и длительностей их обслуживания в приборе двумя первыми моментами – средними значениями $a = 1/\lambda$ и $b = 1/\mu$ и коэффициентами вариации v_a и v_b , представляют собой приближённые зависимости.

Как показал анализ многочисленных опубликованных результатов, одним из наиболее удачных приближений для расчета среднего времени ожидания в СМО G/G/1 является следующая формула [17]:

$$\tilde{w} \approx \frac{\rho b (v_a^2 + v_b^2)}{2(1 - \rho)} f(v_a), \quad (4.6)$$

где $\rho = \lambda b < 1$ – загрузка системы; λ , v_a – интенсивность потока заявок и коэффициент вариации интервалов между поступающими в систему заявками; b , v_b – среднее значение и коэффициент вариации длительности обслуживания заявок; $f(v_a)$ – корректирующая функция, рассчитываемая в зависимости от значения коэффициента вариации v_a :

$$f(v_a) = \begin{cases} \exp\left[-\frac{2(1-\rho)(1-v_a^2)^2}{3\rho(v_a^2+v_b^2)}\right], & v_a < 1 \\ \exp\left[-(1-\rho)\frac{v_a^2-1}{v_a^2+4v_b^2}\right], & v_a \geq 1. \end{cases}$$

При решении многих практических задач выходящий поток заявок из одной СМО является входящим потоком в другую СМО. В этом случае для расчёта характеристик функционирования второй СМО необходимо знать характер входящего потока, наиболее полно описываемый законом распределения интервалов между последовательными заявками. В то же время, для проведения оценочных расчётов во многих случаях достаточно

знание первых двух моментов этих интервалов: математического ожидания и коэффициента вариации.

Очевидно, что в СМО с накопителем неограниченной ёмкости, работающей без перегрузок, интенсивность выходящего потока заявок равна интенсивности входящего потока, то есть математические ожидания интервалов между последовательными заявками на выходе и входе СМО совпадают.

Можно показать, что для экспоненциальной СМО М/М/1 коэффициент вариации выходящего потока равен единице.

В общем случае для СМО G/G/1 коэффициент вариации выходящего потока заявок может быть рассчитан по следующей приближённой формуле [17]:

$$v_c^2 \approx v_a^2 + 2\rho v_b^2 - 2\rho(1-\rho)\frac{\tilde{w}}{b}. \quad (4.7)$$

4.1.5. Анализ свойств одноканальной СМО

«Если факты не подтверждают теорию, от них надо избавиться» (*Закон Майерса*)

Анализ свойств одноканальной СМО с однородным потоком заявок будем проводить с использованием представленных выше математических моделей в виде формул (4.1 – 4.3), определяющих зависимости характеристик обслуживания заявок от параметров поступления и обслуживания заявок для установившегося (стационарного) режима работы системы.

1. Среднее время ожидания заявок в очереди минимально при постоянной (детерминированной) длительности обслуживания заявок, когда коэффициент вариации длительности обслуживания $v_b = 0$, и увеличивается с ростом коэффициента вариации (дисперсии) длительности обслуживания. Заметим, что зависимость среднего времени ожидания от коэффициента вариации v_b носит нелинейный характер. Так, например, при экспоненциально распределенной длительности обслуживания, когда $v_b = 1$, среднее время ожидания заявок увеличивается в 2 раза, а при $v_b = 2$ – в 5 раз, по сравнению с детерминированным обслуживанием.

2. Среднее время ожидания заявок существенно зависит от нагрузки y (загрузки ρ) системы (рис.4.2). При $y \geq 1$ ($\rho \rightarrow 1$) время ожидания заявок возрастает неограниченно: $w \rightarrow \infty$, т.е. заявки могут ожидать обслуживания сколь угодно долго. Отметим, что увеличение нагрузки может быть обусловлено двумя факторами: увеличением интенсивности поступления заявок в систему или увеличением длительности обслуживания заявок (например, в результате уменьшения скорости работы обслуживающего прибора).

3. Можно показать, что для беспriorитетных дисциплин обслуживания в обратном порядке (ООП) и обслуживания в случайном порядке (ОСП) *средние времена ожидания заявок будут такими же, как и при обслуживании в порядке поступления, но дисперсии времени ожидания будут больше.* Это обусловлено, в частности для дисциплины ООП, тем, что часть заявок, поступивших последними, будут ожидать незначительное время, в то время как заявки, попавшие в начало очереди, могут ожидать обслуживания достаточно долго, то есть увеличивается разброс значений времени ожидания относительно среднего значения.

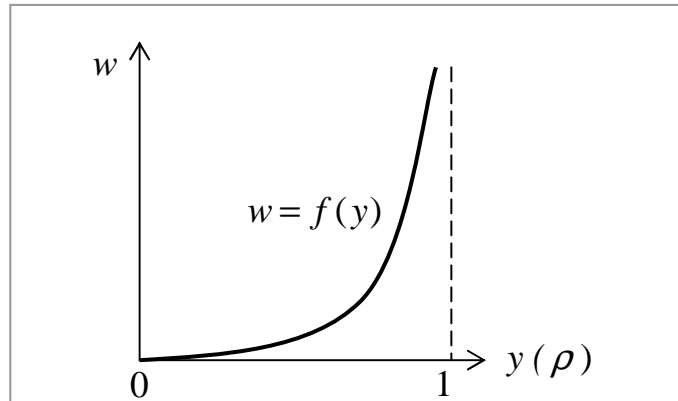


Рис.4.2. Зависимость среднего времени ожидания от нагрузки

4.2. Многоканальные СМО с однородным потоком заявок

«Работая над решением задачи, всегда полезно знать ответ» (*Правило точности*)

Рассмотрим многоканальную СМО с K идентичными обслуживающими приборами и накопителем неограниченной ёмкости, в которую поступают заявки, образующие простейший поток с интенсивностью λ (рис.4.3). Длительность τ_b обслуживания заявок распределена по экспоненциальному закону со средним значением b . Выбор заявок из очереди на обслуживание осуществляется в соответствии с беспriorитетной дисциплиной обслуживания в порядке поступления (ОПП) по правилу «первым пришёл – первым обслужен» (FIFO – First In First Out).

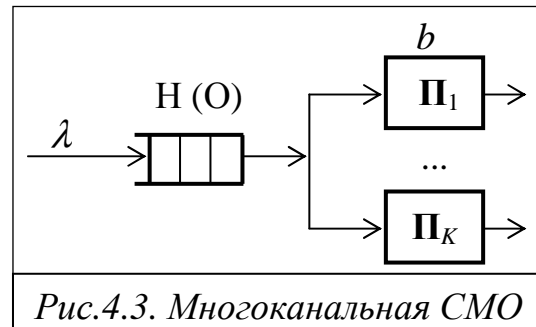


Рис.4.3. Многоканальная СМО

4.2.1. Характеристики многоканальной СМО М/М/К

В качестве основной характеристики функционирования СМО, будем использовать среднее время ожидания w заявок.

Точный метод расчета характеристик обслуживания заявок в многоканальной СМО разработан при следующих предположениях:

- поток заявок – *простейший*;
- длительность обслуживания заявок распределена по *экспоненциальному* закону со средним значением b ;
- все K приборов – *идентичны*, и любая заявка может быть обслужена любым прибором;

- ёмкость накопителя – не ограничена;
- в системе отсутствуют перегрузки, то есть загрузка системы меньше 1: $\rho = \frac{\lambda b}{K} < 1$

При этих предположениях среднее время ожидания заявок определяется следующим образом:

$$w = \frac{Pb}{K(1-\rho)}, \quad (4.8)$$

где P – вероятность того, что все K приборов заняты обслуживанием заявок.

Вероятность P определяется как:

$$P = \frac{(K\rho)^K}{K!(1-\rho)} P_0,$$

где P_0 – вероятность простоя многоканальной СМО, то есть вероятность того, что в системе нет заявок:

$$P_0 = \left[\frac{(K\rho)^K}{K!(1-\rho)} + \sum_{i=0}^{K-1} \frac{(K\rho)^i}{i!} \right]^{-1}.$$

4.2.2. Анализ свойств многоканальной СМО

Анализ свойств многоканальной СМО с однородным потоком заявок и накопителем неограниченной ёмкости может быть выполнен с использованием представленных выше математических моделей, определяющих зависимости характеристик обслуживания заявок от параметров поступления и обслуживания заявок для установившегося (стационарного) режима работы системы.

1. На рис.4.4 показан характер зависимости среднего времени ожидания w и среднего времени пребывания u заявок в системе от числа обслуживающих приборов K . Очевидно, что с увеличением числа обслуживающих приборов времена ожидания и пребывания заявок уменьшаются, при этом в пределе при $K \rightarrow \infty$ время ожидания стремится к нулю, а время пребывания достигает своего наименьшего значения, равного длительности обслуживания заявок.

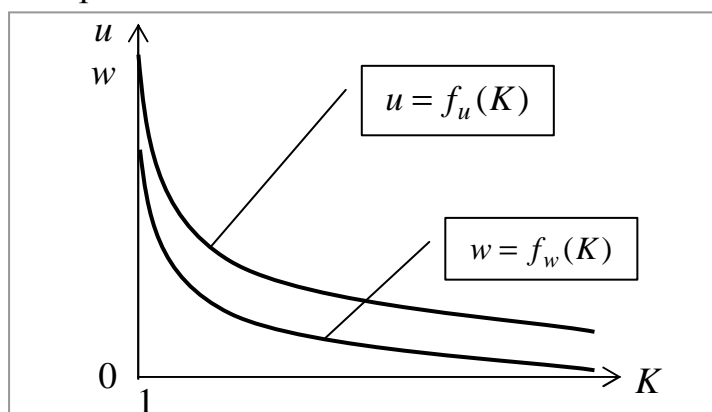


Рис.4.4. Зависимости времени ожидания и пребывания заявок от числа приборов

2. На рис.4.5 показаны аналогичные зависимости, но при условии, что при увеличении числа обслуживающих приборов K их суммарная производительность (скорость работы) остается постоянной, т.е. $V_{\Sigma} = KV_K = const$, где V_K – производительность одного прибора при наличии в системе K обслуживающих приборов. Из представленных графиков

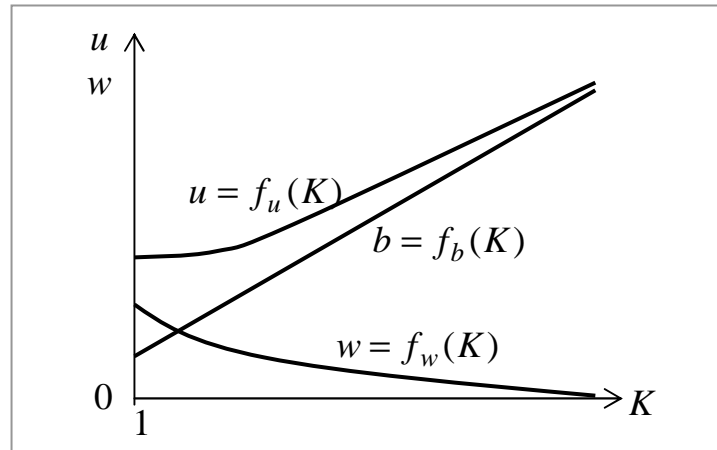


Рис.4.5. Зависимость времени пребывания заявок от числа приборов при $V_{\Sigma} = const$

видно, что среднее время ожидания w заявок, как и в предыдущем случае, уменьшается с увеличением числа приборов, однако время пребывания u заявок в системе увеличивается. Последнее объясняется тем, что с увеличением числа приборов K производительность каждого из них для сохранения суммарной производительности системы уменьшается пропорционально K и, следовательно, линейно увеличивается длительность обслуживания заявки в приборе. При этом скорость увеличения длительности обслуживания больше скорости уменьшения времени ожидания, что в сумме приводит к увеличению времени пребывания заявок в системе. В пределе при $K \rightarrow \infty$ время пребывания заявок асимптотически стремится к длительности обслуживания заявок.

Таким образом, при проектировании систем обслуживания следует иметь в виду, что с точки зрения задержек (времени пребывания заявок) более эффективной является одноканальная система, чем многоканальная, при равенстве суммарной производительности. Основным достоинством многоканальной системы является более высокая надёжность, проявляющаяся в том, что при выходе из строя одного или даже нескольких обслуживающих приборов система продолжает функционировать, хотя и с меньшей эффективностью, что заключается в увеличении времени пребывания заявок в системе.

3. Можно показать, что среднее время ожидания заявок, как и для одноканальных систем, существенно зависит от нагрузки y (загрузки ρ) системы. При $y \geq K$ ($\rho \rightarrow 1$) время ожидания заявок возрастает неограниченно: $w \rightarrow \infty$, то есть заявки могут ожидать обслуживания сколь угодно долго.

4.3. Одноканальные СМО с неоднородным потоком заявок

«Никогда не ставьте задачу, решение которой вам неизвестно» (Правило Берке)

Рассмотрим одноканальную СМО с неоднородным потоком заявок, в которую поступают N классов заявок, образующие простейшие потоки с

интенсивностями $\lambda_1, \dots, \lambda_H$. Длительность τ_{b_k} обслуживания заявок класса k распределена по произвольному закону со средним значением b_k и коэффициентом вариации ν_{b_k} . Выбор заявок из очереди на обслуживание осуществляется в соответствии с заданной дисциплиной обслуживания, в качестве которой будем рассматривать:

- дисциплину обслуживания беспriorитетную (ДО БП), при которой заявки выбираются на обслуживание в порядке поступления;
- дисциплину обслуживания заявок с относительными приоритетами (ДО ОП);
- дисциплину обслуживания заявок с абсолютными приоритетами (ДО АП).

В качестве основной характеристики, описывающей эффективность функционирования системы, будем рассматривать средние времена ожидания заявок разных классов, на основе которых легко могут быть рассчитаны все остальные характеристики с использованием фундаментальных зависимостей, представленных в разделе 3 (п.3.3.5).

При этом следует иметь в виду, что представленные ниже формулы были получены при следующих предположениях:

1) СМО содержит *один обслуживающий прибор*, который в каждый момент времени может обслуживать только одну заявку;

2) СМО имеет *накопитель заявок неограниченной ёмкости*, что означает отсутствие отказов поступающим заявкам при их постановке в очередь, то есть любая поступающая заявка всегда найдёт в накопителе место для ожидания независимо от того, сколько заявок уже находится в очереди;

3) заявки разных классов, поступающие в СМО независимо друг от друга, образуют *простейшие потоки*;

4) длительности обслуживания заявок каждого класса в приборе распределены по *произвольному закону* и не зависят друг от друга;

5) обслуживающий прибор не простаивает, если в системе (накопителе) имеется хотя бы одна заявка любого класса, причем после завершения обслуживания очередной заявки *мгновенно* из накопителя выбирается следующая заявка в соответствии с заданной дисциплиной обслуживания;

6) при использовании ДО БП заявки разных классов выбираются на обслуживание только в зависимости от времени поступления в систему по правилу «раньше пришел – раньше обслужен», независимо от номера класса, к которому принадлежит заявка;

7) при использовании приоритетных дисциплин (ДО ОП и ДО АП) приоритеты классам заявок назначены по принципу «*класс с меньшим номером имеет более высокий приоритет*», то есть наивысшим приоритетом обладают заявки класса 1;

8) в случае ДО АП заявка, обслуживание которой прервано более высокоприоритетной заявкой, *возвращается в накопитель*, где ожидает дальнейшего обслуживания, причем ее обслуживание продолжается с *прерванного места*.

4.3.1. Характеристики и свойства ДО БП

При беспriorитетной ДО средние времена ожидания одинаковы для всех классов заявок и определяются по следующей формуле:

$$w_k^{\text{БП}} = w^{\text{БП}} = \frac{\sum_{i=1}^N \lambda_i b_i^2 (1 + \nu_{b_i}^2)}{2(1 - R)} \quad (k = 1, \dots, N), \quad (4.9)$$

где $R = \sum_{i=1}^N \rho_i = \sum_{i=1}^N \lambda_i b_i$ – суммарная загрузка системы.

Выражение (4.5) получено в предположении, что в системе существует стационарный режим и отсутствует перегрузка: $R < 1$.

Анализ представленной аналитической зависимости (4.5) позволяет выявить **свойства ДО БП** и сформулировать следующие выводы.

1. *Среднее время ожидания заявок разных классов при использовании ДО БП одинаково при любых интенсивностях поступления $\lambda_1, \dots, \lambda_N$ и законах распределений $B_1(\tau), \dots, B_N(\tau)$ длительностей обслуживания заявок:* $w_k^{\text{БП}} = w^{\text{БП}}$ для всех $k = 1, \dots, N$. Отметим, что средние времена пребывания в системе заявок разных классов, в общем случае, различны, так как различны длительности обслуживания: $u_k^{\text{БП}} = w^{\text{БП}} + b_k$ ($k = 1, \dots, N$).

2. *Среднее время ожидания заявок в очереди минимально при постоянной (детерминированной) длительности обслуживания заявок каждого класса, когда коэффициент вариации длительности обслуживания $\nu_{b_k} = 0$, и увеличивается с ростом коэффициента вариации (дисперсии) длительности обслуживания.* Заметим, что зависимость среднего времени ожидания от коэффициента вариации ν_{b_k} носит нелинейный характер. Так, например, при экспоненциально распределенной длительности обслуживания, когда $\nu_{b_k} = 1$, среднее время ожидания заявок увеличивается в 2 раза, а при $\nu_{b_k} = 2$ – в 5 раз, по сравнению с детерминированным обслуживанием.

3. Среднее время ожидания заявок существенно зависит от суммарной нагрузки Y (загрузки R) системы (рис.4.6,а). При $Y \geq 1$ ($R \rightarrow 1$) *время ожидания* заявок всех классов *возрастает неограниченно*: $w^{\text{БП}} \rightarrow \infty$, то есть заявки могут ожидать обслуживания сколь угодно долго. Отметим, что увеличение суммарной нагрузки может быть

обусловлено двумя факторами: увеличением интенсивностей поступления в систему заявок разных классов или увеличением длительности обслуживания заявок (например, за счет уменьшения скорости работы обслуживающего прибора).

Зависимость среднего времени пребывания в системе заявок разных классов от суммарной нагрузки аналогична зависимости времени ожидания (рис.4.6,б). Единственное отличие состоит в том, что *средние времена пребывания в системе заявок разных классов, в общем случае, различны*, то есть $u_i \neq u_j$ ($i \neq j$), поэтому на графике, в отличие от времени ожидания, могут отображаться несколько зависимостей. Это различие обусловлено различием длительностей обслуживания заявок разных классов.

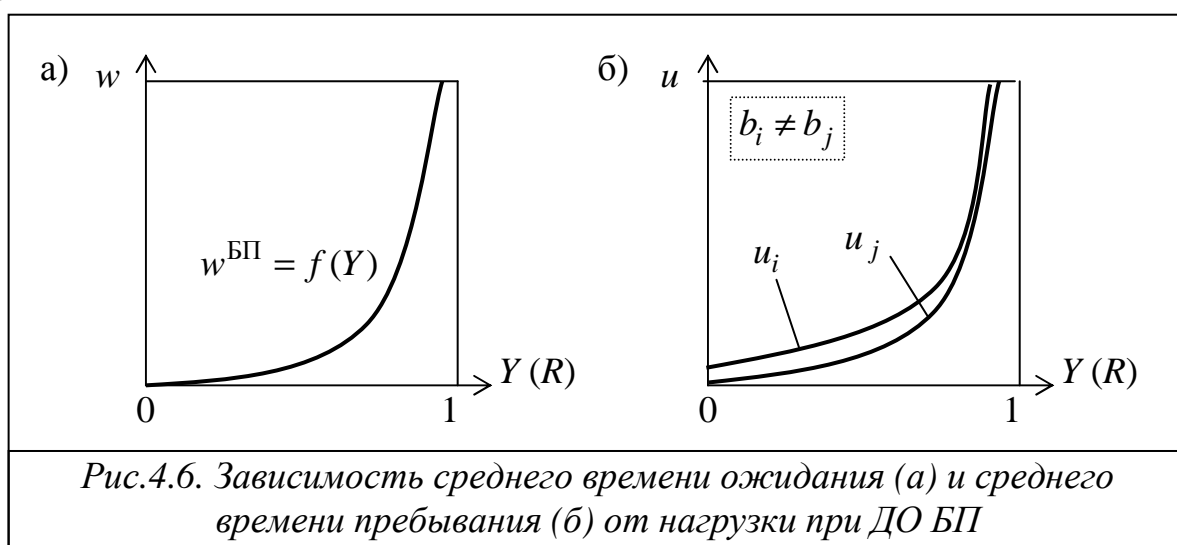


Рис.4.6. Зависимость среднего времени ожидания (а) и среднего времени пребывания (б) от нагрузки при ДО БП

Аналогично, на графиках, отображающих зависимости средних длин очередей и числа заявок в системе от суммарной нагрузки, в общем случае, будут изображаться несколько кривых, соответствующих разным классам заявок. Отметим, что *средние длины очередей заявок разных классов, несмотря на одинаковое время ожидания, в общем случае, различны* и, в соответствии с формулой Литтла ($l_i = \lambda_i w^{\text{БП}}$), совпадают только в случае равенства интенсивностей поступления заявок разных классов в систему.

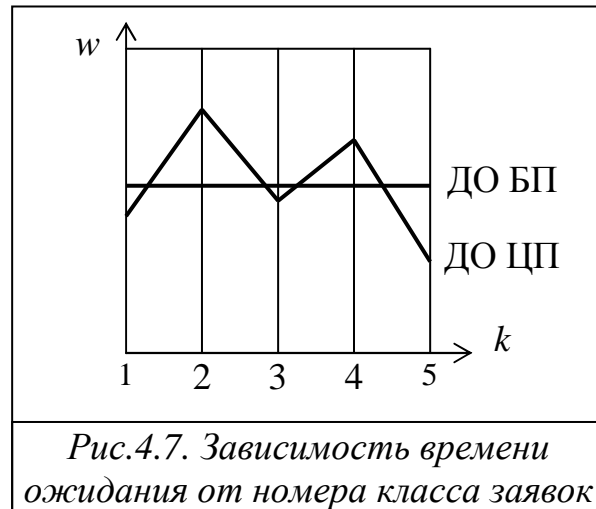
4. Можно показать, что для *бесприоритетной дисциплины обслуживания в обратном порядке (ООП)*, когда заявки на обслуживание выбираются по правилу «последний пришёл – первый обслужен», *средние времена ожидания заявок будут такими же*, как и при обслуживании в порядке поступления (ОПП), *но дисперсия времени ожидания будет больше*. Это обусловлено тем, что заявки, поступившие последними, будут ожидать незначительное время, в то время как заявки, попавшие в начало очереди, могут ожидать обслуживания достаточно долго, что обуславливает большой разброс значений времени ожидания.

5. Аналитическое исследование дисциплины обслуживания в циклическом порядке (ДО ЦП) достаточно сложно и связано с громоздкими математическими выкладками. Поэтому, не выписывая

громоздких формул, отметим лишь наиболее характерные особенности, присущие этой ДО.

Для дисциплины обслуживания в циклическом порядке среднее время ожидания заявок разных классов в общем случае не одинаково (рис.4.7).

Это различие зависит от соотношения параметров потоков ($\lambda_1, \dots, \lambda_H$) и обслуживания ($B_1(\tau), \dots, B_H(\tau)$) заявок разных классов. В некоторых случаях ДО ЦП позволяет обеспечить меньшую суммарную длину очереди заявок, чем ДО БП. Зависимость среднего времени ожидания заявок каждого класса от суммарной нагрузки Y имеет такой же вид, как и для ДО БП (рис.4.6).



4.3.2. Характеристики и свойства ДО ОП

Приоритеты называются *относительными*, если они учитываются только в момент выбора заявки на обслуживание и не сказываются на работе системы в период обслуживания заявки любого класса (приоритета). Относительность приоритета связана со следующим. После завершения обслуживания какой-либо заявки из очереди на обслуживание выбирается заявка класса с наиболее высоким приоритетом, поступившая ранее других заявок этого класса (такого же приоритета). Если в процессе её обслуживания в систему поступят заявки с более высоким приоритетом, то обслуживание рассматриваемой заявки не будет прекращено, то есть эта заявка, захватив прибор, оказывается как бы более приоритетной. Таким образом, приоритет *относителен* в том смысле, что он имеет место лишь в момент выбора заявок на обслуживание и отсутствует, если прибор занят обслуживанием какой-либо заявки.

Введение относительных приоритетов (ОП) позволяет уменьшить по сравнению с ДО БП время ожидания высокоприоритетных заявок.

При описании свойств для определённости будем полагать, что относительные приоритеты назначены по правилу: «более высокий приоритет – классу заявок с меньшим номером».

Для ДО ОП среднее время ожидания заявок класса k определяется по следующей формуле:

$$w_k^{\text{ОП}} = \frac{\sum_{i=1}^H \lambda_i b_i^2 (1 + v_{b_i}^2)}{2(1 - R_{k-1})(1 - R_k)} \quad (k = 1, \dots, H), \quad (4.10)$$

где R_{k-1} и R_k – суммарные загрузки, создаваемые заявками, которые имеют приоритет не ниже $(k-1)$ и k соответственно:

$$R_{k-1} = \sum_{i=1}^{k-1} \rho_i; \quad R_k = \sum_{i=1}^k \rho_i. \quad (4.11)$$

Анализ представленной зависимости (4.10) позволяет выявить свойства ДО ОП и сформулировать следующие выводы.

1. Введение относительных приоритетов по сравнению с ДО БП приводит к уменьшению времени ожидания высокоприоритетных заявок первого класса и к увеличению времени ожидания низкоприоритетных заявок класса H : $w_1^{ОП} < w_1^{БП}$ и $w_H^{ОП} > w_H^{БП}$.

2. При использовании ДО ОП средние времена ожидания заявок монотонно увеличиваются с уменьшением приоритета при любых интенсивностях поступления $\lambda_1, \dots, \lambda_H$ и законах распределения $B_1(\tau), \dots, B_H(\tau)$ длительностей обслуживания: $w_1^{ОП} < w_2^{ОП} < \dots < w_H^{ОП}$.

Отметим, что для средних времён пребывания заявок разных классов последнее соотношение, в общем случае, может и не выполняться.

Свойства, сформулированные выше, иллюстрируются рис.4.8,а, показывающим характер зависимости среднего времени ожидания заявок w_k от номера класса k при использовании ДО БП и ДО ОП.

3. На рис.4.8,б показаны зависимости среднего времени ожидания заявок разных классов от суммарной нагрузки Y системы при использовании ДО ОП. Здесь же для сравнения приведена аналогичная зависимость для ДО БП (штриховая линия). Характер зависимостей свидетельствует о том, что для ДО ОП при $Y \rightarrow 1$ резко увеличивается время ожидания заявок низкоприоритетных классов, в то время как для высокоприоритетных заявок это увеличение незначительно. Более того, для высокоприоритетных заявок обеспечивается достаточно хорошее качество обслуживания, то есть небольшое время ожидания даже при возникновении перегрузок, когда суммарная нагрузка становится больше единицы: $Y \geq 1$. Это свойство, называемое **защитой от перегрузок**, обеспечивается за счет отказа в обслуживании низкоприоритетным заявкам, время ожидания которых при этом резко возрастает. При ДО БП защита от перегрузок *отсутствует* для всех классов заявок.

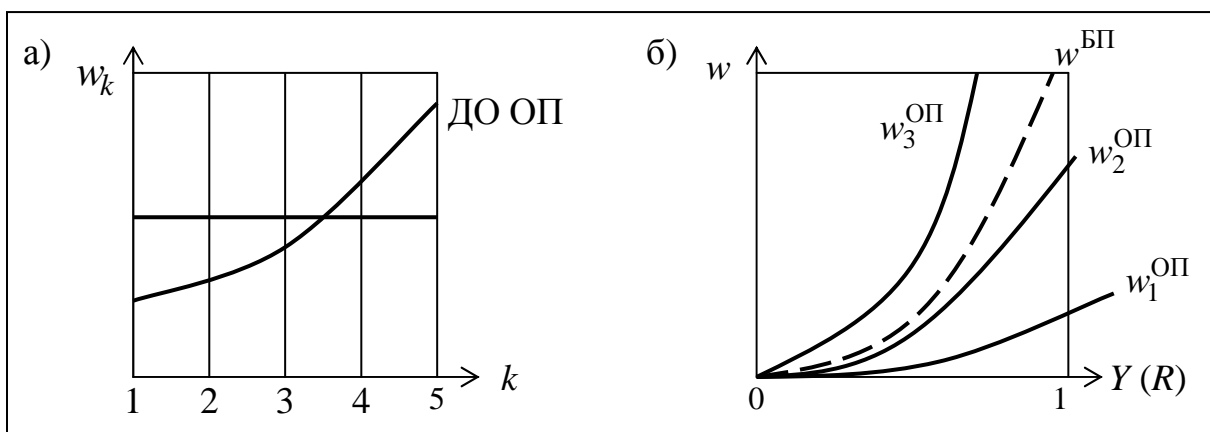
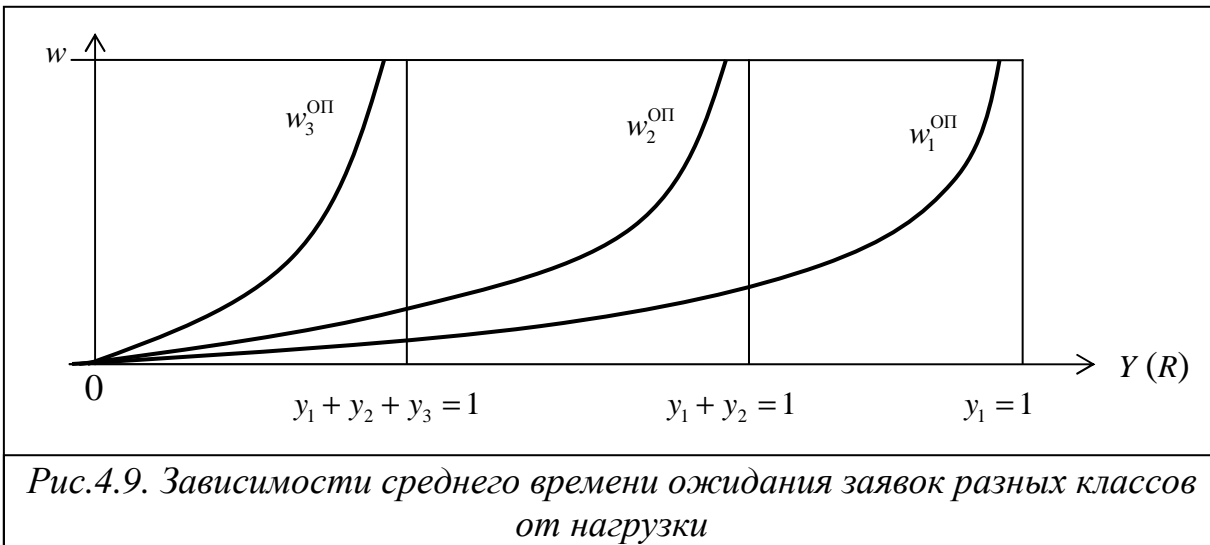


Рис.4.8. Зависимости среднего времени ожидания от номера класса (а) и от нагрузки (б) для ДО ОП и ДО БП

4. Рассмотрим более детально свойство защиты от перегрузок при ДО ОП, для чего построим зависимости среднего времени ожидания заявок трех классов при значительном росте нагрузки Y (рис.4.9).



При достижении суммарной нагрузки, создаваемой заявками всех трех классов, значения 1 ($y_1 + y_2 + y_3 = 1$) время ожидания заявок 3-го класса устремляется в бесконечность, что означает отказ в обслуживании, при этом заявки классов 1 и 2 продолжают обслуживаться и имеют конечное время ожидания. Дальнейшее увеличение нагрузки приводит к отказу в обслуживании заявок второго класса при $y_1 + y_2 = 1$, то есть когда создаваемая заявками 1-го и 2-го классов нагрузка достигнет значения 1. Заявки первого класса получают отказ в обслуживании при $y_1 = 1$. Таким образом, в отличие от ДО БП при ДО ОП система полностью перестаёт обслуживать заявки, то есть функционировать, только в том случае, если нагрузка, создаваемая заявками самого высокоприоритетного (первого) класса, достигнет значения 1.

4.3.3. Характеристики и свойства ДО АП

Иногда время ожидания заявок некоторых классов необходимо уменьшить в такой степени, которая недостижима при использовании ДО ОП. Можно предположить, что время ожидания уменьшится, если при поступлении высокоприоритетной заявки обслуживание ранее поступившей заявки с низким приоритетом прерывается, и прибор переходит к обслуживанию высокоприоритетной заявки. Приоритет, прерывающий обслуживание низкоприоритетной заявки, называется **абсолютным**, а соответствующая дисциплина – **дисциплиной обслуживания с абсолютными приоритетами** (ДО АП).

Прерванная заявка может быть потеряна или возвращена в накопитель, где она будет ожидать дальнейшего обслуживания. В последнем случае возможны два варианта продолжения обслуживания прерванной заявки:

- обслуживание с начала, то есть прерванная заявка будет обслуживаться заново с самого начала;
- дообслуживание, когда обслуживание прерванной заявки в приборе будет выполняться с прерванного места.

В дальнейшем, если не оговорено иное, будем предполагать *дообслуживание* прерванной заявки.

Для ДО АП среднее время ожидания заявок класса k определяется по следующей формуле:

$$w_k^{АП} = \frac{\sum_{i=1}^k \lambda_i b_i (1 + \nu_{b_i}^2)}{2(1 - R_{k-1})(1 - R_k)} + \frac{R_{k-1} b_k}{1 - R_{k-1}} \quad (k = 1, \dots, H) \quad (4.12)$$

где R_{k-1} и R_k – суммарные загрузки, создаваемые заявками, которые имеют приоритет не ниже $(k-1)$ и k соответственно, и определяемые по формулам (4.11).

Анализ выражения (4.12) для расчета среднего времени ожидания заявок при использовании ДО АП и его сопоставление с аналогичным выражением (4.10) для ДО ОП позволяет выявить **свойства ДО АП** и сформулировать следующие выводы.

1. Выражение (4.12) содержит два слагаемых: $w_k^{АП} = s_k + z_k$, отображающих среднее *время ожидания начала обслуживания* s_k и среднее *время ожидания в прерванном состоянии* z_k соответственно:

$$s_k = \frac{\sum_{i=1}^k \lambda_i b_i (1 + \nu_{b_i}^2)}{2(1 - R_{k-1})(1 - R_k)}, \quad z_k = \frac{R_{k-1} b_k}{1 - R_{k-1}} \quad (k = 1, \dots, H).$$

2. *Время ожидания заявок класса k зависит только от значений параметров классов $1, \dots, k$ заявок, имеющих более высокий или такой же приоритет, и не зависит от параметров классов заявок $k+1, \dots, H$, имеющих более низкий приоритет.*

3. *Для заявок класса 1, имеющих самый высокий абсолютный приоритет, обеспечивается минимально возможное время ожидания по сравнению со всеми другими ДО, то есть при любой другой ДО среднее время ожидания заявок первого класса не может быть меньше, чем при ДО АП. Это объясняется тем, что в случае ДО АП заявки первого класса обслуживаются как бы в изоляции, независимо от заявок других классов.*

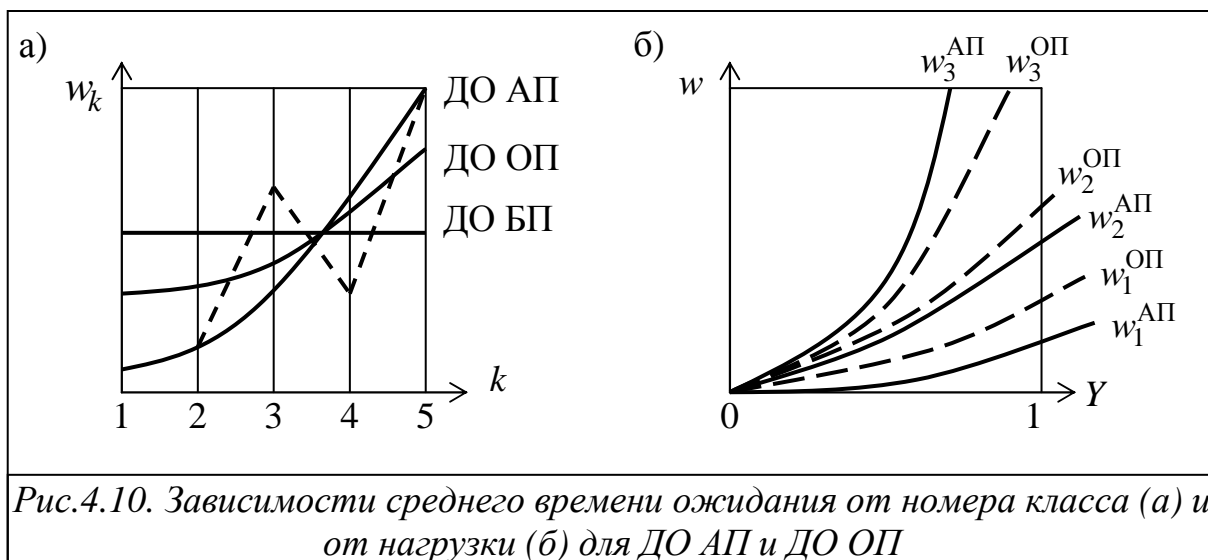
4. *Времена ожидания начала обслуживания s_k монотонно увеличиваются с уменьшением приоритета: $s_1 < s_2 < \dots < s_H$, однако время ожидания высокоприоритетной заявки в прерванном состоянии z_k может оказаться больше времени ожидания z_{k+1} заявки с более низким приоритетом, если длительности обслуживания связаны соотношением $b_k \gg b_{k+1}$, так как количество прерываний заявками более высокого*

приоритета и, следовательно, время ожидания в прерванном состоянии прямо пропорционально зависит от длительности обслуживания заявок данного класса. Вследствие этого, *полное время ожидания заявок высокоприоритетного класса*, складывающееся из времени ожидания начала обслуживания и времени ожидания в прерванном состоянии, *может оказаться больше, чем у заявок класса с низким приоритетом*: $w_k^{\text{АП}} \gg w_{k+1}^{\text{АП}}$. Очевидно, что $w_1^{\text{АП}} < w_2^{\text{АП}} < \dots < w_H^{\text{АП}}$, если длительности обслуживания заявок разных классов связаны соотношением $b_1 \leq b_2 \leq \dots \leq b_H$.

5. Введение АП по сравнению с ОП приводит к уменьшению среднего времени ожидания самых высокоприоритетных заявок первого класса и к его увеличению для заявок класса H : $w_1^{\text{АП}} < w_1^{\text{ОП}}$ и $w_H^{\text{АП}} > w_H^{\text{ОП}}$.

Два последних результата иллюстрируются рис.4.10,а. Для ДО АП пунктиром показан случай, когда $w_3^{\text{АП}} \gg w_4^{\text{АП}}$, из чего следует, что $b_3 \gg b_4$.

Зависимость полного времени ожидания от суммарной нагрузки Y системы при использовании ДО АП аналогична зависимости для ДО ОП (см. рис.4.10,б) с тем лишь отличием, что при ДО АП высокоприоритетные заявки лучше защищены от перегрузок.



4.3.4. Законы сохранения

«Число законов стремится заполнить все доступное для публикации пространство»
(Закон Диджиованни)

Изменение ДО позволяет уменьшить время ожидания высокоприоритетных заявок за счет увеличения времени ожидания низкоприоритетных заявок. Очевидно, что за счет изменения ДО нельзя добиться того, чтобы уменьшилось или увеличилось время ожидания заявок всех классов. Этот факт сформулирован в виде закона сохранения времени ожидания.

Формулировка закона сохранения времени ожидания. Для любой дисциплины обслуживания (ДО)

$$\sum_{i=1}^H \rho_i w_i = \underset{\text{ДО}}{\text{Const}}, \quad (4.13)$$

то есть сумма произведений загрузок ρ_i на среднее время ожидания w_i ($i = \overline{1, H}$) заявок всех классов инвариантна относительно ДО.

Закон сохранения времени ожидания выполняется при следующих условиях:

- система без потерь – все заявки на обслуживание удовлетворяются;
- система простаивает лишь в том случае, когда в ней нет заявок;
- при наличии прерываний длительность обслуживания прерванных заявок распределена по экспоненциальному закону;
- все поступающие потоки заявок – простейшие, и длительности обслуживания не зависят от интенсивностей потоков заявок.

Значение константы в законе сохранения можно определить следующим образом. Поскольку закон сохранения справедлив для любых ДО, удовлетворяющих перечисленным условиям, то он справедлив и для ДО БП, для которой $w_k^{\text{БП}} = w^{\text{БП}}$ для всех ($k = 1, \dots, H$). Отсюда находим значение константы:

$$\text{Const} = w^{\text{БП}} \sum_{i=1}^H \rho_i = R w^{\text{БП}}.$$

Подставив полученное значение константы и формулу (4.9) для расчёта $w^{\text{БП}}$ в закон сохранения, окончательно получим:

$$\sum_{i=1}^H \rho_i w_i = \frac{R \sum_{i=1}^H \lambda_i b_i^2 (1 + v_{bi}^2)}{2(1 - R)}. \quad (4.14)$$

Закон сохранения времени ожидания универсален и справедлив для всех ДО, удовлетворяющих указанным условиям. Его можно использовать для оценки достоверности приближенных результатов, полученных при исследовании сложных ДО и проведении имитационного моделирования, а также при решении задач синтеза.

Модификация закона сохранения. Закон сохранения может быть модифицирован применительно ко времени пребывания заявок в системе с учетом того, что $w_i = u_i - b_i$. Подставив это выражение в закон сохранения времени ожидания (4.14) после некоторых преобразований, получим закон сохранения времени пребывания:

$$\sum_{i=1}^H \rho_i u_i = \frac{R \sum_{i=1}^H \lambda_i b_i^2 (1 + v_{bi}^2)}{2(1 - R)} + \sum_{i=1}^H \rho_i b_i. \quad (4.15)$$

Заметим, что изменение ДО приводит только к изменению времени ожидания и времени пребывания, а остальные величины, входящие в выражения (4.14) и (4.15), не изменяются.

Рассмотрим случай, когда средние длительности обслуживания заявок разных классов одинаковы: $b_i = b = \text{const}$ для всех $i = \overline{1, N}$. Тогда выражение (4.14) может быть преобразовано следующим образом:

$$\sum_{i=1}^N \lambda_i b w_i = b \sum_{i=1}^N \lambda_i w_i = b \sum_{i=1}^N l_i = b L = \text{Const},$$

ДО

откуда получим новую формулировку закона сохранения в виде **закона сохранения суммарной длины очереди заявок**:

$$\sum_{i=1}^N \lambda_i w_i = L = \text{Const}.$$

ДО

Таким образом, если средние длительности обслуживания заявок разных классов одинаковы, то изменение ДО не приводит к изменению суммарной длины L очередей заявок всех классов, которая остается постоянной. В то же время длины очередей l_i ($i = \overline{1, N}$) заявок каждого класса меняются с изменением ДО.

4.4. Разомкнутые экспоненциальные СеМО с однородным потоком заявок

«Чем сложнее и грандиознее план, тем больше шансов, что он провалится»
(Производная от закона Мэрфи)

4.4.1. Описание разомкнутых СеМО

Рассмотрим разомкнутую экспоненциальную сеть массового обслуживания (СеМО) с однородным потоком заявок при следующих предположениях:

1) разомкнутая СеМО (РСеМО) произвольной топологии содержит n узлов;

2) после завершения обслуживания в каком-либо узле передача заявки в другой узел происходит мгновенно;

3) в качестве узлов могут быть как одноканальные, так и многоканальные СМО;

4) все приборы многоканального узла являются идентичными, и любая заявка может обслуживаться любым прибором;

5) заявка, поступившая в многоканальный узел, когда все или несколько приборов свободны, направляется случайным образом в любой свободный прибор;

6) в каждом узле РСеМО имеется накопитель заявок неограниченной ёмкости, что означает отсутствие отказов поступающим заявкам при их постановке в очередь, то есть любая поступающая в узел заявка всегда

найдет в накопителе место для ожидания независимо от того, сколько заявок уже находится в очереди;

7) заявки поступают в РСМО из *внешнего независимого* источника и образуют *простейший поток* заявок;

8) длительности обслуживания заявок во всех узлах сети представляют собой случайные величины, распределенные по *экспоненциальному закону*;

9) обслуживающий прибор любого узла *не простаивает*, если в его накопителе имеется хотя бы одна заявка, причем после завершения обслуживания очередной заявки *мгновенно* из накопителя выбирается следующая заявка;

10) в каждом узле сети заявки из накопителя выбираются в соответствии с *бесприоритетной* дисциплиной обслуживания в порядке поступления (ОПП) по правилу «первым пришел – первым обслужен» (FIFO – First In First Out).

Для описания *линейных разомкнутых однородных экспоненциальных* СеМО необходимо задать следующую совокупность **параметров**:

- число узлов в сети: n ;
- число обслуживающих приборов в узлах сети: K_1, \dots, K_n ;
- матрицу вероятностей передач: $\mathbf{P} = [p_{ij} \mid i, j = 0, 1, \dots, n]$, где вероятности передач p_{ij} должны удовлетворять условию (3.23): сумма элементов каждой строки должна быть равна 1;
- интенсивность λ_0 источника заявок, поступающих в РСМО;
- средние длительности обслуживания заявок в узлах сети: b_1, \dots, b_n .

На основе перечисленных параметров могут быть рассчитаны узловые и сетевые характеристики, описывающие эффективность функционирования соответственно узлов и РСМО в целом.

Расчет характеристик функционирования *линейных разомкнутых однородных экспоненциальных* СеМО базируется на эквивалентном преобразовании сети и проводится в четыре этапа:

- расчет коэффициентов передач α_j и интенсивностей потоков заявок λ_j в узлах $j = \overline{1, n}$ СеМО;
- проверка условия отсутствия перегрузок в СеМО;
- расчет узловых характеристик;
- расчет сетевых характеристик.

4.4.2. Расчет коэффициентов передач и интенсивностей потоков заявок в узлах РСМО

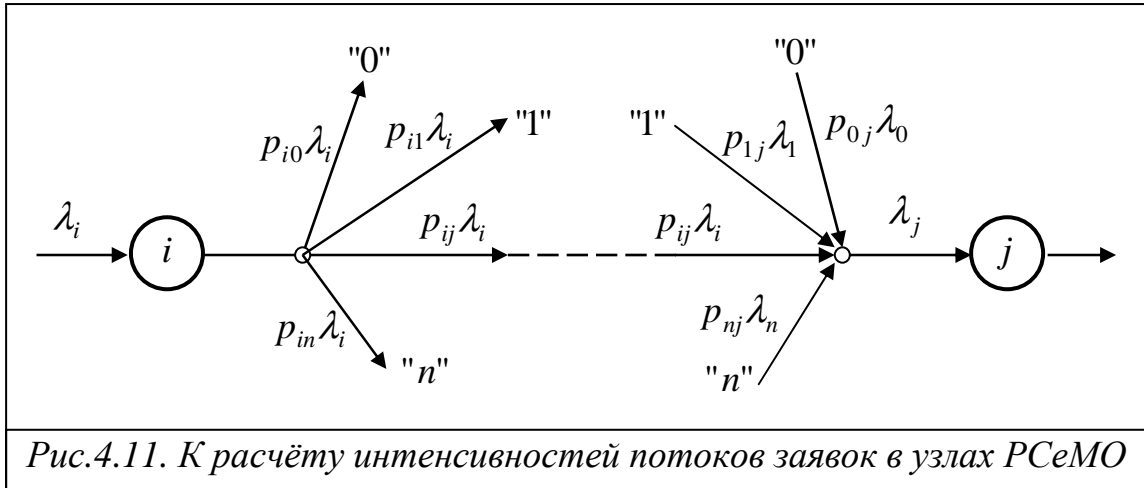
Покажем, что интенсивности $\lambda_0, \dots, \lambda_n$ потоков заявок, поступающих в узлы $0, \dots, n$ сети, однозначно определяются вероятностями передач p_{ij} ($i, j = 1, \dots, n$), задающими маршруты заявок в СеМО.

Будем рассматривать только установившийся режим.

Так как в линейной СеМО заявки не размножаются и не теряются, то интенсивности входящего и выходящего потоков для любого узла будут равны между собой.

Интенсивность потока заявок, входящих в любой узел j сети, равна сумме интенсивностей потоков заявок, поступающих в него из других узлов $i \in \overline{0, n}$ (рис.4.11). Поскольку заявки из узла i поступают в узел j с вероятностью p_{ij} , то интенсивность потока заявок, поступающих из i в j , равна $p_{ij}\lambda_i$, где λ_i - интенсивность выходящего и, следовательно, входящего потока заявок узла i . С учетом этого, на входе узла j имеется поток с интенсивностью

$$\lambda_j = \sum_{i=0}^n p_{ij} \lambda_i \quad (i = 0, 1, \dots, n). \quad (4.16)$$



Выражение (4.16) представляет собой *систему линейных алгебраических уравнений* $(n+1)$ -го порядка, из которой могут быть найдены интенсивности потоков заявок в виде соотношения $\lambda_j = \alpha_j \lambda_0$ ($j = \overline{1, n}$). Коэффициент α_j называется *коэффициентом передачи* и определяет среднее число попаданий заявки в узел j за время ее нахождения в сети, причем $\alpha_0 = 1$.

Для разомкнутой СеМО известна интенсивность источника заявок λ_0 . Можно показать, что система уравнений для расчета интенсивностей имеет единственное решение вида $\lambda_j = \alpha_j \lambda_0$, где λ_0 - заданная величина.

4.4.3. Проверка условия отсутствия перегрузок в СеМО

В п.3.4.2 показано, что в разомкнутой СеМО отсутствуют перегрузки, если выполняется условие (3.25):

$$\lambda_0 < \min\left(\frac{K_1}{\alpha_1 b_1}, \frac{K_2}{\alpha_2 b_2}, \dots, \frac{K_n}{\alpha_n b_n}\right).$$

Если указанное условие не выполняется, то, как следует из него, стационарный режим в разомкнутой СеМО может быть реализован одним из следующих способов:

- уменьшением интенсивности λ_0 внешнего источника заявок до значения, при котором это условие будет выполняться;
- увеличением количества обслуживающих приборов K_j в перегруженных узлах;
- уменьшением длительностей b_j обслуживания заявок в перегруженных узлах;
- уменьшением коэффициентов передач α_j в перегруженных узлах.

4.4.4. Расчет узловых характеристик РСемо

Один и тот же объект, рассматриваемый на разных уровнях детализации, можно представить различными моделями массового обслуживания, характеристики которых одинаковы или отличаются на величину, не превосходящую заданной погрешности. При выполнении определенных условий такие модели легко преобразуются друг в друга.

Для сетевых моделей в виде разомкнутых и замкнутых СеМО могут использоваться два вида преобразований:

- эквивалентное преобразование;
- толерантное преобразование.

Две сетевые модели *эквивалентны*, если сравниваемые характеристики этих моделей не отличаются друг от друга.

Две сетевые модели *толерантны* (подобны), если значения определенных характеристик отличаются друг от друга на величину, не превосходящую заданную.

Использование свойств эквивалентных и толерантных моделей позволяет упростить расчет характеристик моделей путем замены сложных сетевых моделей более простыми. Эквивалентными могут быть сетевые модели одного типа (например, две замкнутые сети), толерантными — модели как одного, так и разных типов [11].

Расчет характеристик функционирования линейных разомкнутых однородных экспоненциальных СеМО базируется на эквивалентном преобразовании сети, заключающемся в представлении *разомкнутой СеМО с n узлами* в виде *n независимых экспоненциальных СМО типа*

М/М/Н (простейший поток заявок, длительность обслуживания распределена по экспоненциальному закону, N обслуживающих приборов). При этом интенсивность входящего потока заявок в СМО, отображающую узел j ($j = \overline{1, n}$) сети, определяется из системы алгебраических уравнений (4.16) через интенсивность входящего в сеть потока и коэффициент передачи узла: $\lambda_j = \alpha_j \lambda_0$, а средняя длительность обслуживания заявок в СМО равна длительности обслуживания b_j заявок в соответствующем узле СеМО.

Характеристики всех n СМО (время ожидания заявок в очереди и пребывания в системе, длина очереди и число заявок в системе, среднее число занятых приборов и т.д.) представляют собой узловые характеристики СеМО.

Среднее время ожидания заявок в очереди может быть рассчитано с использованием выражения (4.8) для многоканальных СМО типа М/М/Н или выражения (4.1) для одноканальных СМО типа М/М/1, остальные характеристики узла j ($j = \overline{1, n}$) – с использованием фундаментальных соотношений, представленных в п.3.4.3, а именно:

- нагрузка в узле j , показывающая среднее число занятых приборов:
 $y_j = \lambda_j b_j$;
- загрузка узла j : $\rho_j = \min(y_j / K_j; 1)$, где K_j – число обслуживающих приборов в узле j ;
- коэффициент простоя узла: $\pi_j = 1 - \rho_j$;
- время пребывания заявок в узле: $u_j = w_j + b_j$;
- длина очереди заявок: $l_j = \lambda_j w_j$;
- число заявок в узле (в очереди и на обслуживании в приборе):
 $m_j = \lambda_j u_j$.

Рассчитанные таким образом характеристики отдельных СМО в точности соответствуют узловым характеристикам исходной СеМО, то есть в отношении своих характеристик модель массового обслуживания, представляющая собой совокупность независимых СМО (каждая СМО рассматривается независимо от других), строго эквивалентна исходной разомкнутой СеМО в целом.

4.4.5. Расчет сетевых характеристик РСеМО

Сетевые характеристики, описывающие эффективность функционирования СеМО в целом, рассчитываются на основе полученных значений узловых характеристик.

В состав сетевых характеристик входят:

- среднее число заявок, ожидающих обслуживания в сети, и среднее число заявок, находящихся в сети:

$$L = \sum_{j=1}^n l_j; \quad M = \sum_{j=1}^n m_j,$$

где l_j – средняя длина очереди и m_j – среднее число заявок в узле j ;

- среднее время ожидания и среднее время пребывания заявок в сети:

$$W = \sum_{j=1}^n \alpha_j w_j; \quad U = \sum_{j=1}^n \alpha_j u_j,$$

где w_j и u_j – соответственно среднее время ожидания и среднее время пребывания заявок в узле j ; α_j – коэффициент передачи для узла j , показывающий среднее число попаданий заявки в узел j за время ее нахождения в сети.

Пример 4.2. Проиллюстрируем изложенный метод расчета характеристик функционирования линейных разомкнутых однородных экспоненциальных СеМО на примере СеМО с четырьмя узлами ($n = 4$), граф которой представлен на рис.4.12. Связи между узлами СеМО описываются следующей матрицей вероятностей передач:

	0	1	2	3	4
0	0	1	0	0	0
1	0,1	0	0,2	0,7	0
2	0	0	0	0	1
3	0	0	0	0	1
4	0	1	0	0	0

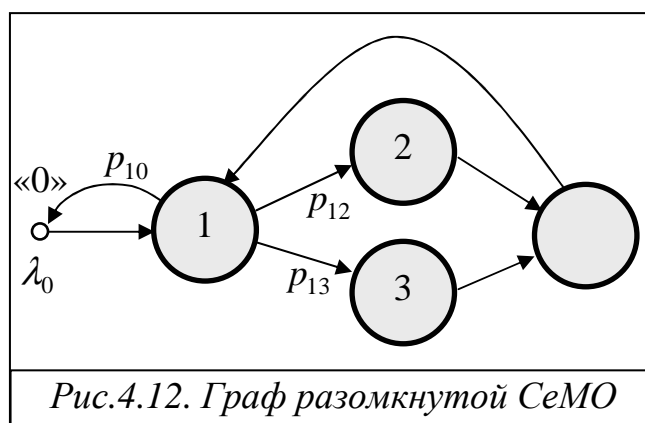


Рис.4.12. Граф разомкнутой СеМО

В РСеМО поступает простейший поток заявок с интенсивностью $\lambda_0 = 0,1 \text{ с}^{-1}$. Положим, что все узлы СеМО – одноканальные, а средние длительности обслуживания заявок в узлах соответственно равны: $b_1 = 0,8 \text{ с}$; $b_2 = 2 \text{ с}$; $b_3 = 0,4 \text{ с}$; $b_4 = 0,3 \text{ с}$.

Система линейных алгебраических уравнений для расчёта интенсивностей потоков заявок в узлах СеМО, согласно (4.16), имеет вид:

$$\left. \begin{aligned} \lambda_0 &= p_{10} \lambda_1 = 0,1 \lambda_1 \\ \lambda_1 &= p_{01} \lambda_0 + p_{41} \lambda_4 = \lambda_0 + \lambda_4 \\ \lambda_2 &= p_{12} \lambda_1 = 0,2 \lambda_1 \\ \lambda_3 &= p_{13} \lambda_1 = 0,7 \lambda_1 \\ \lambda_4 &= p_{24} \lambda_2 + p_{34} \lambda_3 = \lambda_2 + \lambda_3 \end{aligned} \right\}.$$

Решая эту систему уравнений, получим следующие значения интенсивностей: $\lambda_1 = 1 \text{ с}^{-1}$, $\lambda_2 = 0,2 \text{ с}^{-1}$, $\lambda_3 = 0,7 \text{ с}^{-1}$, $\lambda_4 = 0,9 \text{ с}^{-1}$. Тогда

коэффициенты передач будут равны: $\alpha_1 = \lambda_1 / \lambda_0 = 10$; $\alpha_2 = \lambda_2 / \lambda_0 = 2$; $\alpha_3 = \lambda_3 / \lambda_0 = 7$; $\alpha_4 = \lambda_4 / \lambda_0 = 9$.

Определим предельную интенсивность поступления заявок в разомкнутую СеМО, при которой в сети отсутствуют перегрузки. Для этого воспользуемся выражением (3.25), определяющим условие отсутствия перегрузок в РСеМО:

$$\lambda_0 < \min\left(\frac{K_1}{\alpha_1 b_1}, \frac{K_2}{\alpha_2 b_2}, \frac{K_3}{\alpha_3 b_3}, \frac{K_4}{\alpha_4 b_4}\right) = 0,125 \text{ с}^{-1}.$$

РСеМО работает без перегрузок, поскольку данное условие выполняется.

В соответствии с эквивалентным преобразованием представим рассматриваемую экспоненциальную разомкнутую СеМО в виде 4-х независимых СМО типа М/М/1, в которые поступают простейшие потоки заявок соответственно с интенсивностями: $\lambda_1 = 1 \text{ с}^{-1}$, $\lambda_2 = 0,2 \text{ с}^{-1}$, $\lambda_3 = 0,7 \text{ с}^{-1}$, $\lambda_4 = 0,9 \text{ с}^{-1}$, а средние длительности обслуживания заявок в СМО совпадают с длительностями обслуживания в соответствующих узлах СеМО: $b_1 = 0,8 \text{ с}$; $b_2 = 2 \text{ с}$; $b_3 = 0,4 \text{ с}$; $b_4 = 0,3 \text{ с}$.

Значения узловых характеристик СеМО, рассчитанные с использованием выражения (4.1) для среднего времени ожидания заявок в очереди СМО типа М/М/1 и фундаментальных соотношений, представленных в п.3.4.3, приведены в табл.4.1.

Таблица 4.1

Узловые характеристики	Расчётные формулы	Узел 1	Узел 2	Узел 3	Узел 4
Нагрузка	$y_j = \lambda_j b_j$	0,8	0,4	0,28	0,27
Загрузка	$\rho_j = \min(y_j / K_j; 1)$	0,8	0,4	0,28	0,27
Коэф-т простоя	$\pi_j = 1 - \rho_j$	0,2	0,6	0,72	0,73
Время ожидания	$w_j = \rho_j b_j / (1 - \rho_j)$	3,2	1,33	0,16	0,11
Время пребывания	$u_j = w_j + b_j$	4	3,33	0,56	0,41
Длина очереди	$l_j = \lambda_j w_j$	3,2	0,27	0,11	0,10
Число заявок в узле	$m_j = \lambda_j u_j$	4	0,67	0,39	0,37

В табл.4.2 представлены математические зависимости и полученные на их основе значения сетевых характеристик, рассчитанные с учётом найденных значений узловых характеристик.

Таблица 4.2

Сетевые характеристики	Расчётные формулы	Значения
Время ожидания в сети	$W = \sum_{j=1}^n \alpha_j w_j$	36,75
Время пребывания в сети	$U = \sum_{j=1}^n \alpha_j u_j$	54,25
Число заявок в состоянии ожидания	$L = \sum_{j=1}^n l_j$	3,68
Число заявок в сети	$M = \sum_{j=1}^n m_j$	5,43

4.4.6. Анализ свойств разомкнутых СеМО

Свойства разомкнутых СеМО определяются значениями узловых и сетевых характеристик, связанных между собой зависимостями, представленными в разделе 3. Наибольший интерес представляют свойства сети в целом, поскольку свойства отдельных узлов СеМО аналогичны свойствам соответствующих одноканальных и многоканальных СМО.

На рис. 4.13 показана зависимость основной сетевой характеристики РСемо – среднего времени пребывания U заявок в сети от интенсивности λ_0 поступления заявок в сеть. Зависимость $U' = f'(\lambda_0)$ аналогична зависимости среднего времени пребывания заявок в СМО от загрузки системы, изменение которой может быть обусловлено, в частности, изменением интенсивности поступления заявок в СМО. Как и в СМО, имеется некоторое предельное значение интенсивности $\lambda_{0\max}'$, при котором

среднее время пребывания заявок в сети становится бесконечно большим, что свидетельствует о перегрузке в СеМО. Выше (см.п.3.4.2) показано, что в РСемо отсутствуют перегрузки, если они отсутствуют во всех узлах сети, то есть перегрузка в разомкнутой СеМО наступает в том случае, когда загрузка одного из узлов сети становится равной единице. Такой узел называется «узким местом» и характеризуется тем, что очередь заявок перед ним со временем растёт до

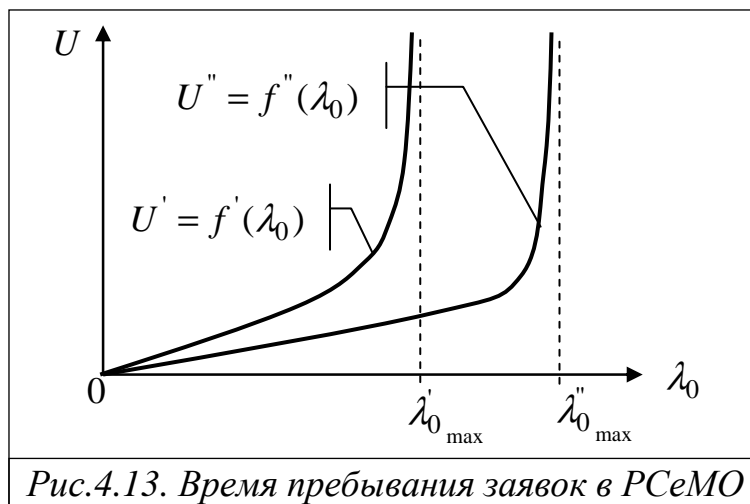


Рис.4.13. Время пребывания заявок в РСемо

бесконечности и, как следствие, становится бесконечным число заявок в разомкнутой СеМО.

Для того чтобы избавиться в РСеМО от перегрузки, необходимо *разгрузить* «узкое место». Это может быть достигнуто следующими способами:

- увеличением скорости работы (быстродействия) обслуживающего прибора;
- увеличением числа обслуживающих приборов в узле.

Любой из этих способов позволяет увеличить производительность СеМО в целом и, как следствие, улучшить характеристики сети. Зависимость среднего времени пребывания U заявок в сети от интенсивности λ_0 поступления заявок в сеть принимает вид $U'' = f''(\lambda_0)$, то есть время пребывания заявок при одной и той же интенсивности λ_0 становится меньше (поскольку сеть имеет большую производительность), а предельное значение интенсивности $\lambda_{0\max}''$, при котором наступает перегрузка СеМО, становится больше: $\lambda_{0\max}'' > \lambda_{0\max}'$. При этом появляется новое узкое место в СеМО, и дальнейшее улучшение сети может быть достигнуто путём разгрузки нового узкого места. Очевидно, что если СеМО является моделью реальной технической системы, разгрузка узкого места за счёт увеличения скорости работы обслуживающего прибора или числа приборов в узле означает увеличение стоимости реальной системы.

Существует ещё один способ разгрузки узкого места СеМО, заключающийся в *уменьшении вероятности передачи* заявок к узлу, являющемуся узким местом. Этот способ часто используется в реальных системах и обычно не связан с увеличением стоимости системы. Например, в вычислительной системе изменение вероятностей передач к накопителям внешней памяти может быть достигнуто за счёт перераспределения файлов между накопителями: наиболее часто используемые файлы, расположенные в наиболее загруженном накопителе, переносятся в наименее загруженный накопитель. При этом уменьшается количество обращений к загруженному накопителю (коэффициент передачи соответствующего узла СеМО).

Характер зависимостей других сетевых характеристик (времени ожидания, числа заявок в сети и в состоянии ожидания) разомкнутой СеМО от интенсивности поступления заявок аналогичен показанному на рис. 4.13.

Пример 4.3. Проиллюстрируем способы разгрузки узкого места и получаемый от этого эффект для четырёхузловой разомкнутой СеМО, рассмотренной в примере 4.2. Там же было показано, что интенсивность поступления заявок в разомкнутую СеМО, при которой в сети отсутствуют перегрузки, должна удовлетворять условию: $\lambda_0 < 0,125 \text{ с}^{-1}$.

1. Рассчитаем сначала характеристики РСeMO, работающей в области загрузок, близких к 1, для чего положим, что интенсивность потока поступающих в сеть заявок равна $\lambda_0 = 0,12 \text{ с}^{-1}$

Тогда интенсивности потоков заявок в узлы РСeMO соответственно будут равны: $\lambda_1 = \alpha_1 \lambda_0 = 1,2 \text{ с}^{-1}$, $\lambda_2 = \alpha_2 \lambda_0 = 0,24 \text{ с}^{-1}$, $\lambda_3 = \alpha_3 \lambda_0 = 0,84 \text{ с}^{-1}$, $\lambda_4 = \alpha_4 \lambda_0 = 1,08 \text{ с}^{-1}$, а средние длительности обслуживания заявок, как и ранее, будут равны: $b_1 = 0,8 \text{ с}$; $b_2 = 2 \text{ с}$; $b_3 = 0,4 \text{ с}$; $b_4 = 0,3 \text{ с}$.

Рассчитанные значения узловых и сетевых характеристик СеМО приведены в табл.4.3.

Таблица 4.3

<i>Характеристики</i>	<i>Узел 1</i>	<i>Узел 2</i>	<i>Узел 3</i>	<i>Узел 4</i>	<i>СеМО</i>
Нагрузка	0,96	0,48	0,336	0,324	2,10
Загрузка	0,96	0,48	0,336	0,324	2,10
Время ожидания	19,2	1,85	0,202	0,144	198,4
Время пребывания	20	3,85	0,602	0,444	215,9
Длина очереди	23,04	0,44	0,170	0,155	23,8
Число заявок	24	0,92	0,506	0,479	25,9

Анализ представленных результатов показывает, что увеличение интенсивности поступления заявок в РСeMO всего лишь на 20% до значения $\lambda_0 = 0,12 \text{ с}^{-1}$, привело к резкому росту значений сетевых характеристик. В частности, среднее время пребывания заявок в сети выросло в 4 раза, а число заявок, находящихся в очередях – почти в 6,5 раз. Это говорит о том, что СеМО работает в области больших загрузок, где незначительное увеличение нагрузки приводит к существенному изменению характеристик обслуживания заявок. Наиболее загруженным узлом СеМО, то есть узким местом, является узел 1, загрузка которого много больше загрузок других узлов и составляет $\rho_1 = 0,96$. Именно в этом узле характеристики обслуживания заявок выросли наиболее существенно: среднее время пребывания заявок в 5 раз (с 4 до 20 секунд), а средняя длина очереди – более чем в 7 раз (с 3,2 до 23 заявок).

2. Для улучшения характеристик обслуживания заявок в РСeMO необходимо разгрузить узкое место сети, которым является узел 1. Для этого увеличим скорость работы обслуживающего прибора в 2 раза, что, в конечном счете, приведёт к уменьшению длительности обслуживания заявок в 2 раза, которая станет равной $b_1 = 0,4 \text{ с}$.

Рассчитанные значения узловых и сетевых характеристик СеМО после разгрузки узкого места приведены в табл.4.4.

Анализ представленных результатов показывает, что разгрузка узкого места позволила существенно уменьшить значения сетевых характеристик: среднее время пребывания заявок в сети уменьшилось более чем в 9

раз, а число заявок, находящихся в очередях – почти в 20 раз. Отметим, что изменение длительности обслуживания заявок в узле 1 привело к изменению узловых характеристик только этого узла; узловые характеристики остальных узлов не изменились. Это является следствием независимого функционирования узлов экспоненциальной разомкнутой СеМО, что фактически и позволяет использовать метод расчёта характеристик сети, основанный на декомпозиции, то есть представлении сети в виде совокупности независимых СМО.

Таблица 4.4

Узловые характеристики	Узел 1	Узел 2	Узел 3	Узел 4	СеМО
Нагрузка	0,48	0,48	0,336	0,324	1,62
Загрузка	0,48	0,48	0,336	0,324	1,62
Время ожидания	0,369	1,846	0,202	0,144	10,1
Время пребывания	0,769	3,846	0,602	0,444	23,6
Длина очереди	0,443	0,443	0,170	0,155	1,21
Число заявок	0,923	0,923	0,506	0,479	2,83

3. Для сравнения выполним разгрузку узкого места другим способом, а именно: увеличим число обслуживающих приборов в узле 1 с одного до двух: $K_1 = 2$, сохранив прежнее значение длительности обслуживания одним прибором: $b_1 = 0,8$ с.

Рассчитанные значения узловых и сетевых характеристик СеМО после разгрузки узкого места приведены в табл.4.5.

Таблица 4.5

Узловые характеристики	Узел 1	Узел 2	Узел 3	Узел 4	СеМО
Нагрузка	0,96	0,48	0,336	0,324	2,10
Загрузка	0,48	0,48	0,336	0,324	1,62
Время ожидания	0,288	1,846	0,202	0,144	9,28
Время пребывания	1,088	3,846	0,602	0,444	26,78
Длина очереди	0,346	0,443	0,170	0,155	1,11
Число заявок	1,306	0,923	0,506	0,479	3,21

Сравним полученные значения сетевых характеристик со значениями, представленными в табл. 4.4 для первого способа разгрузки узкого места за счёт уменьшения длительности обслуживания заявок. При втором способе разгрузки узкого места за счёт увеличения числа обслуживающих приборов ($K_1 = 2$; $b_1 = 0,8$ с) среднее время ожидания заявок в сети несколько уменьшилось по сравнению с первым способом ($K_1 = 1$; $b_1 = 0,4$ с). В то же время, среднее время пребывания заявок в РСМО увеличились более чем на 10%, что обусловлено большей длительностью обслуживания заявок ($b_1 = 0,8$ с) в каждом из приборов двухканального

узла 1 по сравнению с одноканальным узлом при первом способе ($b_1 = 0,4$ с). Как и в предыдущем случае, изменение числа обслуживающих приборов в узле 1 привело к изменению узловых характеристик только этого узла.

4.5. Замкнутые экспоненциальные СеМО с однородным потоком заявок

«Во всякой формуле константы (особенно те, которые взяты из технических справочников) должны рассматриваться как переменные» (*Универсальные законы ...*)

4.5.1. Описание замкнутых СеМО

Рассмотрим замкнутую экспоненциальную сеть массового обслуживания с однородным потоком заявок при следующих предположениях:

1) замкнутая СеМО (ЗСеМО) *произвольной топологии* содержит n узлов;

2) после завершения обслуживания в каком-либо узле передача заявки в другой узел происходит *мгновенно*;

3) все узлы замкнутой СеМО *одноканальные*;

4) в СеМО циркулирует *постоянное число заявок*;

5) длительности обслуживания заявок во всех узлах сети представляют собой случайные величины, распределенные по *экспоненциальному закону*;

6) *ёмкость накопителя* в каждом узле СеМО *достаточна* для хранения всех заявок, циркулирующих в сети, что означает отсутствие отказов поступающим заявкам при их постановке в очередь любого узла (в частности, можно считать, что ёмкость накопителя в каждом узле равна числу заявок, циркулирующих в сети);

7) обслуживающий прибор любого узла *не простаивает*, если в его накопителе имеется хотя бы одна заявка, причем после завершения обслуживания очередной заявки мгновенно из накопителя выбирается следующая заявка;

8) в каждом узле сети заявки из накопителя выбираются в соответствии с *бесприоритетной дисциплиной обслуживания* в порядке поступления (ОПП) по правилу «первым пришел – первым обслужен» (FIFO – First In First Out).

Для описания линейных замкнутых однородных экспоненциальных СеМО необходимо задать такую же совокупность параметров, как и для разомкнутых СеМО, с единственным отличием, заключающимся в том, что вместо интенсивности источника заявок следует задать число заявок, циркулирующих в ЗСеМО. Таким образом, совокупность параметров для замкнутых СеМО будет иметь следующий вид:

- *число узлов* в сети: n ;

- число обслуживающих приборов в узлах сети: K_1, \dots, K_n ;
- матрица вероятностей передач: $\mathbf{P} = [p_{ij} \mid i, j = 0, 1, \dots, n]$, где p_{ij} – вероятность передачи заявки из узла i в узел j ;
- число заявок M , циркулирующих в ЗСеМО;
- средние длительности обслуживания заявок в узлах сети: b_1, \dots, b_n .

На основе перечисленных параметров могут быть рассчитаны узловые и сетевые характеристики, описывающие эффективность функционирования соответственно узлов и ЗСеМО в целом.

Расчёт характеристик функционирования линейных замкнутых однородных экспоненциальных СеМО с одноканальными узлами базируется на так называемой «теореме о прибытии» и проводится с использованием метода средних значений в два этапа:

- расчёт коэффициентов передач в узлах замкнутой СеМО;
- расчёт характеристик ЗСеМО.

4.5.2. Расчёт коэффициентов передач в узлах ЗСеМО

Для замкнутой СеМО на первом этапе рассчитываются только коэффициенты передач. Интенсивности потоков заявок в узлах ЗСеМО не могут быть рассчитаны, как в РСеМО, поскольку для ЗСеМО изначально не известна интенсивность λ_0 , которая является не параметром, задаваемым в составе исходных данных, а характеристикой, представляющей собой производительность ЗСеМО и определяемой в процессе анализа эффективности функционирования ЗСеМО.

Для расчёта коэффициентов передач $\alpha_1, \dots, \alpha_n$ после некоторых преобразований можно воспользоваться той же системой линейных алгебраических уравнений (4.16). Для этого в левой и правой части выражения (4.16) представим интенсивности в виде $\lambda_j = \alpha_j \lambda_0$. Разделив левую и правую часть выражения (4.16) на λ_0 , окончательно получим систему линейных алгебраических уравнений относительно $\alpha_1, \dots, \alpha_n$:

$$\alpha_j = \sum_{i=0}^n p_{ij} \alpha_i \quad (i = 0, 1, \dots, n). \quad (4.17)$$

Полагая $\alpha_0 = 1$, можно найти корни системы уравнений, численно определяющие значения $\alpha_1, \dots, \alpha_n$.

4.5.3. Расчёт характеристик ЗСеМО

Характеристики ЗСеМО могут быть рассчитаны с использованием марковских процессов, поскольку количество состояний марковского процесса, в отличие от РСеМО, не бесконечно и равно числу сочетаний C_{M+n-1}^M , где n – число узлов в ЗСеМО и M – число заявок, циркулирующих в ЗСеМО. При этом основная трудность заключается в определении веро-

ятности состояний сети $P(M_1, \dots, M_n)$ в случае большой ее размерности ($n > 5; M > 5$), когда число состояний оказывается значительным. При выполнении расчетов на ЭВМ это, во многих случаях, приводит к потере значимости в процессе промежуточных вычислений и, следовательно, к невозможности получения конечных результатов.

От указанного недостатка свободен *метод средних значений*, позволяющий вычислять средние характеристики функционирования экспоненциальных СеМО на основе сравнительно простых рекуррентных соотношений.

Положим, что замкнутая однородная СеМО содержит n **одноканальных** узлов, длительности обслуживания заявок в которых распределены по экспоненциальному закону со средними значениями b_1, \dots, b_n соответственно. Пусть для каждого узла i сети известно среднее число попаданий заявки в данный узел за время ее нахождения в сети, то есть коэффициент передачи α_i , который, если конфигурация сети задана матрицей вероятностей передач $P = [p_{ij} | i, j = 0, 1, \dots, n]$, определяется в результате решения системы линейных алгебраических уравнений (4.17).

Обозначим: u_i - среднее время пребывания заявки в узле i за время пребывания в сети; m_i - среднее число заявок в узле i ($i = 1, \dots, n$); λ_0 - производительность замкнутой сети. Очевидно, что эти величины зависят от числа заявок M , циркулирующих в замкнутой сети, то есть $u_i = u_i(M)$; $m_i = m_i(M)$; $\lambda_0 = \lambda_0(M)$.

Можно показать, что имеют место следующие соотношения:

$$u_i(M) = b_i [1 + m_i(M - 1)]; \quad (4.18)$$

$$U(M) = \sum_{i=1}^n \alpha_i u_i(M); \quad (4.19)$$

$$\lambda_0(M) = \frac{M}{U(M)}; \quad (4.20)$$

$$m_i(M) = \alpha_i \lambda_0(M) u_i(M), \quad (4.21)$$

где $U(M)$ - среднее время пребывания заявок в сети при условии нахождения в ней M заявок; $m_i(0) = 0$.

Выражение (4.18) получено на основе так называемой *теоремы о прибытии* [1], утверждающей, что в замкнутой экспоненциальной сети с одноканальными узлами, в которой циркулируют M заявок, стационарная вероятность состояния любого узла в момент поступления в него новой заявки совпадает со стационарной вероятностью того же состояния рассматриваемого узла в сети, в которой циркулирует на одну заявку меньше, то есть $(M - 1)$ заявок. Это означает, что в сети с M заявками среднее число заявок $m_i(M)$, находящихся в узле i в момент поступления в этот узел новой заявки, равно $m_i(M - 1)$. Тогда среднее время пребывания

в узле i поступившей заявки будет складываться из среднего времени обслуживания всех $m_i(M-1)$ ранее поступивших и находящихся в узле i заявок и средней длительности обслуживания рассматриваемой заявки:

$$u_i(M) = b_i m_i(M-1) + b_i = b_i [1 + m_i(M-1)].$$

В этом выражении учтено, что среднее время дообслуживания заявки, находящейся в приборе на момент поступления рассматриваемой заявки, равно средней длительности обслуживания b_i в силу свойства отсутствия последействия, присущего экспоненциальному закону. Среднее время пребывания заявки в узле i за время ее нахождения в сети, учитывающее число попаданий α_i заявки в данный узел, равно $U_i(M) = \alpha_i u_i(M)$.

Выражения (4.19) и (4.20) представляют собой формулы Литтла для сети, а выражение (4.21) – для узла i , где $\lambda_i(M) = \alpha_i \lambda_0(M)$ – интенсивность потока заявок в узел i ($i = 1, \dots, n$).

На основе рекуррентных соотношений (4.18) – (4.21) последовательно для $M = 1, 2, \dots, M^*$, где M^* – заданное число заявок в замкнутой сети, могут быть рассчитаны средние значения характеристик замкнутой экспоненциальной СеМО.

Заметим, что приведенный метод расчета является *точным* для замкнутых экспоненциальных СеМО с *одноканальными* узлами.

Пример 4.4. Рассчитаем характеристики замкнутой однородной экспоненциальной СеМО, полученной путём преобразования разомкнутой СеМО (рис. 4.12), рассмотренной в Примере 4.2, в замкнутую. Положим, что «нулевая точка», отображающая завершение обслуживания заявок в сети и мгновенное формирование новой заявки, выбрана на дуге, выходящей из узла 1 и входящей снова в этот же узел (рис.4.14). Напомним, что в ЗСеМО относительно «нулевой точки» рассчитываются временные сетевые характеристики: время нахождения в состоянии ожидания и время пребывания заявок в сети, а также производительность ЗСеМО.

ЗСеМО содержит $n = 4$ одноканальных узла, связи между которыми описываются той же матрицей вероятностей передач:

	0	1	2	3	4
0	0	1	0	0	0
1	0,1	0	0,2	0,7	0
2	0	0	0	0	1
3	0	0	0	0	1
4	0	1	0	0	0

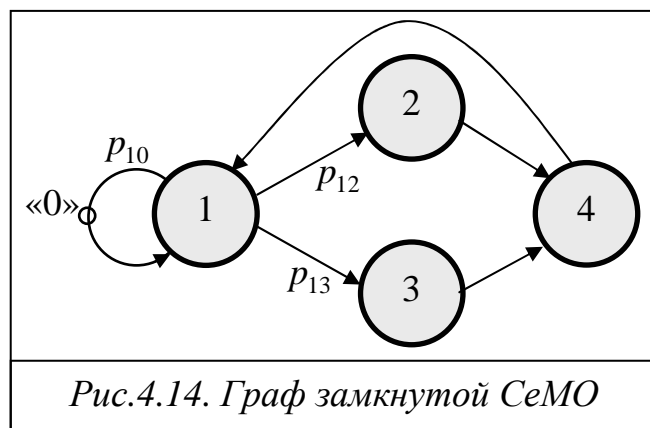


Рис.4.14. Граф замкнутой СеМО

Следовательно, коэффициенты передач для всех узлов, рассчитываемые путём решения системы линейных алгебраических уравнений (4.17), будут иметь те же самые значения: $\alpha_1 = 10$; $\alpha_2 = 2$; $\alpha_3 = 7$; $\alpha_4 = 9$.

В ЗСеМО циркулирует M заявок, средние длительности обслуживания которых в узлах равны: $b_1 = 0,8$ с; $b_2 = 2$ с; $b_3 = 0,4$ с; $b_4 = 0,3$ с.

Ниже в табл.4.6 представлены значения времени пребывания $u_i(M)$ и числа заявок $m_i(M)$ в узлах сети, а также среднего времени пребывания $U(M)$ заявок в сети и производительности $\lambda_0(M)$, рассчитанные на основе выражений (4.18) – (4.21), для числа циркулирующих в сети заявок $M = 1, 2, \dots, 6$. Корректность выполненных расчетов подтверждается тем, что для всех $M = 1, 2, \dots, 6$ выполняется проверочное условие:

$$\sum_{i=1}^4 m_i(M) = M.$$

Таблица 4.6

M	i	$u_i(M)$	$U(M)$	$\lambda_0(M)$	$m_i(M)$
1	1	0,8	17,5	0,057	0,46
	2	2,0			0,23
	3	0,4			0,16
	4	0,3			0,15
2	1	1,17	22,94	0,087	1,02
	2	2,46			0,43
	3	0,46			0,28
	4	0,35			0,27
3	1	1,61	28,87	0,104	1,68
	2	2,86			0,59
	3	0,51			0,37
	4	0,38			0,36
4	1	2,14	35,29	0,113	2,43
	2	3,19			0,72
	3	0,55			0,44
	4	0,41			0,42
5	1	2,74	42,14	0,119	3,25
	2	3,45			0,82
	3	0,57			0,48
	4	0,42			0,45
6	1	3,40	49,35	0,122	4,14
	2	3,63			0,88
	3	0,59			0,50
	4	0,44			0,48

На рис.4.15 представлены зависимости производительности рассматриваемой замкнутой СеМО и среднего времени пребывания заявок в сети

от количества $M = \overline{1,10}$ циркулирующих заявок. Анализ полученных результатов показывает, что все характеристики, включая производительность λ_0 , растут с увеличением M .

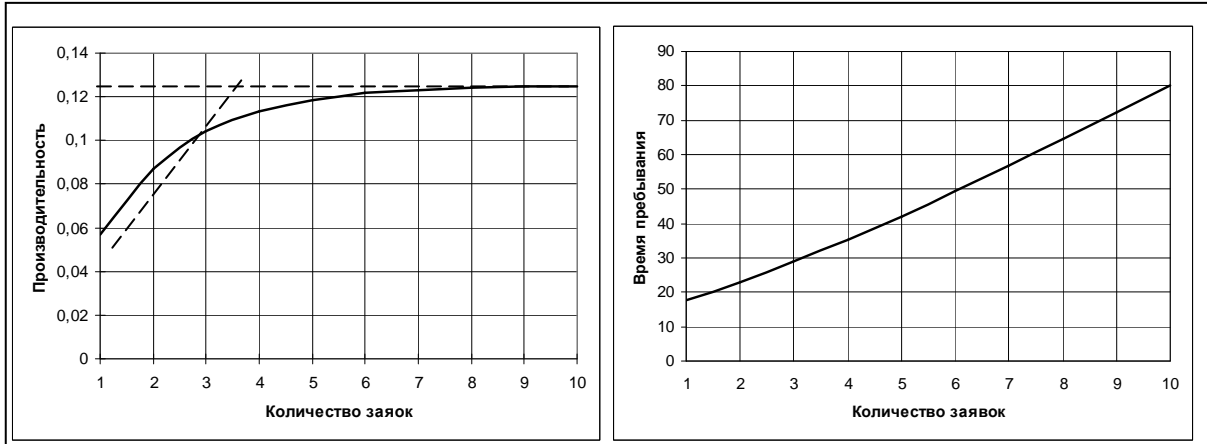


Рис.4.15. Производительность и время пребывания заявок в ЗСеМО

Производительность сети асимптотически приближается к максимально возможной производительности (пропускной способности ЗСеМО), совпадающей с предельно допустимой интенсивностью поступления заявок в аналогичной разомкнутой СеМО (см. Пример 4.1), при которой в сети отсутствуют перегрузки, и равна $\lambda_0 = 0,125 \text{ с}^{-1}$.

Среднее время пребывания заявок в ЗСеМО растёт неограниченно с увеличением количества заявок с сети.

Остальные характеристики замкнутой СеМО (загрузки и коэффициенты простоя узлов, время ожидания, длины очередей и число заявок в узлах сети, полное время ожидания в сети) могут быть рассчитаны с использованием фундаментальных соотношений, представленных в разделе 3 (п.3.4.3).

4.5.4. Анализ свойств замкнутых СеМО

Для замкнутых СеМО, как и для разомкнутых, наибольший интерес представляют свойства сети в целом, в частности, влияние циркулирующих в ЗСеМО числа заявок, на такие сетевые характеристики как производительность λ_0 замкнутой СеМО и среднее время пребывания U заявок в сети.

Анализ представленных на рис.4.16, зависимостей позволяет сформулировать следующие выводы.

1. Зависимость $\lambda_0 = f(M)$ производительности ЗСеМО λ_0 от числа M циркулирующих заявок вначале растёт с увеличением M до некоторого значения M_0 , после которого рост производительности замедляется, а с дальнейшим увеличением M производительность сети асимптотически стремится к некоторому предельному значению $\hat{\lambda}_0$, представляющему

собой пропускную способность ЗСеМО. Для объяснения этой зависимости вспомним, что производительность замкнутой сети измеряется как интенсивность потока заявок, проходящих через некоторую условную точку, обозначаемую как «0» и расположенную на одной из дуг СеМО, отображающей завершение обслуживания заявок в сети и мгновенное формирование новой заявки, поступающей в сеть. Выше (см. пример 4.4) было показано, что увеличение числа заявок в замкнутой СеМО приводит к увеличению значений всех сетевых характеристик, включая производительность λ_0 . В свою очередь, увеличение производительности приводит к увеличению загрузок узлов СеМО, связанных с интенсивностью λ_0 зависимостью:

$$\rho_j = \frac{\alpha_j \lambda_0 b_j}{K_j},$$

где α_j, b_j и K_j – соответственно коэффициент передачи, средняя длительность обслуживания и количество приборов в узле $j = \overline{1, n}$.

Когда число заявок в ЗСеМО достигает некоторого значения M_0 , загрузка одного из узлов становится близкой к 1, при этом практически прекращается рост производительности, которая при $M \rightarrow \infty$ достигает своего предельного значения – пропускной способности $\hat{\lambda}_0$. Такой узел представляет собой «узкое место» сети, и значение пропускной способности $\hat{\lambda}_0$ определяется пропускной способностью узкого места из условия, что загрузка ρ_y этого узла равна 1:

$$\rho_y = \frac{\alpha_y \lambda_0 b_y}{K_y} = 1.$$

Отсюда пропускная способность замкнутой СеМО:

$$\hat{\lambda}_0 = \frac{K_y}{\alpha_y b_y},$$

где α_y, b_y и K_y – соответственно коэффициент передачи, средняя длительность обслуживания и количество обслуживающих приборов в узле, являющимся узким местом.

Правая часть последнего выражения представляет собой пропускную способность узла, являющегося узким местом сети: $\mu_y = \frac{K_y}{\alpha_y b_y}$.

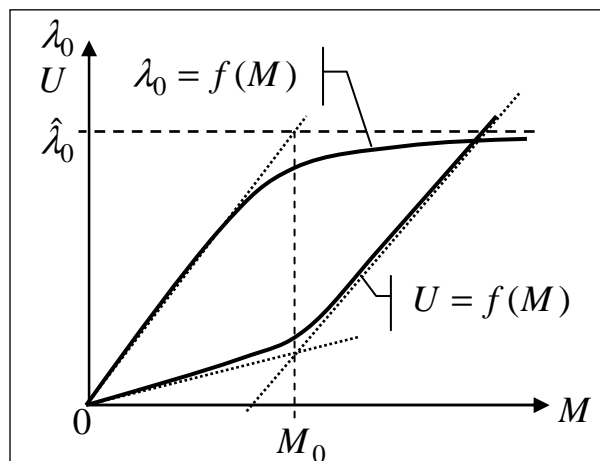


Рис.4.16. Характеристики ЗСеМО

Действительно, $\alpha_y b_y$ представляет собой полное время обслуживания одной заявки в данном узле с учётом того, что заявка за время нахождения в сети в среднем α_y раз побывает в данном узле. Тогда величина, обратная $\alpha_y b_y$, представляет собой интенсивность обслуживания заявок одним прибором в данном узле: $\mu_1 = 1/\alpha_y b_y$, а $\mu_y = K_y \mu_1$ – интенсивность обслуживания заявок узлом, то есть всеми приборами.

Этот же результат можно получить следующими рассуждениями. Если загрузка некоторого узла, являющегося узким местом СеМО, становится равной 1, то это означает, что все приборы данного узла постоянно обслуживают заявки, то есть не простаивают. Тогда интенсивность выходящего из этого узла потока заявок будет равна интенсивности обслуживания: $\lambda_y = \mu_y = K_y \mu_1$. Напомним, что интенсивность потока заявок в узле λ_y связана с производительностью ЗСеМО λ_0 зависимостью $\lambda_y = \alpha_y \lambda_0$. Отсюда вытекает, что производительность ЗСеМО равна $\lambda_0 = \frac{\lambda_y}{\alpha_y} = \frac{K_y \mu_1}{\alpha_y} = \frac{K_y}{\alpha_y b_y}$.

2. Среднее время пребывания заявок (рис.4.16) в замкнутой СеМО, как и производительность, растёт с увеличением числа M циркулирующих в сети заявок, причём вначале наблюдается незначительный рост, а затем, после значения $M = M_0$, наблюдается линейный рост времени пребывания.

Действительно, если в сети циркулирует только одна заявка, то в такой сети не может быть очередей, и время пребывания заявок в СеМО складывается только из времён обслуживания заявок в узлах с учётом коэффициентов передач:

$$U = \sum_{i=1}^n \alpha_i b_i.$$

С увеличением числа заявок M в узлах ЗСеМО появляются очереди, причём очевидно, что чем больше заявок в сети, тем более длинные очереди образуются в узлах и тем больше время ожидания, а, следовательно, и время пребывания заявок в ЗСеМО.

Сопоставляя зависимости производительности и среднего времени пребывания заявок от их числа в ЗСеМО, можно сделать следующий вывод: увеличение числа заявок в сети, с одной стороны, приводит к увеличению производительности, что может рассматриваться как положительный фактор, а, с другой стороны, – к увеличению времени пребывания заявок в сети, что является нежелательным фактором.

Точка $M = M_0$ характеризует некоторое граничное значение числа заявок в ЗСеМО. Дальнейшее увеличение числа заявок в сети оказывается нецелесообразным, поскольку приводит к резкому увеличению времени

пребывания заявок в ЗСеМО при незначительном увеличении производительности сети.

3. Когда загрузка узкого места становится равной единице, дальнейший рост производительности за счёт увеличения числа заявок в ЗСеМО невозможен. Для увеличения производительности ЗСеМО, как и в РСеМО, необходимо разгрузить узкое место, то есть уменьшить загрузку:

$$\rho_y = \frac{\alpha_y \lambda_0 b_y}{K_y} = 1, \text{ что при одной и той же производительности может быть}$$

достигнуто:

- уменьшением длительности обслуживания заявок b_y , например за счет увеличения скорости работы (быстродействия) обслуживающего прибора;
- увеличением числа обслуживающих приборов K_y в узле;
- уменьшением коэффициента передачи α_y или, что то же самое, вероятности передачи заявок к узлу, являющемуся узким местом.

Если до разгрузки узкого места зависимость производительности ЗСеМО от числа заявок в сети имела вид $\lambda'_0 = f(M)$ (рис.4.17), а пропускная способность была равна $\hat{\lambda}'_0$, то после разгрузки – зависимость производительности от числа заявок будет иметь вид $\lambda''_0 = f(M)$, а пропускная способность станет равной $\hat{\lambda}''_0 > \hat{\lambda}'_0$. При этом граничное значение числа заявок в ЗСеМО увеличится: $M''_0 > M'_0$

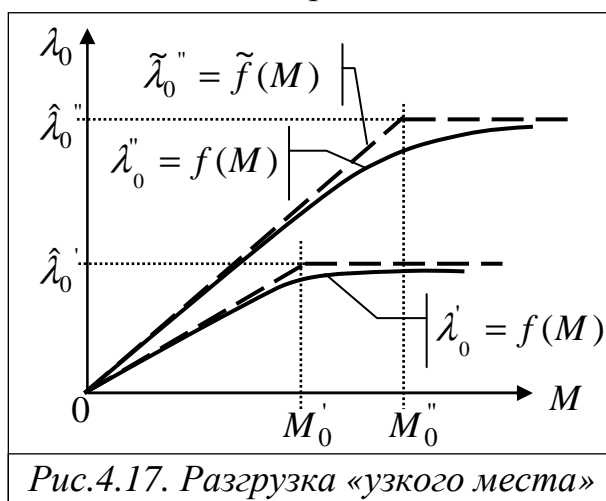


Рис.4.17. Разгрузка «узкого места»

Следует отметить, что к рассматриваемой зависимости производительности ЗСеМО λ_0 от числа M циркулирующих в сети заявок может быть применена линейная аппроксимация $\tilde{\lambda}_0 = \tilde{f}(M)$, показанная на рис.4.17 в виде пунктирных линий и представляющая собой верхнюю границу производительности ЗСеМО. Последнее означает, что производительность ЗСеМО будет не больше, чем рассчитанное верхнее значение.

Нетрудно представить себе и изобразить на графике, как изменится зависимость среднего времени пребывания заявок в замкнутой СеМО от числа циркулирующих в сети заявок после разгрузки узкого места.

Отметим, что в некоторых случаях разгрузка узкого места не приводит к улучшению характеристик СеМО, в частности, к увеличению производительности. Обычно это связано с тем, что в СеМО может существовать несколько узлов, являющихся «узкими местами». Условием

этого является равенство загрузок узлов: $\rho_i = \rho_j$ или $\frac{\alpha_i \lambda_0 b_i}{K_i} = \frac{\alpha_j \lambda_0 b_j}{K_j}$,

откуда окончательно получим: $\frac{\alpha_i b_i}{K_i} = \frac{\alpha_j b_j}{K_j}$ ($i \neq j$). В этом случае для

улучшения характеристик ЗСеМО необходимо одновременно разгрузить все «узкие места».

Последовательно разгружая узкие места СеМО, мы можем прийти к некоторой «идеальной» сети, в которой загрузки всех узлов одинаковы.

СеМО, в которой загрузки всех узлов равны, называется **сбалансированной**. Сбалансированная СеМО обладает наилучшими характеристиками по сравнению с несбалансированной.

При построении реальных систем, моделями которых служат СеМО, необходимо, по-возможности, строить сбалансированные системы, хотя на практике по многим причинам достичь этого не удаётся.

4.6. Резюме

1. Одноканальная экспоненциальная СМО М/М/1 является наиболее *простой* с точки зрения аналитического расчета. Средние времена ожидания и пребывания заявок в СМО М/М/1 рассчитываются по сравнительно простым формулам:

$$w = \frac{\rho b}{1 - \rho} \quad \text{и} \quad u = \frac{b}{1 - \rho},$$

где $\rho = \lambda b < 1$ – загрузка системы; λ – интенсивность поступления заявок в систему; b – средняя длительность обслуживания заявок в приборе.

Для СМО М/Г/1 среднее время ожидания заявок определяется по формуле Поллачека-Хинчина:

$$w = \frac{\lambda b^2 (1 + v_b^2)}{2(1 - \rho)},$$

где v_b – коэффициент вариации длительности обслуживания.

Для общего случая одноканальных СМО типа Г/Г/1 с однородным потоком применяются *приближённые* аналитические методы расчёта.

Свойства одноканальной СМО с однородным потоком заявок:

- среднее время ожидания заявок в очереди *минимально при детерминированной длительности обслуживания* заявок с коэффициентом вариации $v_b = 0$ и увеличивается нелинейно с ростом коэффициента вариации (дисперсии) длительности обслуживания;
- среднее время ожидания заявок существенно зависит от нагрузки y (загрузки ρ) системы и при $y \geq 1$ ($\rho \rightarrow 1$) возрастает *неограниченно*: $w \rightarrow \infty$, т.е. заявки могут ожидать обслуживания сколь угодно долго;

- для дисциплин обслуживания в обратном порядке и обслуживания в случайном порядке средние времена ожидания заявок будут такими же, как и при обслуживании в порядке поступления, но *дисперсии времени ожидания будут больше*.

2. В случае многоканальных СМО с однородным потоком заявок *точный* метод расчета среднего времени ожидания заявок разработан только для СМО типа М/М/К:

$$w = \frac{Pb}{K(1-\rho)},$$

где $\rho = \frac{\lambda b}{K}$ – загрузка системы; P – вероятность того, что все K приборов заняты обслуживанием заявок:

$$P = \frac{(K\rho)^K}{K!(1-\rho)} P_0,$$

где P_0 – вероятность простоя многоканальной СМО, то есть вероятность того, что в системе нет заявок:

$$P_0 = \left[\frac{(K\rho)^K}{K!(1-\rho)} + \sum_{i=0}^{K-1} \frac{(K\rho)^i}{i!} \right]^{-1}.$$

Свойства многоканальной СМО с однородным потоком заявок:

- с увеличением числа обслуживающих приборов времена ожидания и пребывания заявок уменьшаются, при этом в пределе при $K \rightarrow \infty$ время ожидания стремится к нулю, а время пребывания становится равным длительности обслуживания заявок;
- при увеличении числа обслуживающих приборов K и сохранении их суммарной производительности (скорости работы) время ожидания заявок уменьшается, однако время пребывания заявок в системе увеличивается и в пределе (при $K \rightarrow \infty$) асимптотически стремится к длительности обслуживания заявок, то есть с точки зрения задержек (времени пребывания заявок) более эффективной является одноканальная система, чем многоканальная, при равенстве суммарной производительности; достоинством многоканальной системы является более высокая надежность;
- среднее время ожидания заявок, как и для одноканальных систем, зависит от нагрузки y (загрузки ρ) системы и при $y \geq K$ ($\rho \rightarrow 1$) время ожидания заявок возрастает *неограниченно*: $w \rightarrow \infty$, то есть заявки могут ожидать обслуживания сколь угодно долго.

3. Для одноканальных СМО с неоднородным потоком заявок и беспriorитетной ДО средние времена ожидания одинаковы для всех классов заявок и определяются по следующей формуле:

$$w_k^{\text{БП}} = w^{\text{БП}} = \frac{\sum_{i=1}^H \lambda_i b_i^2 (1 + \nu_{b_i}^2)}{2(1 - R)} \quad (k = 1, \dots, H),$$

где $R = \sum_{i=1}^H \rho_i = \sum_{i=1}^H \lambda_i b_i$ – суммарная загрузка системы ($R < 1$).

Свойства одноканальной СМО с беспriorитетной ДО:

- среднее время ожидания заявок разных классов при использовании ДО БП *одинаково*;
- среднее время ожидания заявок в очереди *минимально* при детерминированной длительности обслуживания заявок каждого класса и увеличивается нелинейно с ростом коэффициента вариации (дисперсии) длительности обслуживания;
- среднее время ожидания заявок зависит от суммарной нагрузки Y (загрузки R) системы и при $Y \geq 1$ ($R \rightarrow 1$) время ожидания заявок всех классов возрастает *неограниченно*: $w^{\text{БП}} \rightarrow \infty$, однако средние времена пребывания в системе и средние длины очередей заявок разных классов, в общем случае, различны, поскольку различны длительности обслуживания и интенсивности поступления заявок разных классов;
- для беспriorитетной дисциплины обслуживания в обратном порядке средние времена ожидания заявок будут такими же, как и при обслуживании в порядке поступления, но дисперсия времени ожидания будет больше;
- для дисциплины обслуживания в циклическом порядке средние времена ожидания заявок разных классов в общем случае *не равны*.

4. Для ДО ОП средние времена ожидания заявок k -го класса определяются по следующей формуле:

$$w_k^{\text{ОП}} = \frac{\sum_{i=1}^H \lambda_i b_i^2 (1 + \nu_{b_i}^2)}{2(1 - R_{k-1})(1 - R_k)} \quad (k = 1, \dots, H),$$

где R_{k-1} и R_k – суммарные загрузки системы со стороны заявок, которые имеют приоритет не ниже $(k - 1)$ и k соответственно.

Свойства ДО ОП:

- введение относительных приоритетов по сравнению с ДО БП приводит к уменьшению времени ожидания высокоприоритетных заявок и к увеличению времени ожидания низкоприоритетных заявок;

- средние времена ожидания заявок при использовании ДО ОП *монотонно увеличиваются* с уменьшением приоритета: $w_1^{\text{ОП}} < w_2^{\text{ОП}} < \dots < w_H^{\text{ОП}}$;
- ДО ОП обладает свойством *защиты от перегрузок*, заключающемся в том, что высокоприоритетные заявки даже при возникновении перегрузки ($Y \geq 1$) имеют конечное время ожидания за счёт отказа в обслуживании низкоприоритетным заявкам, время ожидания которых при этом резко возрастает и стремится к бесконечности.

5. Для ДО АП средние времена ожидания заявок k -го класса определяются по следующей формуле:

$$w_k^{\text{АП}} = \frac{\sum_{i=1}^k \lambda_i b_i (1 + \nu \frac{b_i^2}{b_i})}{2(1 - R_{k-1})(1 - R_k)} + \frac{R_{k-1} b_k}{1 - R_{k-1}} \quad (k = 1, \dots, H),$$

где R_{k-1} и R_k – суммарные загрузки системы со стороны заявок, которые имеют приоритет не ниже $(k - 1)$ и k соответственно.

Свойства ДО АП:

- среднее время ожидания заявок складывается из двух составляющих: *среднего времени ожидания начала обслуживания и среднего времени ожидания в прерванном состоянии*;
- время ожидания заявок класса k зависит только от значений параметров классов заявок, имеющих более высокий или такой же приоритет, и *не зависит* от параметров классов заявок, имеющих более низкий приоритет;
- для заявок с самым высоким абсолютным приоритетом обеспечивается *минимально возможное время ожидания* по сравнению со всеми другими ДО;
- *полное время ожидания у заявок высокоприоритетного класса k может оказаться больше*, чем у заявок класса $(k + 1)$ с более низким приоритетом, если длительности обслуживания заявок этих классов связаны соотношением $b_k \gg b_{k+1}$;
- введение АП по сравнению с ОП приводит к уменьшению среднего времени ожидания самых высокоприоритетных заявок и к его увеличению для заявок класса самого низкого приоритета;
- при ДО АП высокоприоритетные заявки лучше защищены от перегрузок, чем при ДО ОП.

6. В соответствии с **законом сохранения** времени ожидания *изменение ДО позволяет уменьшить время ожидания высокоприоритетных заявок за счет увеличения времени ожидания низкоприоритетных заявок*:

$$\sum_{i=1}^H \rho_i w_i = \underset{\text{ДО}}{\text{Const.}}$$

Закон сохранения выполняется при следующих условиях:

- система без потерь;
- система простаивает лишь при отсутствии в системе заявок;
- при наличии прерываний длительность обслуживания прерванных заявок распределена по экспоненциальному закону;
- все поступающие потоки заявок – простейшие, и длительность обслуживания не зависит от параметров потоков заявок.

7. В качестве параметров линейных разомкнутых однородных экспоненциальных СеМО необходимо задать:

- число узлов в сети: n ;
- число обслуживающих приборов в узлах сети: K_1, \dots, K_n ;
- матрицу вероятностей передач: $\mathbf{P} = [p_{ij} \mid i, j = 0, 1, \dots, n]$, где p_{ij} – вероятность передачи заявки из узла i в узел j ;
- интенсивность λ_0 источника заявок, поступающих в РСеМО;
- средние длительности обслуживания заявок в узлах сети: b_1, \dots, b_n .

Условие отсутствия перегрузок в разомкнутой СеМО:

$$\lambda_0 < \min \left(\frac{K_1}{\alpha_1 b_1}, \frac{K_2}{\alpha_2 b_2}, \dots, \frac{K_n}{\alpha_n b_n} \right).$$

Расчет характеристик функционирования линейных разомкнутых однородных экспоненциальных СеМО базируется на эквивалентном преобразовании сети и проводится в три этапа:

- расчет интенсивностей потоков заявок λ_j в узлах $j = \overline{1, n}$ РСеМО путём решения системы линейных алгебраических уравнений:

$$\lambda_j = \sum_{i=0}^n p_{ij} \lambda_i \quad (j = 0, 1, \dots, n);$$

- расчет узловых характеристик:
 - нагрузка в узле j , показывающая среднее число занятых приборов: $y_j = \lambda_j b_j$;
 - загрузка узла j : $\rho_j = \min(y_j / K_j; 1)$, где K_j – число обслуживающих приборов в узле j ;
 - коэффициент простоя узла: $\pi_j = 1 - \rho_j$;
 - время пребывания заявок в узле: $u_j = w_j + b_j$;
 - длина очереди заявок: $l_j = \lambda_j w_j$;

□ число заявок в узле (в очереди и на обслуживании в приборе):
 $m_j = \lambda_j u_j$.

- расчет сетевых характеристик:

□ среднее число заявок, ожидающих обслуживания в сети, и среднее число заявок, находящихся в сети:

$$L = \sum_{j=1}^n l_j; \quad M = \sum_{j=1}^n m_j,$$

где l_j - средняя длина очереди и m_j - среднее число заявок в узле j ;

□ среднее время ожидания и среднее время пребывания заявок в сети:

$$W = \sum_{j=1}^n \alpha_j w_j; \quad U = \sum_{j=1}^n \alpha_j u_j,$$

где w_j и u_j - соответственно среднее время ожидания и среднее время пребывания заявок в узле j ; $\alpha_j = \lambda_j / \lambda_0$ ($j = \overline{1, n}$) - коэффициент передачи для узла j , показывающий среднее число попаданий заявки в узел j за время ее нахождения в сети.

Свойства разомкнутых СеМО:

- свойства отдельных узлов СеМО аналогичны свойствам соответствующих одноканальных и многоканальных СМО;
- с увеличением интенсивности λ_0 поступления заявок в сеть сетевые характеристики увеличиваются, причём имеется предельное значение интенсивности $\lambda_{0 \max}'$, при котором, в частности, среднее время пребывания заявок в сети становится бесконечно большим, что свидетельствует о перегрузке в СеМО;
- узел сети, загрузка которого с увеличением λ_0 стремится к единице, называется «узким местом» и характеризуется бесконечным ростом очереди заявок перед ним и, как следствие, бесконечным ростом числа заявок в СеМО;
- способы разгрузки «узкого места»:
 - увеличение скорости работы (быстродействия) обслуживающего прибора;
 - увеличение числа обслуживающих приборов в узле;
 - уменьшение вероятности передачи заявок к узлу, являющемуся узким местом.

8. В качестве параметров линейных замкнутых однородных экспоненциальных СеМО необходимо задать:

- число узлов в сети: n ;
- число обслуживающих приборов в узлах сети: K_1, \dots, K_n ;

- матрица вероятностей передач: $\mathbf{P} = [p_{ij} \mid i, j = 0, 1, \dots, n]$, где p_{ij} – вероятность передачи заявки из узла i в узел j ;
- число заявок M^* , циркулирующих в ЗСеМО;
- средние длительности обслуживания заявок в узлах сети: b_1, \dots, b_n .

В замкнутых СеМО всегда существует установившийся режим.

Расчет характеристик функционирования линейных замкнутых однородных экспоненциальных СеМО с одноканальными узлами проводится с использованием метода средних значений в два этапа:

- расчет коэффициентов передач в узлах замкнутой СеМО путём решения системы линейных алгебраических уравнений относительно $\alpha_1, \dots, \alpha_n$ с учётом того, что $(\alpha_0 = 1)$:

$$\alpha_j = \sum_{i=0}^n p_{ij} \alpha_i \quad (i = 0, 1, \dots, n);$$

- расчет характеристик ЗСеМО с использованием следующих рекуррентных соотношений для значений $M = 1, 2, \dots, M^*$:

$$u_i(M) = b_i [1 + m_i(M - 1)];$$

$$U(M) = \sum_{i=1}^n \alpha_i u_i(M);$$

$$\lambda_0(M) = \frac{M}{U(M)};$$

$$m_i(M) = \alpha_i \lambda_0(M) u_i(M),$$

где M^* – заданное число заявок в замкнутой сети; $m_i(0) = 0$.

Свойства замкнутых СеМО:

- зависимость производительности ЗСеМО λ_0 от числа M циркулирующих заявок растёт с увеличением M и стремится к некоторому предельному значению $\hat{\lambda}_0$, представляющему собой пропускную способность ЗСеМО;
- среднее время пребывания заявок в замкнутой СеМО, как и производительность, растёт с увеличением числа циркулирующих в сети заявок, причём рост времени пребывания вначале незначителен, а затем принимает линейный характер;
- увеличение числа заявок в сети, с одной стороны, приводит к увеличению производительности (положительный фактор), а, с другой стороны – к увеличению времени пребывания заявок в сети (нежелательный фактор).
- для каждой замкнутой СеМО существует некоторое граничное значение числа заявок в сети, после которого резко увеличивается время пребывания заявок в ЗСеМО при незначительном увеличении производительности сети;

- когда загрузка узкого места становится равной единице, дальнейший рост производительности за счёт увеличения числа заявок в ЗСеМО невозможен; для увеличения производительности ЗСеМО необходимо разгрузить узкое место одним из следующих способов:
 - уменьшением длительности обслуживания заявок (увеличением скорости работы обслуживающего прибора);
 - увеличением числа обслуживающих приборов в узле;
 - уменьшением коэффициента передачи.
- если в СеМО существует несколько узлов, одновременно являющихся «узким местом», для улучшения характеристик функционирования ЗСеМО необходимо одновременно разгрузить все узкие места;
- при построении реальных систем, моделями которых служат СеМО, следует, по-возможности, строить сбалансированные системы, в которых загрузки всех узлов одинаковы.

4.7. Практикум: решение задач

Задача 1. В одноканальную систему обслуживания поступают заявки двух классов с интенсивностями 0,3 и 1 заявок в секунду. Интенсивности их обслуживания соответственно равны 0,5 и 5 заявок в секунду.

а) Сформулировать условия, при которых время пребывания заявок 1-го класса будет равно 2 секунды?

б) Чему будет равно время пребывания заявок 1-го класса, если при тех же условиях интенсивность их поступления увеличится в два раза?

в) Чему будет равно время пребывания заявок 1-го класса, если при тех же условиях интенсивность их обслуживания увеличится в два раза?

Дано: Одноканальная СМО:

- количество классов заявок: $N = 2$;
- интенсивности потоков: $\lambda_1 = 0,3 \text{ с}^{-1}$; $\lambda_2 = 1 \text{ с}^{-1}$;
- интенсивности обслуживания: $\mu_1 = 0,5 \text{ с}^{-1}$; $\mu_2 = 5 \text{ с}^{-1}$.

Требуется: а) сформулировать условия, при которых $u_1 = 2 \text{ с}$;

б) определить $u_1' = ?$ при $\lambda_1' = 2\lambda_1$;

в) определить $u_1'' = ?$ при $\mu_1'' = 2\mu_1$;

Решение.

а) Время пребывания заявок класса 1: $u_1 = w_1 + b_1$, где w_1 – время ожидания; $b_1 = 1/\mu_1 = 2 \text{ с}$ – длительность обслуживания. Очевидно, что $u_1 = 2 \text{ с}$, если $w_1 = 0$, то есть заявки 1-го класса не должны образовывать очередь. Для этого необходимо, чтобы:

- заявки 1-го класса имели абсолютный приоритет по отношению к заявкам 2-го класса; это означает, что заявки 2-го класса не смогут влиять на характеристики обслуживания заявок 1-го класса, однако это не исключает образования очереди заявок 1-го класса;
- для того чтобы заявки 1-го класса не образовывали очередь, процессы поступления и обслуживания заявок 1-го класса должны быть детерминированными, то есть интервалы между поступающими в систему заявками 1-го класса и длительности их обслуживания должны быть детерминированными (не случайными) величинами;
- нагрузка, создаваемая заявками 1-го класса не должна превышать 1, в противном случае система будет перегружена и не сможет справиться с обслуживанием заявок 1-го класса, время ожидания которых будет расти до бесконечности.

Проверим выполнение последнего условия: $y_1 = \lambda_1 / \mu_1 = 0,6$ – система работает без перегрузок. Таким образом, для того чтобы $u_1 = 2$ с, необходимо выполнение двух первых условий.

б) Определим $u_1' = ?$ при $\lambda_1' = 2\lambda_1$. Если интенсивность поступления заявок 1-го класса увеличится в 2 раза, то загрузка, то создаваемая заявками нагрузка тоже увеличится в 2 раза и станет равной $y_1' = 2\lambda_1 / \mu_1 = 1,2$, что означает перегрузку системы, следовательно, время ожидания и время пребывания заявок 1-го класса вырастут до бесконечности: $u_1' = \infty$.

Заметим, что если бы нагрузка не превысила значение 1, то время пребывания заявок 1-го класса осталось бы прежним: $u_1 = 2$ с.

в) Определим $u_1'' = ?$ при $\mu_1'' = 2\mu_1$. Увеличение интенсивности обслуживания заявок 1-го класса приведёт к уменьшению нагрузки в 2 раза: $y_1'' = \lambda_1 / (2\mu_1) = 0,3$, то есть система будет работать без перегрузки. С другой стороны, длительность обслуживания заявок тоже уменьшится в 2 раза: $b_1 = 1 / (2\mu_1) = 1$ с, следовательно, время пребывания станет равным $u_1'' = b_1'' = 1$ с.

Задача 2. Интенсивность поступления заявок в разомкнутую трехузловую СеМО равна 2 заявки в секунду. Среднее число заявок в узлах СеМО соответственно равно: 2, 4 и 6. Определить среднее время пребывания заявок в сети.

Дано: РСемо; $n = 3$; $\lambda_0 = 2$ с⁻¹; $m_1 = 2$; $m_2 = 4$; $m_3 = 6$.

Требуется: определить U .

Решение.

Среднее время пребывания заявок в СеМО определяется по формуле:

$$U = \sum_{j=1}^n \alpha_j u_j = \sum_{j=1}^n \alpha_j \frac{m_j}{\lambda_j} = \sum_{j=1}^n \alpha_j \frac{m_j}{\alpha_j \lambda_0} = \sum_{j=1}^3 \frac{m_j}{\lambda_0} = \frac{2+4+6}{2} = 6 \text{ с.}$$

Здесь последовательно применены формулы (3.29), (3.15) и (3.5) для узла j .

Этот же результат можно получить, исходя из формулы Литтла (3.31), связывающей среднее время пребывания и число заявок в сети:

$$U = \frac{M}{\lambda_0} = \frac{1}{\lambda_0} \sum_{j=1}^3 m_j = 6 \text{ с.}$$

4.8. Самоконтроль: перечень вопросов и задач**СМО с однородным потоком заявок**

1. Как зависит среднее время ожидания заявок в СМО от коэффициента вариации длительности обслуживания? Во сколько раз изменится среднее время ожидания заявок при переходе от постоянной длительности обслуживания к экспоненциально распределенной? Во сколько раз изменится среднее время ожидания при переходе от экспоненциального распределения длительности обслуживания к гиперэкспоненциальному распределению с коэффициентом вариации, равным 2?

2. Какое распределение длительности обслуживания заявок в СМО является предпочтительным для уменьшения среднего времени ожидания заявок?

3. Изменится ли разность между средним временем пребывания и средним временем ожидания заявок в СМО при изменении: а) скорости работы (быстродействия) прибора; б) интенсивности потока заявок; в) количества приборов?

4. Изменится ли разность между средним числом заявок в системе и средней длиной очереди при изменении: а) скорости работы (быстродействия) прибора; б) интенсивности потока заявок; в) количества приборов?

5. Заявки поступают в одноканальную СМО с интервалом 2,5 минуты, длительность обслуживания заявок в приборе 45 секунд. Определить загрузку и коэффициент простоя системы.

6. Интенсивность поступления заявок в трехканальную СМО – 21 заявка в секунду, интенсивность обслуживания – 10 заявок в секунду. Определить: а) вероятность того, что обслуживающий прибор работает; б) вероятность того, что обслуживающий прибор простаивает; в) среднее число заявок, находящихся на обслуживании; г) среднее число работающих приборов?

7. Интенсивность поступления заявок в четырехканальную СМО равна интенсивности обслуживания заявок одним прибором. Определить:

а) вероятность того, что система простаивает; б) среднее число простаивающих приборов; в) на какую величину среднее число заявок в системе отличается от средней длины очереди.

8. Длительность обслуживания заявок в одном приборе четырехканальной СМО равна 4 минуты. Определить предельную интенсивность поступления заявок в систему, при которой в системе существует стационарный режим?

9. Интенсивность поступления заявок в СМО – 15 заявок в секунду, длительность обслуживания одной заявки – 5 секунд. Определить число обслуживающих приборов, при котором в системе существует стационарный режим?

10. Заявки поступают в одноканальную СМО с интервалом 0,5 секунд, интенсивность обслуживания – 2,5 заявки в секунду, среднее время пребывания заявок в системе – 2 секунды. Определить среднюю длину очереди заявок.

11. Интенсивности поступления и обслуживания заявок в СМО соответственно равны 4 и 5 заявок в секунду. Определить среднее время пребывания заявок в системе, если известно, что средняя длина очереди равна 6.

12. Интенсивность обслуживания заявок в СМО равна 5 заявок в секунду. Определить предельную интенсивность поступления заявок в систему, при которой среднее время пребывания заявок в системе не превысит 1 секунду и средняя длина очереди не превысит 2.

СМО с неоднородным потоком заявок

13. При какой дисциплине обслуживания заявок средние времена ожидания заявок разных классов одинаковы?

14. При каких условиях среднее время ожидания заявок для ДО ОП является возрастающей (убывающей) функцией от номера класса заявок?

15. При каких условиях среднее время пребывания заявок для ДО ОП является возрастающей функцией от номера класса заявок?

16. Может ли среднее время пребывания заявок для ДО ОП быть убывающей функцией от номера класса заявок, если среднее время ожидания заявок является возрастающей функцией от номера класса заявок? Ответ пояснить.

17. Может ли заявка с более высоким относительным приоритетом иметь большее время пребывания, чем низкоприоритетная? Ответ обосновать.

18. При каких условиях среднее время ожидания заявок для дисциплины обслуживания с абсолютными приоритетами является возрастающей функцией от номера класса заявок?

19. В каких случаях среднее время ожидания заявок более высокого приоритета при ДО ОП может иметь большее значение, чем заявок низкого приоритета?

20. Может ли среднее время ожидания заявок быть убывающей функцией от номера класса (приоритета) заявок?

21. Почему среднее время ожидания заявок наивысшего приоритета при использовании ДОАП отличается от нуля?

22. При каких условиях характер зависимости среднего времени пребывания заявок от номера класса совпадает с характером зависимости среднего времени ожидания? Могут ли зависимости среднего времени пребывания и среднего времени ожидания заявок от номера класса иметь противоположный характер?

23. В каких случаях характер зависимости средней длины очереди заявок от номера класса отличается от характера зависимости среднего времени ожидания заявок? При каких условиях характер зависимости средней длины очереди заявок от номера класса совпадает с характером зависимости среднего времени ожидания заявок?

24. Может ли среднее число заявок в СМО отличаться от средней длины очереди заявок больше, чем на единицу? Изменится ли разность между ними при изменении дисциплины обслуживания?

25. Изменится ли разность между средним временем пребывания и средним временем ожидания для заявок суммарного потока при изменении дисциплины обслуживания?

26. Можно ли утверждать, что если при какой-либо дисциплине обслуживания значение одной из характеристик суммарного потока (например, время ожидания) меньше значений аналогичной характеристики при другой дисциплине обслуживания, то и значения остальных характеристик при первой дисциплине также будут меньше? Проиллюстрировать на примере.

27. Изменяются ли характеристики обслуживания всех классов заявок или только некоторых из них при ДО БП (ДО ОП, ДО АП), если изменить: а) интенсивность поступления заявок какого-либо класса; б) среднюю длительность обслуживания какого-либо класса заявок?

28. Нарисовать зависимости среднего времени ожидания и среднего времени пребывания заявок в СМО от номера класса заявок для беспriorитетной дисциплины обслуживания. Объяснить характер этих зависимостей.

29. Нарисовать зависимости среднего времени ожидания заявок в СМО от суммарной загрузки в случае трех классов заявок и дисциплины обслуживания с относительными и абсолютными приоритетами. Провести анализ этих зависимостей.

30. Нарисовать зависимости среднего времени ожидания и среднего времени пребывания заявок от номера класса заявок для дисциплин обслуживания с относительными и с абсолютными приоритетами. Объяснить характер этих зависимостей.

31. Физический смысл и математическая запись закона сохранения времени ожидания. Сформулировать условия, сопровождающие закон

сохранения. При каком условии в случае ДО АП не выполняется закон сохранения времени ожидания?

32. Как получается значение константы в законе сохранения времени ожидания?

33. При каком условии закон сохранения времени ожидания вырождается в закон сохранения: а) суммарной длины очереди заявок: $L = \text{Const}$; б) суммарного числа заявок в системе: $M = \text{Const}$; в) среднего времени ожидания; г) среднего времени пребывания?

34. Можно ли, изменив дисциплину обслуживания, изменить: а) время ожидания заявок всех классов; б) время пребывания заявок всех классов; в) длительность обслуживания заявок; г) время ожидания заявок только одного класса; д) время ожидания заявок только двух классов? Показать на примерах.

35. В каких случаях следует применять приоритетные дисциплины обслуживания заявок с относительными и абсолютными приоритетами?

36. Что такое "защита от перегрузок" и для каких дисциплин обслуживания она существует? Проиллюстрировать на графике свойство "защиты от перегрузок".

37. В одноканальную СМО поступают 2 простейших потока заявок с интенсивностями 0,1 и 0,2 заявок в секунду; длительности их обслуживания соответственно 2 и 4 секунды. Чему будет равно среднее время ожидания заявок 1-го класса при использовании бесприоритетной дисциплины?

38. В одноканальную СМО поступают 2 класса заявок с интенсивностями 0,1 и 0,2 заявок в секунду. Длительности их обслуживания соответственно 2 и 3 секунды. Среднее время ожидания заявок при использовании бесприоритетной дисциплины обслуживания – 5 секунд. После введения приоритетов среднее время ожидания заявок 1-го класса стало равно 2 секундам. Чему равно среднее время ожидания заявок 2-го класса?

39. В систему поступают заявки трех классов с интенсивностями 2, 1 и 0,5 заявок в секунду соответственно. Сформулировать условия, при которых среднее время пребывания заявок всех классов будет одинаково.

40. В одноканальную систему обслуживания поступают заявки двух классов с интенсивностями 0,5 и 2 заявки в секунду. Интенсивности их обслуживания соответственно равны 5 и 1,25 заявок в секунду. а) При каких условиях время пребывания заявок 2-го класса будет равно 0,8 секунды? б) Чему будет равно время пребывания заявок 2-го класса, если при тех же условиях интенсивность поступления заявок 1-го класса увеличится в два раза? в) Чему будет равно время пребывания заявок 2-го класса, если при тех же условиях интенсивность их поступления увеличится в два раза? г) Чему будет равно время пребывания заявок 2-го класса, если при тех же условиях интенсивность их обслуживания увеличится в два раза?

Разомкнутые и замкнутые СеМО

41. Что показывает коэффициент передачи в СеМО?
42. Дать физическое толкование значения коэффициента передачи узла СеМО, равное: а) 4; б) 0,4.
43. В чём различие между эквивалентным и толерантным преобразованиями СеМО? Привести пример эквивалентного преобразования СеМО.
44. Что такое «узкое место» и какие способы используются для разгрузки «узкого места»?
45. Показать на графике, как изменится зависимость производительности и среднего времени пребывания заявок в замкнутой СеМО от числа циркулирующих в сети заявок после разгрузки узкого места. В каких случаях разгрузка узкого места может не дать положительного эффекта?
46. В двухузловой замкнутой СеМО циркулирует 1 заявка. Определить загрузку узлов 1 и 2, если известно, что загрузка узла 1 в 3 раза больше, чем загрузка узла 2.
47. В замкнутой двухузловой СеМО циркулирует одна заявка, которая последовательно переходит из одного узла в другой. Длительность обслуживания в узлах распределена по экспоненциальному закону с одним и тем же средним значением, равным 5 минут. По какому закону распределено время пребывания заявки в сети? Определить производительность замкнутой СеМО.
48. В замкнутой двухузловой СеМО циркулирует одна заявка, которая последовательно переходит из одного узла в другой. Длительности обслуживания в узлах 1 и 2 сети соответственно равны 2 и 3 с. Определить: а) коэффициенты простоя узлов замкнутой СеМО; б) среднее число заявок, находящихся в каждом из узлов СеМО.
49. В замкнутой двухузловой СеМО циркулирует 4 заявки, которые последовательно переходят из одного узла в другой. Длительности обслуживания заявок в узлах сети одинаковы и равны 2 с. Среднее время ожидания заявок в узле 1 равно 3 с. Определить: а) производительность замкнутой СеМО; б) загрузку узлов сети; в) среднее число заявок, находящихся в состоянии ожидания.
50. В разомкнутую СеМО поступают заявки с интервалом 5 секунд. Время пребывания заявок в сети равно 15 секунд. Определить среднее число заявок в сети и интенсивность выходящего из сети потока заявок.
51. Средние времена ожидания заявок в узлах трехузловой СеМО соответственно равны: 1, 2 и 4 секунды, а коэффициенты простоя узлов равны 0,8; 0,4; 0,7. Определить среднее время ожидания заявок в сети, если известно, что длительности обслуживания заявок во всех узлах одинаковы и коэффициент передачи узла 1 равен 2.
52. Известны вероятности состояний двухузловой замкнутой СеМО: $P(0,4)=0,1$; $P(1,3)=0,4$; $P(2,2)=0,2$; $P(3,1)=0,1$; $P(4,0)=0,2$, где состояние (i_1, i_2) задает число заявок в одноканальном узле 1 и трехканальном узле 2

соответственно. Определить среднее число заявок в СеМО, находящихся в состоянии ожидания.

53. Известны вероятности состояний трехузловой замкнутой СеМО: $P(0,0,2)=0,1$; $P(0,1,1)=0,2$; $P(0,2,0)=0,15$; $P(1,0,1)=0,35$; $P(1,1,0)=0,05$; $P(2,0,0)=0,15$, где состояние (i_1, i_2, i_3) задает число заявок в узле 1, 2, 3 соответственно. Определить среднее число параллельно работающих узлов сети.

54. Известны вероятности состояний трехузловой замкнутой СеМО: $P(0,0,2)=0,1$; $P(0,1,1)=0,3$; $P(0,2,0)=0,4$; $P(1,0,1)=0,05$; $P(1,1,0)=0,05$; $P(2,0,0)=0,1$. Длительности обслуживания заявок во всех одноканальных узлах одинаковы. Определить значения коэффициентов передач второго и третьего узлов сети, если известно, что коэффициент передачи первого узла равен 2.

55. Известны вероятности состояний трехузловой замкнутой СеМО: $P(0,0,2)=0,1$; $P(0,1,1)=0,3$; $P(0,2,0)=0,4$; $P(1,0,1)=0,05$; $P(1,1,0)=0,05$; $P(2,0,0)=0,1$. Определить производительность СеМО, если известно, что коэффициент передачи первого узла (четырехканального) равен 2, а средняя длительность обслуживания заявок в этом узле равна 0,1 с.