ТЕМА 1
ОСНОВНЫЕ ПОНЯТИЯ И ВЕРОЯТНОСТНЫЙ АППАРАТ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

1.1. Основные понятия и определения ТМО

Теория массового обслуживания – прикладная математическая дисциплина, целью исследований которой является рациональный выбор структуры системы обслуживания и процесса обслуживания.

Определение

Заявкой называется спрос на удовлетворение какой-либо потребности (далее потребности предполагаются однотипными). Выполнение заявки называется обслуживанием заявки.

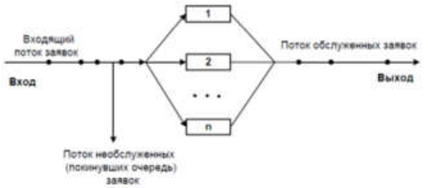
Определение

Системой массового обслуживания (СМО) называется любая система для выполнения заявок, поступающих в неё в случайные моменты времени.

Определение

Каналом обслуживания называется устройство в СМО, обслуживающее заявку.

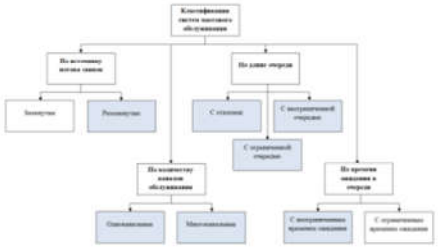
Схема СМО изображена на рисунке 1.



СМО считается заданной, если описаны следующие ее параметры:

- входящий поток заявок;
- время обслуживания заявок;
- структура системы;
- емкости накопителей (буферов);
- дисциплина обслуживания,

Классификация СМО представлена на схеме



2.1. Вероятностный аппарат ТМО. Случайные процессы

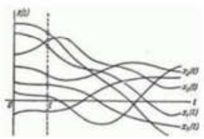
Определение

Случайным процессом $X(t)$ называют функцию, аргументом которой является время $t \in T$, а принимаемые ею значения являются случайными величинами.

Определение

Функция, полученная в результате наблюдения над случайным процессом, называется реализацией случайного процесса.

Пусть получено n реализаций случайного процесса: $x_1(t), x_2(t), \ldots, x_n(t)$.



Определение

Случайная величина $X(t_0)$, в которую обращается случайный процесс при фиксированном $t = t_0$, называется сечением случайного процесса.

Определение

Случайный процесс $X(t)$ называется **процессом с дискретным временем**, если система, в которой он протекает, меняет свои состояния только в заранее известные моменты времени, число которых конечно или счётно.

Случайный процесс называется процессом **с непрерывным временем**, если переход из состояния в состояние может происходить в любой случайный момент времени.

Определение

Случайный процесс называется **процессом с непрерывными состояниями**, если сечением случайного процесса является непрерывная случайная величина.

Случайный процесс называется **случайным процессом с дискретными состояниями**, если сечением случайного процесса является дискретная случайная величина

2.2. Вероятностные характеристики случайного процесса

Определение

Математическим ожиданием случайного процесса $X(t)$ называется неслучайная функция $m_X(t)$, которая при каждом значении аргумента t равна математическому ожиданию соответствующего сечения случайной функции.

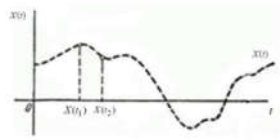
Определение

Дисперсией случайного процесса $X(t)$ называют неслучайную неотрицательную функцию $D_X(t)$, значение которой для каждого аргумента t равно дисперсии соответствующего сечения случайной функции.

Часто вместо дисперсии случайного процесса рассматривают среднее квадратическое отклонение случайного процесса

$$\sigma_x(t) = \sqrt{D_X(t)}\,.$$

Пусть имеется случайный процесс $X(t)$. Рассмотрим два ее сечения $X(t_1)$ и $X(t_2)$.



Определение

Корреляционной функцией случайного процесса $X(t)$ называется неслучайная функция $K_X(t_1,t_2)$, которая для каждой пары моментов времени t_1 и t_2 равна математическому ожиданию произведения соответствующих сечений $X(t_1)$ и $X(t_2)$:

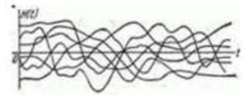
$$K_x(t_1,t_2) = M\left[\,(X(t_1) - m_x(t_1))\cdot(X(t_2) - m_x(t_2))\right]$$

Свойства корреляционной функции:

1. $K_x(t_1,t_1) = D_x(t_1)$.
2. $K_x(t_1,t_2) = K_x(t_2,t_1)$

2.3. Стационарные случайные процессы

На практике очень часто встречаются случайные процессы, протекающие во времени приблизительно однородно. Например, колебания напряжения в электрической осветительной сети; случайные шумы в радиоприемнике.



Стационарные случайные процессы очень часто встречаются в технических задачах. По своей природе эти процессы проще, чем нестационарные, и описываются более простыми характеристиками.

Определение

Случайный процесс $X(t)$ называют стационарным, если все его вероятностные характеристики не меняются при любом изменении аргумента t, т.е.

1. $m_x(t) = m_x = const$,
2. $D_x(t) = D_x = const$,
3. $K_x(t,t+\tau) = k_x(\tau)$.

Полагая в условии (3) $\tau = 0$ имеем

$$K_x(t,t) = D_x(t) = k_x(0) = const$$

Условие (3) есть единственное существенное условие, которому должна удовлетворять стационарная случайная функция.

Т.к. корреляционная функция любой случайной функции обладает свойством симметрии, то

$$K_x(t,t+\tau) = K_x(t+\tau,t).$$

Отсюда

$$k_x(\tau) = k_x(-\tau)\,.$$

Для количественной характеристики степени зависимости стационарных случайных процессов вводят безразмерную величину, называемую **нормированной корреляционной функцией**. Она определяется выражением:

$$r_x(t_1,t_2) = \frac{K_x(t_1,t_2)}{\sigma_x(t_1)\cdot\sigma_x(t_2)}$$

Нормированная корреляционная функция связи при фиксированных значениях аргументов представляет собой коэффициент корреляции соответствующих сечений случайного процесса.

Тема 2
ДИСКРЕТНЫЕ МАРКОВСКИЕ ЦЕПИ

1. Понятие марковского случайного процесса. Дискретная цепь Маркова.

Определение

Случайный процесс, протекающий в некоторой системе, называется марковским или процессом без последействия, если для каждого момента времени поведение системы в будущем зависит только от состояния системы в данный момент времени и не зависит от того, каким образом система пришла в это состояние.

Иными словами, в марковских случайных процессах «будущее» состояние целиком определяется его «настоящим», «прошлое» на него никак не влияет.

Это свойство называется свойством отсутствия последействия или марковским свойством.

Это свойство было названо в честь русского математика Андрея Маркова.

Пусть имеется физическая система S , которая может находиться в состояниях:

$$\boldsymbol{S_1, S_2, \ldots, S_n} \text{ – состояния системы}$$

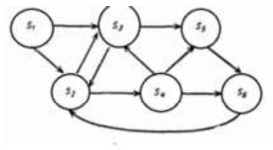
Пусть переходы системы из одного состояния в другое происходят мгновенно (скачкообразно) и только в заранее фиксированные моменты времени (шаги), которые можно пронумеровать: $k = 0, 1, 2, \ldots$

Таким образом, в системе протекает случайный процесс с дискретными состояниями и дискретным временем.

Определение

Марковский случайный процесс с дискретными состояниями и дискретным временем называют дискретной марковской цепью (или дискретной цепью Маркова).

Марковскую цепь изображают в виде графа переходов, вершины которого соответствуют состояниям цепи, а дуги — переходам между ними.



Случайный процесс, который происходит в системе, состоит в том, что в последовательные моменты времени (шаги)

$$k = 0,1,2,\ldots$$

система оказывается в тех или других состояниях, ведя себя, например, следующим образом

$$S_1 \to S_3 \to S_2 \to S_4 \to S_6 \to S_2 \to$$

В общем случае в эти моменты времени система может не только менять состояние, но и оставаться в прежнем.

Обозначим через

A_{ki} – событие, состоящее в том, что на k-м шаге система будет находиться в состоянии S_i ($i = 1,2,\ldots,n$, $k = 0,1,2,\ldots$)

Тогда процесс, происходящий в системе, можно представить как последовательную цепочку событий, например:

$$A_{01}, A_{13}, A_{22}, A_{34}, A_{46}, A_{52}\ldots$$

Определение

Вероятность события A_{ki}, т.е. вероятность того, что система на i-м шаге будет находиться в состоянии S_i , называют вероятностью i-го состояния системы на k-м шаге и обозначают $p_i(k)$.

Вероятности состояний системы на k-м шаге в совокупности образуют вектор:

$$P(k) = [\ p_1(k),\ p_2(k), \ldots, p_n(k)\ ].$$

Определение

Вектор $P(0) = [\ p_1(0),\ p_2(0), \ldots, p_n(0)\ ]$, где $p_i(0)$ – вероятность появления состояния S_i в начальном испытании, называют вектором начальных вероятностей.

Очевидно, что на любом шаге k процесс может находиться в одном и только одном из n возможных состояний

$$S_1, S_2, \ldots, S_n.$$

Следовательно, при любом k события

$$A_{k1},\ A_{k2}, \ldots, A_{kn}$$

единственно возможны и несовместны, т.е. образуют полную группу.

Поэтому для каждого шага k имеет место равенство

$$p_1(k) + p_2(k) + \cdots + p_n(k) = 1.$$

Для любого шага существуют какие-то вероятности перехода системы из любого состояния в любое другое (некоторые из них равны нулю, если непосредственный переход за один шаг невозможен), а также вероятность задержки системы в данном состоянии.

Будем называть эти вероятности переходными вероятностями марковской цепи.

Определение

Марковскую цепь называют однородной, если переходные вероятности не зависят от номера шага. В противном случае марковскую цепь называют неоднородной.

Будем считать марковскую цепь однородной.

Обозначим через p_{ij} вероятность перехода системы S за один шаг из состояния S_i в состояние S_j . Из переходных вероятностей p_{ij} можно составить матрицу.

Определение

Матрицей перехода за один шаг называют матрицу, которая содержит все переходные вероятности этой системы за один шаг:

$$\pi_1 = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}.$$

Свойства матрицы перехода:

1. Некоторые из переходных вероятностей p_{ij} могут быть равны нулю, что означает невозможность перехода системы из i-го состояния в j-е за один шаг.

2. По главной диагонали матрицы переходных вероятностей стоят вероятности того, что система останется в текущем состоянии.

3. В каждой строке матрицы помещены вероятности событий, которые образуют полную группу. Сумма вероятностей этих событий равна единице. Поэтому сумма элементов каждой строки матрицы равна единице.

$$\sum_{j=1}^{n} p_{ij} = 1 \qquad (i = 1,2,\dots,n).$$

2. Равенство Маркова

Рассмотрим следующую задачу:

Имеется физическая система S, которая может находиться в состояниях:

$$S_1, S_2, \dots, S_n$$

В системе протекает дискретный марковский процесс.

Известны p_{ij} – вероятности перехода системы из состояния S_i в состояние S_j за один шаг.

Требуется найти вероятности перехода системы из состояния S_i в состояние S_j за k шагов.

Обозначим эти вероятности через $p_{ij}(k)$. По формуле полной вероятности

$$p_{ij}(k) = \sum_{r=1}^{n} p_{ir}(m) \cdot p_{rj}(k-m) \qquad \text{– равенство Маркова}$$

В общем случае

$$\pi_k = \pi_1^k \ , \forall k$$

Если известны вектор начального распределения вероятностей и матрица переходных вероятностей, то можно вычислить вероятности состояния системы для любого шага k:

$$P(k) = P(0) \cdot \pi_1^k$$

Цепь Маркова считается заданной, если:

1. Имеется вектор начальных вероятностей, описывающий начальное состояние системы.
2. Известна матрица переходных вероятностей.

3. Достаточное условие эргодичности

Теорема:

Если существует такое число $m > 0$, при котором все элементы матрицы π_m переходов за m шагов положительны, то существуют такие постоянные числа p_j ($j = 1,2,\dots n$), что

$$\lim_{k\to\infty} p_{ij}(k) = p_j.$$

Величины p_j называются предельными (финальными) вероятностями системы. Предельные вероятности характеризуют среднюю долю времени, в течение которого система находится в данном состоянии при наблюдении в течение достаточно продолжительного времени.

Заметим, что

$$p_1 + p_2 + \dots + p_n = 1 \tag{1}$$

Обозначим $\mathrm{P} = (p_1, p_2, \dots, p_n)$.

При $k \to \infty$ имеем

$$\mathrm{P} = \mathrm{P} \cdot \pi_1. \tag{2}$$

Записанная в матричном виде система (2) является системой линейных уравнений с n неизвестными $p_1, p_2, \dots, p_n$. Учитывая условие (1), одно из уравнений системы (2.2) можно отбросить.

Отбросим последнее уравнение и запишем систему уравнений в явном виде

$$\begin{cases} \sum_{i=1}^{n} p_i \cdot p_{ij} = p_j & (j = 1,2,\dots,n-1), \\ \sum_{i=1}^{n} p_i = 1. \end{cases} \tag{2.3}$$

Решая данную систему, определяют значения предельных вероятностей $p_1, p_2, \dots, p_n$.

Тема 3

НЕПРЕРЫВНЫЕ МАРКОВСКИЕ ЦЕПИ

1. Понятие непрерывной марковской цепи

Определение

Марковский случайный процесс с дискретными состояниями и непрерывным временем называется непрерывной цепью Маркова.

Рассмотрим некоторую системы при следующих допущениях:

- система в процессе функционирования может принимать состояния $S_1, S_2, \dots, S_n$;
- переход системы из состояния в состояние может осуществляться в любой момент времени;
- процесс смены состояний является марковским.

Таким образом, процесс смены состояний системы описывается непрерывной марковской цепью.

Пусть $S_1, S_2, \dots, S_n$ – всевозможные состояния некоторой физической системы S. Переход системы из одного состояния в другое может осуществляться в любой момент времени.

Определение

Вероятность того, что система в момент времени t будет находиться в состоянии S_i, называют вероятностью i-го состояния системы в момент времени t и обозначают как $p_i(t)$, $i = 1, 2, \dots, n$

Так как в любой момент времени t система будет находиться в одном и только одном из состояний $S_1, S_2, \dots, S_n$, то имеет место нормировочное условие

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

Придадим t приращение $\Delta t \neq 0$. Обозначим через $p_{ij}(\Delta t)$ вероятность перехода системы из состояния S_i в состояние S_j за время $t + \Delta t$.

Определение.

Интенсивностью перехода системы из состояния S_i в состояние S_j за время Δt называют предел отношения переходной вероятности к приращению Δt при условии, что $\Delta t \to 0$:

$$\lambda_{ij}(t) = \lim_{\Delta t\to 0} \frac{p_{ij}(\Delta t)}{\Delta t}.$$

$$\frac{p_{ij}(\Delta t)}{\Delta t} = \lambda_{ij}(t) + \alpha(\Delta t),$$

где $\alpha(\Delta t)$ - бесконечно малая величина при $\Delta t \to 0$, откуда

$$p_{ij}(\Delta t) = \lambda_{ij}(t) \cdot \Delta t + \alpha(\Delta t) \cdot \Delta t,$$

Тогда с точностью до бесконечно малых высших порядков можно записать

$$p_{ij}(\Delta t) \approx \lambda_{ij}(t) \cdot \Delta t.$$

Из определения интенсивностей переходов видно, что они в общем случае зависят от времени, неотрицательны и в отличие от вероятностей могут быть больше 1.

Определение

Непрерывная цепь Маркова называется однородной, если для любых $i, j = \overline{1,n}$

$$\lambda_{ij}(t) = \lambda_{ij} = const$$

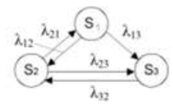
В дальнейшем будем рассматривать непрерывную однородную цепь Маркова.

5. Уравнения Колмогорова

Рассмотрим вывод уравнений Колмогорова на примере.

Пусть система S имеет три возможных состояния S_1, S_2, S_3

Пусть для всех пар состояний S_i и S_j известны интенсивности переходов λ_{ij} и λ_{ji}.



Требуется определить для любого момента времени t вероятности состояний: $p_1(t), p_2(t), \dots, p_n(t)$.

Зафиксируем момент времени t и найдем вероятность $p_1(t + \Delta t)$ того, что в момент $t + \Delta t$ система будет находиться в состоянии S_1. Это может произойти двумя способами: $A = A_1 \cdot A_2$ или $B = B_1 \cdot B_2$.

По теореме сложения вероятностей

$$p_1(t + \Delta t) = P(A) + P(B).$$

$$P(A) = P(A_1) \cdot P_{A_1}(A_2),$$

$$P(A) = p_1(t) \cdot \left(1 - \left(p_{12}(\Delta t) + p_{13}(\Delta t)\right)\right)$$

$$P(A) = p_1(t) \cdot \left(1 - (\lambda_{12} \cdot \Delta t + \lambda_{13} \cdot \Delta t)\right)$$

Аналогично найдем вероятность события B:

$$P(B) = P(B_1 \cdot B_2) = P(B_1) \cdot P_{B_1}(B_2) = p_2(t) \cdot p_{21}(\Delta t).$$

$$P(B) = p_2(t) \cdot \lambda_{21} \cdot \Delta t$$

Отсюда

$$p_1(t + \Delta t) = p_1(t) \cdot \left(1 - (\lambda_{12} \cdot \Delta t + \lambda_{13} \cdot \Delta t)\right) + p_2(t) \cdot \lambda_{21} \cdot \Delta t$$

$$p_1(t + \Delta t) = p_1(t) - \lambda_{12} \cdot \Delta t \cdot p_1(t) - \lambda_{13} \cdot \Delta t \cdot p_1(t) + \lambda_{21} \cdot \Delta t \cdot p_2(t)$$

$$\frac{p_1(t + \Delta t) - p_1(t)}{\Delta t} = -\lambda_{12} \cdot p_1(t) - \lambda_{13} \cdot p_1(t) + \lambda_{21} \cdot p_2(t)$$

Перейдем к пределу при $\Delta t \to 0$:

$$\lim_{\Delta t\to 0} \frac{p_1(t + \Delta t) - p_1(t)}{\Delta t} =$$

$$= p_1'(t) = -\lambda_{12} \cdot p_1(t) - \lambda_{13} \cdot p_1(t) + \lambda_{21} \cdot p_2(t)$$

Аналогичные дифференциальные уравнения могут быть составлены и для других вероятностей состояний.

В общем случае, вероятности состояний $p_i(t)$, $i = 1, 2, \dots, n$ являются решением следующей системы n линейных однородных дифференциальных уравнений:

$$p_i'(t) = -\left(\sum_{\substack{j=1 \\ j\neq i}}^{n} \lambda_{ij}\right) p_i(t) + \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(\lambda_{ji} \cdot p_j(t)\right),$$

$$i = 1, 2, \dots, n.$$

Эта система называется системой дифференциальных уравнений Колмогорова.

Для однозначного решения системы должны быть заданы начальные значения: $p_1(0), p_2(0), \dots, p_n(0)$, причем

$$\sum_{i=1}^{n} p_i(0) = 1.$$

Составить систему Колмогорова удобно по размеченному графу состояний.

Правило построения

1. В левой части каждого уравнения стоит производная от вероятности состояния, а правая часть содержит столько членов, сколько стрелок связано с данным состоянием.
2. Если стрелки направлены из состояния, соответствующий член имеет знак «–» ,если в состоянии то «+».
3. Каждый член равен произведению интенсивности перехода на вероятность того состояния, из которого выходит стрелка.

3. Предельные вероятности состояния

Для эргодических однородных марковских цепей существует стационарный режим при $t \to \infty$. При стационарном режиме вероятности состояний стремятся к некоторым установившимся значениям – предельным вероятностям, которые постоянны и не зависят от начального состояния системы.

Теорема

Если число n состояний системы конечно и из каждого состояния можно перейти в любое другое за конечное время, то существуют предельные (финальные) вероятности состояний:

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

Очевидно, предельные вероятности состояний в сумме должны давать единицу:

$$\sum_{i=1}^{n} p_i = 1.$$

Предельные вероятности определяют среднее относительное время пребывания системы в данном состоянии.

Для того чтобы вычислить предельные вероятности состояний, достаточно в системе уравнений Колмогорова, описывающих вероятности состояний, приравнять все левые части (производные) к нулю (поскольку в установившемся режиме все вероятности состояний постоянны).

Для вычисления предельных вероятностей нужно в уравнениях Колмогорова положить $p_i'(t) = 0$. Получаем систему однородных линейных алгебраических уравнений:

$$\left(\sum_{\substack{j=1 \\ j\neq i}}^{n} \lambda_{ij}\right) p_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(\lambda_{ji} \cdot p_j\right), \qquad i = 1,2,\dots,n$$

которые совместно с нормировочным условием

$$\sum_{i=1}^{n} p_i = 1.$$

позволяют вычислить все предельные вероятности состояний

Тема 4

ПРОЦЕССЫ «ГИБЕЛИ И РАЗМНОЖЕНИЯ»

1. Понятие процесса «гибели и размножения»

В теории массового обслуживания широко распространен специальный класс случайных процессов – так называемые процессы гибели и размножения. Название это связано с рядом биологических задач, где этот процесс служит математической моделью изменения численности биологических популяций.

Определение

Непрерывная марковская цепь называется процессом «гибели и размножения», если ее граф состояний можно представить в виде одной цепочки, в которой каждые из средних состояний $S_1, S_2, \dots, S_{n-1}$ связано прямой и обратной связью с каждым из средних состояний, а крайние состояния S_0, S_n – только с одним соседним состоянием.

Пример

Техническое устройство состоит из трех одинаковых узлов, каждый из которых может выходить из строя (отказывать). Отказавший узел начинает немедленно восстанавливаться. Состояние системы нумеруем по числу неисправных узлов.

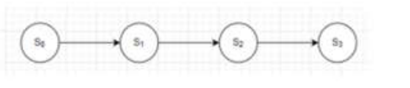
S_0 – все три узла исправны;

S_1 один узел отказал (восстанавливается), два исправны;

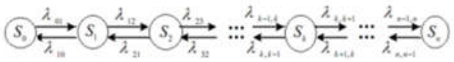
S_2 два узла восстанавливаются, один исправен;

S_3 — все три узла восстанавливаются.

Граф состояний показан на рис. 1. Из графа видно, что процесс, протекающий в системе, представляет собой процесс «гибели и размножения».



В общем случае граф состояний процесса «гибели и размножения» имеет следующий вид (рис. 1).



2. Описание процесса «гибели и размножения»

Пусть для всех пар состояний S_i и S_j известны интенсивности переходов λ_{ij} и λ_{ji}. Составим по графу состояний систему уравнений Колмогорова.

$$\begin{cases} p_0'(t) = -\lambda_{01}\cdot p_0(t) + \lambda_{10}\cdot p_1(t) \\ p_1'(t) = \lambda_{01}\cdot p_0(t) - (\lambda_{10}+\lambda_{12})\cdot p_1(t) + \lambda_{21}\cdot p_2(t) \\ \dots\dots\dots \\ p_k'(t) = \lambda_{k-1,k}\cdot p_{k-1}(t) - (\lambda_{k,k-1}+\lambda_{k,k+1})\cdot p_k(t) + \lambda_{k+1,k}\cdot p_{k+1}(t) \\ \dots\dots\dots \\ p_n'(t) = \lambda_{n-1,n}\cdot p_{n-1}(t) - \lambda_{n,n-1}\cdot p_n(t) \end{cases}$$

В стационарном режиме при $t\to\infty$ имеем: $p_k(t)\to p_k$, $p_k'(t)\to 0$. Поэтому система дифференциальных уравнений вырождается в систему линейных алгебраических уравнений.

Составим алгебраические уравнения для вероятностей состояний. Для состояния S_0 имеем:

$$\lambda_{01}\cdot p_0 = \lambda_{10}\cdot p_1 . \qquad (*)$$

Для состояния S_1

$$\lambda_{10}\cdot p_1 + \lambda_{12}\cdot p_1 = \lambda_{01}\cdot p_0 + \lambda_{21}\cdot p_2$$

В силу (*), можно сократить справа и слева равные друг другу члены $\lambda_{01}p_0$ и $\lambda_{10}p_1$; получим:

$$\lambda_{12}\cdot p_1 = \lambda_{21}\cdot p_2 .$$

Рассуждая аналогично, получаем систему однородных линейных алгебраических уравнений

$$\begin{cases} \lambda_{01}p_0 = \lambda_{10}p_1 \\ \lambda_{12}p_1 = \lambda_{21}p_2 \\ \dots\dots\dots \\ \lambda_{k-1,k}p_{k-1} = \lambda_{k,k-1}p_k \\ \dots\dots\dots \\ \lambda_{n-1,n}p_{n-1} = \lambda_{n,n-1}p_n \end{cases}$$

к которой добавляется нормировочное условие

$$\sum_{i=0}^{n} p_i = 1.$$

Решим эту систему уравнений. Из первого уравнения выразим p_1 через p_0 :

$$p_1 = \frac{\lambda_{01}}{\lambda_{10}} p_0$$

Из второго получим:

$$p_2 = \frac{\lambda_{12}}{\lambda_{21}} p_1 = \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} p_0$$

и вообще, для любого $k = \overline{1,n}$:

$$p_k = \frac{\lambda_{k-1,k}\dots\lambda_{12}\lambda_{01}}{\lambda_{k,k-1}\dots\lambda_{21}\lambda_{10}} p_0$$

В формулах для вероятностей $p_1, p_2, \dots, p_n$ числители представляют собой произведения всех интенсивностей, стоящих у стрелок, ведущих слева направо (от начала и до данного состояния S_k); знаменатели – произведения всех интенсивностей, стоящих у стрелок, ведущих справа налево (из состояния S_k и до начала).

Таким образом, все предельные вероятности состояний выражены через p_0. Подставим эти выражения в нормировочное условие.

Получим, вынося за скобки p_0:

$$p_0\left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \dots + \frac{\lambda_{n-1,n}\dots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\dots\lambda_{21}\lambda_{10}}\right) = 1$$

откуда можно получить выражение для p_0:

$$p_0 = \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \dots + \frac{\lambda_{n-1,n}\dots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\dots\lambda_{21}\lambda_{10}}\right)^{-1}$$

Таким образом, найдены предельные вероятности состояний.

ПРОСТЕЙШИЙ ПОТОК ЗАЯВОК. ВРЕМЯ ОБСЛУЖИВАНИЯ. ВРЕМЯ ОЖИДАНИЯ

1. Простейший поток заявок

Входящий поток заявок во многом определяет характеристики производительности функционирования СМО. Поэтому правильное описание потока заявок, поступающих в случайные моменты времени в реальную систему, является весьма важной задачей.

Определение

Случайным потоком называется последовательность случайных моментов наступления некоторых событий (например, поступления вызовов на АТС, прибытия самолетов в аэропорт, отказов элементов и т.д.).

Случайный поток событий по существу является случайным процессом. Заявки, поступающие в СМО или покидающие систему, образуют случайный поток. Такой поток можно изобразить как последовательность точек на оси времени, соответствующих случайным моментам появления заявок (рис. 4.1):

Рисунок 4.1 – Поток заявок

Среди свойств, которыми могут обладать случайные потоки событий, выделяют свойства стационарности, отсутствия последействия и ординарности.

Определение

Поток событий называется **стационарным**, если вероятность наступления какого-либо числа событий на интервале времени зависит только от длины этого интервала и не зависит от того, где на оси времени взят этот интервал.

Условию стационарности удовлетворяет поток заявок, вероятностные характеристики которого не зависят от времени.

Определение

Поток событий называется **потоком без последействия**, если для любой пары непересекающихся интервалов времени число событий, попадающих на один из них, не зависит от числа событий, попадающих на другой.

Условие отсутствия последействия означает, что заявки поступают в СМО в те или иные моменты времени независимо друг от друга.

Определение

Поток событий называется **ординарным**, если пренебрежимо мала вероятность того, что на малый интервал времени попадет больше одного события.

Условие ординарности означает, что заявки поступают в СМО поодиночке, а не группами.

Определение

Поток событий называется **простейшим** (или **стационарным пуассоновским**), если он одновременно стационарен, ординарен и не имеет последействия.

Определение

Интенсивностью простейшего потока называют среднее число заявок, поступающих в систему в единицу времени.

Для простейшего потока справедливы две теоремы.

Теорема 1

Число событий, произошедших в простейшем потоке, есть случайная величина, распределённая по закону Пуассона. Вероятность того, что в простейшем потоке с интенсивностью λ за интервал времени τ поступит ровно k заявок, равна

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

В частности, вероятность того, что за интервал времени τ не поступит ни одной заявки, равна

$$P_0(\tau) = e^{-\lambda\tau}$$

Теорема 2

Интервал времени l между событиями в простейшем потоке есть случайная величина, распределенная по показательному закону. Вероятность того, что интервал времени l между поступлениям заявок меньше некоторой величины t, равна:

$$P(l < t) = F(t) = 1 - e^{-\lambda t} \qquad (t > 0) .$$

Если входящий поток заявок и все потоки обслуживания простейшие, то процесс, протекающий в СМО, является марковским случайным процессом.

2. Время обслуживания. Время ожидания

Одной из важнейших характеристик СМО, которая определяет пропускную способность всей системы, является время обслуживания, т.е. время пребывания одной заявки в канале обслуживания,

В теории массового обслуживания время обслуживания ($t_{обсл}$) считают случайной величиной. Время обслуживания зависит как от стабильности работы самих каналов обслуживания, так и от параметров поступающих в систему заявок (к примеру, различной грузоподъемности транспортных средств, поступающих под погрузку). В общем случае время обслуживания может изменяться в большом диапазоне.

Случайная величина $t_{обсл}$ полностью характеризуется законом распределения вероятностей, который устанавливает соответствие между возможными значениями случайной величины и вероятностями их появления. Закон распределения вероятностей определяется на основе статистических испытаний.

Однако $t_{обсл}$ – непрерывная случайная величина, а множество значений непрерывной случайной величины бесконечно и несчетно. Поэтому вероятность того, что непрерывная случайная величина примет какое-то конкретное значение, равна нулю.

Однако можно вычислить вероятность того, что непрерывная случайная величина примет значение, принадлежащее заданному интервалу (a,b), и задать закон распределения вероятностей с помощью функции плотности распределения вероятностей. Эта функция характеризует плотность, с которой распределяются значения случайной величины в данной точке.

В теории массового обслуживания обычно считают, что непрерывная случайная величина $t_{обсл}$ имеет, как правило, показательный закон распределения с плотностью распределения вероятностей

$$g(t) = \mu e^{-\mu t} \qquad (t > 0).$$

Величину $\mu = \frac{1}{\bar{t}_{обсл}}$, где $\bar{t}_{обсл}$ - среднее время обслуживания, называют интенсивностью обслуживания.

Показательный закон распределения времени обслуживания имеет место тогда, когда плотность распределения вероятностей резко убывает с возрастанием времени.

Например, когда основная масса заявок обслуживается быстро, а продолжительное обслуживание встречается редко. Наличие показательного закона распределения времени обслуживания устанавливается на основе статистических наблюдений.

Время ожидания (время пребывания заявки в очереди, если последняя существует), также считают случайной величиной, имеющей, как правило, показательное распределение с плотностью распределения вероятностей

$$h(t) = \nu e^{-\nu t} \qquad (t > 0).$$

Величина $\nu = \frac{1}{\bar{t}_{ож}}$, где $\bar{t}_{ож}$ – среднее время ожидания обслуживания, характеризует интенсивность потока «уходов заявок», стоящей в очереди.

Тема 6

СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С ОТКАЗАМИ

1. Понятие системы массового обслуживания с отказами. Уравнения Эрланга

В СМО с отказами заявка, поступившая в момент, когда все каналы заняты, немедленно получает отказ, покидает систему и в дальнейшем процессе обслуживания не участвует.

Рассмотрим СМО с отказами, удовлетворяющую следующим условиям:

1. Система имеет n каналов обслуживания.
2. На вход СМО поступает простейший поток заявок с интенсивностью λ.
3. Время обслуживания одной заявки – случайная величина, имеющая показательное распределение. Обслуживание заявок **каждым каналом** СМО осуществляется с интенсивностью $\mu = \frac{1}{\bar{t}_{обсл}}$, где $\bar{t}_{обсл}$ - среднее время обслуживания.
4. При занятости всех n каналов вновь пришедшая заявка получает отказ и покидает СМО.

Так как поток заявок и поток освобождений канала – простейшие, то процесс, протекающий в системе, будет марковским. Состояния СМО нумеруются по числу заявок, находящихся в СМО (в силу отсутствия очереди, оно совпадает с числом занятых каналов):

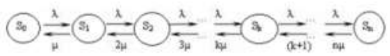
S_0 – свободны все каналы,

S_1 – занят один канал,

……………………………………

S_n – заняты все n каналов.

Граф состояний СМО с отказами имеет вид



Определим интенсивности поток событий, переводящих систему из одного состояния в другое. Поток заявок последовательно переводит систему из любого левого состояния в соседнее правое с одной и той же интенсивностью λ. Интенсивность же потока обслуживаний, переводящих систему из любого правого состояния в соседнее левое состояние, постоянно меняется в зависимости от состояния. Действительно, если СМО находится в состоянии S_2 (два канала заняты), то она может перейти в состояние S_1 (один канал занят), когда закончит обслуживание либо первый, либо второй канал, т.е. суммарная интенсивность их потоков обслуживании будет равна 2μ. Аналогично, суммарный поток обслуживаний, переводящий СМО из состояния S_3 (три канала заняты) в S_2, будет иметь интенсивность 3μ, т.е. может освободиться любой из трех каналов и т.д.

Из рисунка видно, что процесс, протекающий в СМО, представляет собой частный случай процесса «гибели и размножения». Пользуясь **общими правилами, можно составить уравнения Колмогорова** для вероятностей состояний $p_k(t)$, $k=0,1,...,n$.

$$\begin{cases} p_0'(t) = -\lambda \cdot p_0(t) + \mu \cdot p_1(t), \\ \dots\dots\dots \\ p_k'(t) = \lambda \cdot p_{k-1}(t) - (\lambda + k\cdot\mu)\cdot p_k(t) + (k+1)\cdot\mu\cdot p_{k+1}(t) \\ \qquad (1 \le k \le n-1), \\ \dots\dots\dots \\ p_n'(t) = \lambda \cdot p_{n-1}(t) - n\cdot\mu\cdot p_n(t). \end{cases} \qquad (*)$$

Уравнения (*) называются уравнениями Эрланга. Интегрировать систему уравнений (*) надо при начальных условиях, описывающих состояние S_0 – «в начальный момент все каналы свободны»:

$$p_0(0)=1;$$

$$p_1(0)=p_2(0)=...=p_n(0).$$

Для любого момента времени должно выполняться условие

$$\sum_{k=0}^{n} p_k(t) = 1.$$

Вероятности $p_k(t)$ характеризуют среднюю загрузку системы и ее изменение с течением времени.

В частности, $p_n(t)$ есть вероятность того, что заявка, пришедшая в момент t, застанет все каналы занятыми (получит отказ):

$$P_{отк} = p_n(t).$$

Величина $q(t) = 1 - p_n(t)$ называется относительной пропускной способностью системы. Для данного момента t это есть отношение среднего числа обслуженных за единицу времени заявок к среднему числу поданных.

2. Стационарный режим обслуживания СМО с отказами. Формулы Эрланга

Пусть дана n-канальная система массового обслуживания с отказами. На вход системы поступает простейший поток заявок с интенсивностью λ; время обслуживания - показательное, с интенсивностью потока освобождений занятого канала μ.

В начале работы в системе массового обслуживания (как и в любой динамической системе) возникнет так называемый «переходный», нестационарный процесс. Однако, спустя некоторое время, этот переходный процесс затухнет, и система перейдет на стационарный режим, вероятностные характеристики которого уже не будут зависеть от времени.

Доказано, что для любой СМО с отказами существует стационарный режим обслуживания, т.е. что при $t \to \infty$ вероятности $p_k(t) \to p_k$, а $p_k'(t) \to 0$, $k = 0,1,..n$.

Определение

Вероятности p_0, p_1, ..., p_n состояний СМО в стационарном режиме функционирования называют предельными вероятностями.

Для нахождения предельных вероятностей заменим в уравнениях Эрланга все вероятности $p_k(t)$ их пределами p_k, а все производные положим равными нулю. Получим систему уже не дифференциальных, а алгебраических уравнений

$$\begin{cases} -\lambda \cdot p_0 + \mu \cdot p_1 = 0\,, \\ \dots\dots \\ \lambda \cdot p_{k-1} - (\lambda + k\cdot\mu)\cdot p_k + (k+1)\cdot\mu\cdot p_{k+1} = 0, \\ \qquad (0 < k < n), \\ \dots\dots \\ \lambda \cdot p_{n-1} - n\cdot\mu\cdot p_n = 0. \end{cases}$$

$$\sum_{k=0}^{n} p_k = 1.$$

Решая данную систему относительно неизвестных p_0, p_1, ..., p_n, получаем выражения для предельных вероятностей состояний (формулы Эрланга):

$$p_0 = \left[\sum_{k=0}^{n} \frac{\rho^k}{k!}\right]^{-1},$$

$$p_k = \frac{\rho^k}{k!}\cdot p_0 \qquad (k = 1,2,...,n)$$

где $\rho = \dfrac{\lambda}{\mu}$ – коэффициент загрузки канала, представляет собой среднее число заявок, приходящих в СМО за среднее время обслуживания.

Формулами Эрланга дают предельный закон распределения числа занятых каналов в зависимости от характеристик потока заявок и производительности системы обслуживания. Формулы Эрланга остаются справедливыми при любом законе распределения времени обслуживания, лишь бы входной поток был простейшим.

Предельные характеристики эффективности многоканальной СМО с отказами сведены в таблицу.

Предельные характеристики	Формулы
Вероятность отказа, т.е. вероятность того, что поступившая заявка не будет обслужена	$P_{отк.} = p_n$
Вероятность обслуживания поступившей заявки:	$P_{обсл.} = 1 - p_n$
Абсолютная пропускная способность СМО (среднее число заявок, обслуживаемое СМО в единицу времени)	$Q = \lambda\cdot(1 - p_n)$
Относительная пропускная способность СМО (средняя доля заявок, обслуживаемых системой)	$q = \dfrac{Q}{\lambda} = 1 - P_{отк.} = P_{обсл.}$
Среднее число занятых каналов	$\bar{k} = \rho\cdot(1 - p_n)$

Тема 7

СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С ОЖИДАНИЕМ

1. Понятие СМО с ожиданием

Система массового обслуживания называется системой с ожиданием, если заявка, заставшая все каналы занятыми, становится в очередь и ждет, пока не освободится какой-нибудь канал.

Если время ожидания заявки в очереди ничем не ограничено, то система называется «чистой системой с ожиданием». Если оно ограничено какими-то условиями, то система называется «системой смешанного типа».

В системах с ожиданием существенную роль играет так называемая «дисциплина очереди». Ожидающие заявки могут вызываться на обслуживание как в порядке очереди (раньше прибывший раньше и обслуживается), так и в случайном, неорганизованном порядке. Существуют системы массового обслуживания «с преимуществами», где некоторые заявки обслуживаются предпочтительно перед другими («генералы и полковники вне очереди»).

2. СМО с ограниченным временем ожидания

Рассмотрим СМО смешанного типа с ограниченным временем ожидания при следующих допущениях.

1. Система имеет n каналов обслуживания.
2. На вход СМО поступает простейший поток заявок с интенсивностью λ.
3. Время обслуживания одной заявки представляет собой случайную величину, имеющую показательное распределение. Обслуживание заявок каждым каналом СМО осуществляется с интенсивностью $\mu = \dfrac{1}{\bar{t}_{обсл.}}$, где $\bar{t}_{обсл.}$ - среднее время обслуживания.
4. Заявка, заставшая все каналы занятыми, становится в очередь и ожидает обслуживания. Число мест в очереди не ограничено, но время ожидания ограничено некоторым сроком. Если до истечения этого срока заявка не будет принята к обслуживанию, то она покидает очередь и остается необслуженной.
5. Срок ожидания представляет собой случайную величину, имеющую показательное распределение. На каждую заявку, стоящую в очереди, действует «поток уходов из очереди» с интенсивностью $\nu = \dfrac{1}{\bar{t}_{ож}}$, где $\bar{t}_{ож}$ - среднее время ожидания.
6. Все потоки событий, приводящие к изменениям состояний СМО, носят пуассоновский характер. Тогда процесс, протекающий в системе, будет марковским.

При $\nu \to \infty$ система смешанного типа превращается в чистую систему с отказами; при $\nu \to 0$ она превращается в чистую систему с ожиданием.

Напишем уравнения для вероятностей состояний системы. Для этого, прежде всего, перечислим эти состояния. Будем их нумеровать не по числу занятых каналов, а по числу связанных с системой заявок. Заявку будем называть «связанной с системой», если она либо находится в состоянии обслуживания, либо ожидает очереди.

Возможные состояния системы будут:

S_0 - ни один канал не занят (очереди нет),

S_1 - занят ровно один канал (очереди нет),

....

S_k - занято ровно k каналов (очереди нет),

....

S_n - заняты все n каналов (очереди нет),

S_{n+1} - заняты все n каналов, одна заявка стоит в очереди,

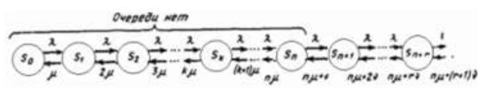
....

S_{n+l} - заняты все n каналов, l заявок стоят в очереди,

....

Число l заявок, стоящих в очереди, может быть сколь угодно большим. Таким образом, система S имеет бесконечное (хотя и счетное) множество состояний.

Схема возможных переходов дана на рис. 1



Напишем дифференциальные уравнения для вероятностей состояний системы.

$$\begin{cases} p_0'(t) = -\lambda\cdot p_0(t) + \mu\cdot p_1(t), \\ p'_1(t) = \lambda\cdot p_0(t) - (\lambda+\mu)\cdot p_1(t) + 2\cdot\mu\cdot p_2(t) \\ \dots\dots\dots \\ p'_k(t) = \lambda\cdot p_{k-1}(t) - (\lambda + k\cdot\mu)\cdot p_k(t) + (k+1)\cdot\mu\cdot p_{k+1}(t) \\ \qquad (1 \le k \le n-1), \\ \dots\dots \\ p_n'(t) = \lambda\cdot p_{n-1}(t) - (\lambda + n\cdot\mu)\cdot p_n(t) + (n\cdot\mu + \nu)\cdot p_{n+1}(t) \\ \dots\dots\dots \\ p_{n+l}'(t) = \lambda\cdot p_{n+l-1}(t) - (\lambda + n\cdot\mu + l\cdot\nu)\cdot p_{n+l}(t) + \\ \qquad + [n\cdot\mu + (l+1)\cdot\nu]\cdot p_{n+l+1}(t) \end{cases} \qquad (*)$$

Уравнения (*) являются естественным обобщением уравнений Эрланга на случай системы смешанного типа с ограниченным временем ожидания. Параметры λ, μ, ν в этих уравнениях могут быть как постоянными, так и переменными. При интегрировании системы (*) нужно учитывать, что хотя теоретически число возможных состояний системы бесконечно, но на практике вероятности $p_{n+s}(t)$ при возрастании s становятся пренебрежимо малыми, и соответствующие уравнения могут быть отброшены.

Выведем формулы, аналогичные формулам Эрланга, для вероятностей состояний системы при установившемся режиме обслуживания.

При $t \to \infty$ для установившегося режима, когда $p_k(t) \to p_k$, $p_k'(t) \to 0$, получим следующую систему алгебраических уравнений для определения вероятностей состояний p_k:

$$\begin{cases} -\lambda\cdot p_0 + \mu\cdot p_1 = 0, \\ \lambda\cdot p_0 - (\lambda+\mu)\cdot p_1 + 2\cdot\mu\cdot p_2 = 0 \\ \dots\dots\dots \\ \lambda\cdot p_{k-1} - (\lambda + k\cdot\mu)\cdot p_k + (k+1)\cdot\mu\cdot p_{k+1} = 0 \\ \qquad (1 \le k \le n-1), \\ \lambda\cdot p_{n-1} - (\lambda + n\cdot\mu)\cdot p_n + (n\cdot\mu + \nu)\cdot p_{n+1} = 0 \\ \dots\dots\dots \\ \lambda\cdot p_{n+l-1} - (\lambda + n\cdot\mu + l\cdot\nu)\cdot p_{n+l} + \\ \qquad + [n\cdot\mu + (l+1)\cdot\nu]\cdot p_{n+l+1} = 0 \end{cases}$$

К этим уравнениям нужно присоединить условие

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

Решая полученную систему, найдем

$$p_0 = \left[\sum_{k=0}^{n}\frac{\alpha^k}{k!} + \frac{\alpha^n}{n!}\cdot\sum_{l=1}^{\infty}\frac{\alpha^l}{\prod_{m=1}^{l}(n + m\cdot\rho)}\right]^{-1},$$

$$p_k = \frac{\alpha^k}{k!}p_0 \qquad (1 \le k \le n),$$

$$p_{n+l} = \frac{\alpha^{n+l}\cdot p_0}{n!}\left[\prod_{m=1}^{l}(n + m\cdot\rho)\right]^{-1}, \qquad (l \ge 1),$$

где

$$\alpha = \frac{\lambda}{\mu}, \quad \rho = \frac{\nu}{\mu} \qquad (**)$$

Параметры α и ρ выражают соответственно среднее число заявок и среднее число уходов заявки, стоящей в очереди, приходящиеся на среднее время обслуживания одной заявки.

Очевидно, что пропускная способность системы с ожиданием, при тех же λ и μ, будет всегда выше, чем пропускная способность системы с отказами, т.к. в случае наличия ожидания необслуженными уходят не все заявки, заставшие n

каналов занятыми, а только некоторые. Пропускная способность увеличивается при увеличении среднего времени ожидания $\bar{t}_{ожид} = \frac{1}{\nu}$.

Математическое ожидание числа заявок, находящихся в очереди

$$M_l = \frac{\alpha}{\rho} \cdot p_0$$

3. Чистая СМО

Посмотрим, во что превратятся формулы (**) при $\rho \to 0$ и $\rho \to \infty$.

При $\rho \to \infty$ система с ожиданием должна превратиться в систему с отказами (заявка мгновенно уходит из очереди).

При $\rho \to 0$ получим чистую систему с ожиданием. В такой системе заявки вообще не уходят из очереди, и поэтому $P_н = 0$: каждая заявка рано или поздно дождется обслуживания.

В чистой системе с ожиданием не всегда имеется предельный стационарный режим при $t \to \infty$. Можно доказать, что такой режим существует только при $\alpha < n$, т.е. когда среднее число заявок, приходящееся на время обслуживания одной заявки, не выходит за пределы возможностей n-канальной системы. Если же $\alpha \ge n$, число заявок, стоящих в очереди, будет с течением времени неограниченно возрастать.

Пусть $\alpha < n$. Тогда все заявки будут обслужены и очередь не будет возрастать до бесконечности. В этом случае

$$p_0 = \left[\sum_{k=0}^{n} \frac{\alpha^k}{k!} + \frac{\alpha^{n+1}}{n!(n-\alpha)}\right]^{-1},$$

$$p_k = \frac{\alpha^k}{k!} \cdot p_0 \quad (1 \le k \le n),$$

$$p_{n+l} = \frac{\alpha^{n+l}}{n! n^l} \cdot p_0 \qquad (l \ge 1).$$

Среднее число заявок, находящихся в очереди, при $\rho \to 0$ определяется формулой

$$m_l = \frac{\alpha^{n+1}}{n \cdot n!}\left(1 - \frac{\alpha}{n}\right)^{-2} \cdot p_0$$

4. СМО с ограниченной длиной очереди

Рассмотрим СМО с ограниченной длиной очереди при следующих допущениях.

1. Система имеет n каналов обслуживания.
2. На вход СМО поступает простейший поток заявок с интенсивностью λ.
3. Время обслуживания одной заявки представляет собой случайную величину, имеющую показательное распределение. Обслуживание заявок каждым каналом СМО осуществляется с интенсивностью $\mu = \frac{1}{\bar{t}_{обсл.}}$, где $\bar{t}_{обсл.}$ - среднее время обслуживания.
4. Заявка, заставшая все каналы занятыми, становится в очередь и ожидает обслуживания при условии, что число заявок в очереди меньше m. Если же число заявок в очереди равно m (больше m оно быть не может), то последняя прибывшая заявка в очередь не становится и покидает систему необслуженной.
5. Все потоки событий, приводящие к изменениям состояний СМО, носят пуассоновский характер. Тогда процесс, протекающий в системе, будет марковским.

Перечислим состояния системы:

S_0 - все каналы свободны, очереди нет,

S_1 - занят один канал, очереди нет,

....

S_k - занято k каналов, очереди нет,

....

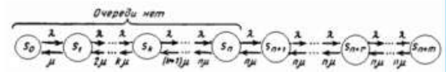
S_{n-1} - занято $n-1$ каналов, очереди нет,

S_n - заняты все n каналов, очереди нет,

S_{n+1} - заняты все n каналов, одна заявка стоит в очереди,

....

S_{n+m} - заняты все n каналов, m заявок стоит в очереди.



Составим дифференциальные уравнения для вероятностей состояний системы. В данном случае число состояний системы будет конечно, так как общее число заявок, связанных с системой, не может превышать $n+m$ (n обслуживаемых и m стоящих в очереди). Таким образом, получена система $(n+m+1)$ дифференциальных уравнений:

$$\begin{cases} p_0'(t) = -\lambda \cdot p_0(t) + \mu \cdot p_1(t), \\ \dots\dots\dots \\ p'_k(t) = \lambda \cdot p_{k-1}(t) - (\lambda + k \cdot \mu) \cdot p_k(t) + (k+1) \cdot \mu \cdot p_{k+1}(t) \\ \qquad (1 \le k \le n-1), \\ \dots\dots\dots \\ p_n'(t) = \lambda \cdot p_{n-1}(t) - (\lambda + n \cdot \mu) \cdot p_n(t) + n \cdot \mu \cdot p_{n+1}(t) \\ \dots\dots\dots \\ p_{n+s}'(t) = \lambda \cdot p_{n+s-1}(t) - (\lambda + n \cdot \mu) \cdot p_{n+s}(t) + n \cdot \mu \cdot p_{n+s+1}(t) \\ \qquad (1 \le s \le m-1), \\ p_{n+m}'(t) = \lambda \cdot p_{n+m-1}(t) - n \cdot \mu \cdot p_{n+m}(t) \end{cases}$$

Начальными условиями являются:

$$p_0(0) = 1, \; p_k(0) = 0 \quad (k = 1,2,\dots,n+m).$$

При $t \to \infty$ для установившегося режима, когда $p_k(t) \to p_k$, $p_k'(t) \to 0$ уравнения для определения p_k будут иметь вид:

$$\begin{cases} -\lambda \cdot p_0 + \mu \cdot p_1 = 0, \\ \dots\dots\dots \\ \lambda \cdot p_{k-1} - (\lambda + k \cdot \mu) \cdot p_k + (k+1) \cdot \mu \cdot p_{k+1} = 0 \\ \qquad (1 \le k \le n-1), \\ \dots\dots\dots \\ \lambda \cdot p_{n-1} - (\lambda + n \cdot \mu) \cdot p_n + n \cdot \mu \cdot p_{n+1} = 0 \\ \dots\dots\dots \\ \lambda \cdot p_{n+s-1} - (\lambda + n \cdot \mu) \cdot p_{n+s} + n \cdot \mu \cdot p_{n+s+1} = 0 \\ \qquad (1 \le s \le m-1), \\ \lambda \cdot p_{n+m-1} - n \cdot \mu \cdot p_{n+m} = 0 \end{cases} \qquad (*)$$

Добавочное условие к системе (*):

$$\sum_{k=0}^{n+m} p_k = 1.$$

Решая систему (*), получим:

$$p_0 = \left\{\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^n}{n!} \cdot \sum_{l=1}^{m}\left(\frac{\rho}{n}\right)^l\right\}^{-1},$$

$$p_k = \frac{\rho^k}{k!} p_0, \qquad (1 \le k \le n),$$

$$p_{n+l} = \frac{\rho^n}{n!} \cdot \left(\frac{\rho}{n}\right)^l \cdot p_0, \qquad (1 \le l \le m),$$

где $\rho = \frac{\lambda}{\mu}$.

Вероятность того, что заявка покинет систему необслуженной:

$$P_{отк} = p_{n+m}$$

равна вероятности P_{n+m} того, что в очереди уже стоят m заявок.

Относительная пропускная способность системы определяется формулой

$$q = 1 - P_{отк}.$$

Абсолютная пропускная способность:

$$Q = \lambda \cdot q$$

Предельные характеристики	Формулы
Вероятность отказа	$P_{отк.} = p_{n+m}$
Вероятность того, что любая заявка будет обслужена	$P_{обсл.} = 1 - P_{отк.}$
Абсолютная пропускная способность	$Q = \lambda(1 - p_{n+m})$
Относительная пропускная способность	$q = 1 - p_{n+m}$
Среднее число заявок в очереди	$\overline{N}_{оч.} = \frac{\rho^{n+1}}{n \cdot n!}\left(\frac{1-(m+1)\cdot\left(\frac{\rho}{n}\right)^m + m\cdot\left(\frac{\rho}{n}\right)^{m+1}}{\left(1-\frac{\rho}{n}\right)^2}\right) \cdot p_0$
Среднее число занятых каналов	$\overline{N}_{кан.} = \rho(1 - p_{n+m})$
Среднее число заявок в СМО (обслуживаемых или ожидающих в очереди)	$\overline{N}_{сист.} = \overline{N}_{оч.} + \overline{N}_{кан.}$
Среднее время пребывания заявки в СМО (в очереди или под обслуживанием)	$\bar{t}_{сист.} = \frac{\overline{N}_{сист.}}{\lambda}$
Среднее время пребывания заявки в очереди	$\bar{t}_{оч.} = \frac{\overline{N}_{оч.}}{\lambda}$

Тема 8

ЗАМКНУТАЯ СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ

1. Понятие замкнутой СМО

Замкнутая СМО – это система массового обслуживания, в которой есть фиксированное число источников заявок.

СМО содержит n обслуживающих каналов. На вход поступает простейший поток заявок с интенсивностью λ. Если в момент поступления требования имеется свободный канал, то он немедленно приступает к обслуживанию поступившего требования. Каждый канал может одновременно обслуживать только одно требование. Все каналы функционируют независимо. Если заявка застала все обслуживающие каналы занятыми, то она встает в очередь и ожидает начала обслуживания. Требования на обслуживание поступают от m обслуживаемых объектов, то есть поток поступающих требований ограничен. При этом предполагается, что $m > n$. Таким образом, максимальная длина очереди равна

$$m - n$$

Интенсивность обслуживания заявок

$$\mu = \frac{1}{t_{обс}},$$

где $t_{обс}$ - среднее время обслуживания объекта (источника заявок).

Интенсивность потока требований каждого источника заявок равна

$$\lambda = \frac{1}{t_{раб}}$$

где $t_{раб}$ - среднее время безотказной работы каждого объекта.

Если под обслуживанием находятся k объектов, то интенсивность потока заявок в СМО будет равна

$$(m-k)\lambda.$$

Состояния системы:

S0 – заявок нет, все каналы обслуживания свободны;

S1 – один канал занят, очереди нет;

S2 – два канала заняты, очереди нет;

… ;

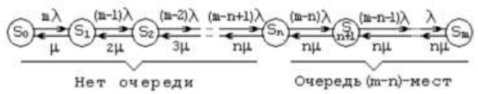
Sn – все n каналов заняты, очереди нет;

Sn+1 – все n каналов заняты, одна заявка стоит в очереди;

… ;

Sn+(m-n) = Sm – все n каналов заняты, m-n заявок стоят в очереди.

Граф замкнутой СМО представлен на рис.1.



Интенсивность потока заявок уменьшается от $m\lambda$ до λ по мере того, как ТУ одно за другим выходят из строя. Интенсивность потока обслуживания нарастает по мере подключения новых каналов обслуживания от μ до $n\mu$. Как только загружены все каналы, интенсивность потока обслуживания остается постоянной и равной $n\mu$ для состояний от Sn до Sm.

Система дифференциальных уравнений имеет вид:

$$\begin{cases} p_0'(t) = -(n+m)\lambda p_0(t) + \mu p_1(t) \\ p_1'(t) = (n+m)\lambda p_0(t) - [(n+m-1)\lambda + \mu]p_1(t) + 2\mu p_2(t) \\ p_2'(t) = (n+m-1)\lambda p_1(t) - [(n+m-2)\lambda + 2\mu]p_2(t) + 3\mu p_3(t) \\ \dots \\ p_k'(t) = (n+m-k+1)\lambda p_{k-1}(t) - [(n+m-k)\lambda + k\mu]p_k(t) + (k+1)\mu p_{k+1}(t) \\ \dots \\ p_n'(t) = (m+1)\lambda p_{n-1}(t) - (m\lambda + n\mu)p_n(t) + n\mu p_{n+1}(t) \\ p_{n+1}'(t) = m\lambda p_n(t) - [(m-1)\lambda + n\mu]p_{n+1}(t) + n\mu p_{n+2}(t) \\ \dots \\ p_{n+r}'(t) = (m-r+1)\lambda p_{n+r-1}(t) - [(m-r)\lambda + n\mu]p_{n+r}(t) + n\mu p_{n+r+1}(t) \\ \dots \\ p_{n+m-1}'(t) = 2\lambda p_{n+m-2}(t) - (\lambda + n\mu)p_{n+m-1}(t) + n\mu p_{n+m}(t) \\ p_{n+m}'(t) = \lambda p_{n+m-1}(t) - n\mu p_{n+m}(t) \\ \sum_{i=0}^{n+m} p_i(t) = 1 \\ p_0(0) = 1, p_1(0) = 0, p_2(0) = 0, \dots, p_{n+m}(0) = 0 \end{cases}$$

Предельные вероятности:

$$p_k = \begin{cases} \frac{m!\rho^k}{k!(m-k)!}p_0, & 1 \le k \le n \\ \frac{m!\rho^k}{n!n^{k-n}(m-k)!}p_0, & n < k \le m \end{cases}$$

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

где $\rho = \frac{\lambda}{\mu}$.

2. Характеристики эффективности замкнутой СМО

№	Предельные характеристики	Формулы
1.	Среднее число заявок в очереди	$\overline{N}_{оч} = \sum_{k=n}^{m} (k-n) \cdot p_k$
2.	Среднее число занятых каналов	$\overline{N}_{зан} = n - \sum_{k=0}^{n-1} (n-k) \cdot p_k$
3.	Среднее число заявок в СМО (обслуживаемых или ожидающих в очереди)	$\overline{N}_{смо} = \sum_{k=1}^{N} k \cdot p_k$
4.	Абсолютная пропускная способность	$Q = \overline{N}_{зан.кан} \cdot \mu$
5.	Относительная пропускная способность	$q = 1$
6.	Среднее время пребывания заявки в очереди	$\bar{t}_{оч} = \frac{\overline{N}_{смо}}{\lambda(m - \overline{N}_{смо})} - \frac{1}{\mu}$
7.	Коэффициент простоя обслуживаемой заявки в очереди (характеризует потери времени из-за ожидания начала обслуживания)	$K_{пр.зав.} = \frac{\overline{N}_{оч}}{m}$
8.	Коэффициент простоя обслуживающих каналов (показывает, какую долю времени смены каждый из обслуживающих каналов простаивает)	$K_{пр.кан} = \frac{\overline{N}_{зан}}{n}$
9.	Коэффициент использования обслуживаемых объектов (источников заявок)	$K_{исп.об.} = 1 - \frac{\overline{N}_{смо}}{m}$