Теория массового обслуживания (ТМО) изучает процессы, в которых возникают требования на выполнение каких-либо видов услуг, и происходит обслуживание этих требований. Объектами (ТМО) могут быть производственные процессы, процессы снабжения, транспорт, торговля, военные операции.

Общей особенностью всех задач (ТМО) является случайный характер исследуемых явлений:

1) количество требований, поступающих на обслуживание, и временной интервал между моментами их поступления являются случайными величинами;

2) длительность обслуживания каждой заявки носит случайный характер.

§5.1. Системы массового обслуживания

Условно система массового обслуживания (СМО) делится на две части: та часть, в которой возникают требования, называется ***обслуживаемой системой*** (население, часть производства), а та часть, которая принимает требования и удовлетворяет их, называется ***обслуживающей системой*** (транспорт, почта, магазины).

Система массового обслуживания включает в себя 1) источник, 2) входящий поток требований, 3) очередь, 4) обслуживающее устройство, 5) выходящий обслуженных требований.

1) **Источник.** Это устройство или группа устройств, люди или группа людей, от которых поступают требования в обслуживающую систему.
2) **Входящий поток требований.** Это последовательность появления во времени требований, поступающих от источника. Правильная работа системы зависит от того, как хорошо изучены характеристики входящего потока.
3) **Очередь.** Это совокупность требований, которые не могут быть сразу удовлетворены. Очередь присуща не всякой системе. Существуют системы, в которых очередь не допускается.
4) **Обслуживающее устройство.** Это обслуживающий аппарат, канал обслуживания. От организации обслуживающего устройства, его характеристик зависит время обслуживания требований, длина очереди, время ожидания в очереди.
5) **Выходящий поток обслуженных требований.** Это поток требований, выходящий из обслуживающего устройства.

(СМО) могут быть классифицированы по различным признакам. В соответствии с этими признаками рассматривают следующие (СМО):

1) (СМО) с отказами. Это системы, у которых требования, поступающие в момент, когда все приборы обслуживания заняты, получают отказ и теряются (телефонная сеть);
2) (СМО) с ожиданием. Это системы, у которых возможно появление как угодно длинной очереди требований (пропускная система в метро);
3) (СМО) с ограниченной длиной очереди. Это системы, допускающие очередь с ограниченным числом мест ней;
4) (СМО) с ограниченным временем ожидания. Это системы, допускающие очередь с ограниченным сроком пребывания каждого требования в ней.

Одним из признаков, по которому классифицируются (СМО), является количество обслуживающих устройств или каналов: если система имеет один прибор или один канал, то она называется одноканальной, если же их более одного, то она называется многоканальной. (СМО) классифицируются так же по местонахождению источника требований: если источник требований находится вне (СМО), то система называется ***разомкнутой***, если же она находится внутри системы, то система называется ***замкнутой***.

§5.2. Марковские случайные процессы

Часто при исследовании различных явлений природы, экономических и технических процессов приходится иметь со случайными величинами, изменяющимися во времени. В основе (ТМО) лежит математический аппарат теории вероятностей и теории случайных функций.

Случайной функцией (СФ) называется такая функция, значение которой при каждом данном значении аргумента является случайной величиной. Полной характеристикой (СФ) является ее закон распределения. Общий вид закона распределения (СФ) характеризуется многомерной функцией распределения:

$$F(t_1,t_2,..t_n,x_1,x_2,..x_n) = P(X(t_1) < x_1,X(t_2) < x_2,..X(t_n) < x_n),$$

где $x_1,x_2,..,x_n$ – текущие значения величины x в моменты времени $t_1,t_2,..,t_n$, а n может быть как угодно велико.

Общий анализ случайных функций требует установления вероятностных зависимостей между значениями этой функции в различные моменты времени. Для этого используются ***автоковариационная*** и ***взаимная ковариационная функция***.

Автоковариационной функцией (СФ) $X(t)$ называется ковариация значений этой функции при различных значениях ее аргумента

$K_x(t_1,t_2) = cov(x_1,x_2) = M[(X(t_1) - M_x(t_1))(X(t_2) - M_x(t_2))]$.

Пример. Найдем: а) математическое ожидание; б) автоковариационную функцию; в) дисперсию случайной функции $X(t) = Ucos(2t)$, где U – случайная величина, причем $M(U) = 5, D(U) = 6$.

а) Найдем искомое математическое ожидание, вынося неслучайный множитель $cos(2t)$ за знак математического ожидания,

$M\big(X(t)\big) = M(U)\,cos(2t) = 5cos(2t)$.

б) Найдем искомую автоковариационную функцию

$K_x(t_1,t_2) = M\big[\big((U-5)cos(2t_1)\big)\big((U-5)cos(2t_2)\big)\big] =$

$= cos(2t_1)cos(2t_2)M[(U-5)^2] = 6cos(2t_1)cos(2t_2)$.

в) Найдем искомую дисперсию, для чего положим $t_1 = t_2 = t$:

$D_x(t) = K_x(t,t) = 6cos^2(2t)$.

Автокорреляционной функцией (СФ) $X(t)$ называется функция вида

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

где $\sigma_x(t_1), \sigma_x(t_2)$ – средние квадратичные отклонения значений (СФ) $X(t)$ в сечениях t_1 и t_2.

Взаимной ковариационной функцией между двумя (СФ) $X(t)$ и $Y(t)$ называется ковариация значений этих функций в различные моменты времени:

$$K_{xy}(t_1,t_2) = cov(x_1,x_2) = M\big[(X(t_1) - M_x(t_1))\big(Y(t_2) - M_y(t_2)\big)\big].$$

Пример. Найдем взаимную ковариационную функцию 2 случайных функций: $X(t) = t^2U$ и $Y(t) = t^3U$, где U – случайная величина, причем $D(U) = 5$.

Найдем математические ожидания:

$M_x(t) = M(t^2U) = t^2M(U), M_y(t) = M(t^3U) = t^3M(U)$.

Найдем взаимную ковариационную функцию:

$K_{xy}(t_1,t_2) = cov(x_1,x_2) = M\left[\left(t_1^2(U - M(U))\right)\left(t_2^3(U - M(U))\right)\right] =$

$= t_1^2t_2^2M\left[\left(U - M(U)\right)^2\right] = t_1^2t_2^2D(U) = 5t_1^2t_2^2$.

Взаимной корреляционной функцией двух (СФ) $X(t)$ и $Y(t)$ называется функция двух аргументов

$\rho_{xy}(t_1,t_2) = \dfrac{K_{xy}(t_1,t_2)}{\sigma_x(t_1)\sigma_y(t_2)}$.

Существенную роль в исследовании (СФ) играют стационарные процессы.

(СФ) $X(t)$ называется ***стационарной в узком смысле,*** если вероятностные характеристики (СФ) $X(t+\Delta)$ при любом Δ тождественно совпадают с соответствующими характеристиками (СФ) $X(t)$.

(СФ) $X(t)$ называется ***стационарной в широком смысле,*** если ее математическое ожидание, дисперсия, автоковариационная и автокорреляционная функция не меняются со временем.

Особенностью автоковариационной функции является то, что она зависит только от разности значений аргумента $\tau = t_2 - t_1$
$K_x(t_1 - t_2) = K_x(t_2 - t_1) = K_x(\tau)$.

Пример. Задана случайная функция $X(t) = cos(t+\varphi)$, где φ – случайная величина, распределенная равномерно в интервале $(0,2\pi)$. Доказать, что $X(t)$ – стационарная функция в широком смысле.

Найдем математическое ожидание $X(t)$:
$$M_x(t) = M[cos(t+\varphi)] = M[cos(t)cos(\varphi) - sin(t)sin(\varphi)] =$$
$$= cos(t)M[cos(\varphi)] - sin(t)M[sin(\varphi)],$$
где
$$M[cos(\varphi)] = \frac{1}{2\pi}\int_0^{2\pi} cos(\varphi)d\varphi = 0, M[sin(\varphi)] = \frac{1}{2\pi}\int_0^{2\pi} sin(\varphi)d\varphi = 0,$$
так что окончательно $M_x(t) = 0$. Найдем автоковариационную функцию
$$K_x = M[cos(t_1+\varphi)cos(t_2+\varphi)] =$$
$$= \frac{1}{2}cos(t_2 - t_1) + \frac{1}{2}M[cos(t_1+t_2+\varphi)] = \frac{1}{2}cos(t_2 - t_1).$$

Итак, $M_x(t) = 0$ при всех значениях t и автоковариационная функция зависит только от разности аргументов. Следовательно, $X(t)$– стационарная функция.

Пусть имеется некоторая физическая система S, состояние которой меняется с течением времени случайным образом. Это значит, что в системе протекает случайный процесс.

Случайный процесс называется ***марковским процессом,*** если он обладает следующим свойством: для каждого момента времени t_0 вероятность любого состояния системы в будущем $(t > t_0)$ зависит только от ее состояния в настоящем $S(t_0)$ и не зависит от того, когда и каким образом система перешла в это состояние.

Состояния системы могут изменяться либо дискретно, либо непрерывно. Случайный марковский процесс называется процессом с ***дискретными состояниями***, если возможные состояния можно представить в виде числовой последовательности $S_1, S_2, ..S_n, ...$, а сам процесс состоит в том, что время от времени система S мгновенно перескакивает из одного состояния в другое.

Случайный марковский процесс с непрерывным изменением состояний называется марковским процессом с ***непрерывными состояниями.***

В системе с дискретными состояниями переход из состояния в состояние может происходить либо в определенные моменты времени, либо в случайные моменты.

Случайный процесс называется процессом с ***дискретным временем,*** если переходы системы из состояния в состояние возможны только в определенные моменты времени $t_1, t_2, .., t_n ...$. В промежутки времени между этими моментами система S сохраняет свое состояние.

Случайный процесс называется процессом с ***непрерывным временем,*** если переходы системы из состояния в состояние возможны в любой наперед неизвестный случайный момент времени.

Рассмотрим марковский случайный процесс с дискретным состоянием и дискретным временем.

Пусть имеется система S, которая может находиться в состояниях $S_1, S_2, ..S_n,$ причем переходы из состояния в состояние возможны только в определенные моменты $t_1, t_2, .., t_n, ...$. Случайный процесс происходящий в системе, состоит в том, что в последовательные моменты времени $t_1, t_2, .., t_n$ система оказывается в разных состояниях, например, следующим образом:

$S_1 \to S_1 \to S_3 \to S_5 \to S_5 \to S_2 \to S_1 \to \cdots,$

где стрелками указано направление перехода из состояния в состояние. Процесс, протекающий в такой схеме, можно рассматривать как последовательность состояний

$$S_1^{(1)}, S_1^{(2)}, S_3^{(3)}, S_5^{(4)}, S_5^{(5)}, S_2^{(6)}, S_1^{(7)},$$

где число в скобках обозначает номер шага, нижний индекс обозначает номер состояния. Такую последовательность называют марковской цепью. При этом в общем случае существует разная вероятность перехода из одного состояния в другое.

Марковская цепь с фиксированным шагом называется ***дискретной марковской цепью,*** если для каждого шага вероятность перехода из любого состояния S_i в любое другое состояние S_j не зависит от того, когда и как система перешла в состояние S_i.

В тех случаях, когда переход системы из одного дискретного состояния в другое происходит в случайные моменты времени, применяется схема марковского процесса с дискретными состояниями и непрерывным временем. Такая схема и такой процесс называется ***непрерывной марковской цепью.***

При анализе случайных процессов с дискретными состояниями оказывается удобным использование графов состояний. Граф состояний геометрически изображает возможные состояния системы (изображаются квадратом) и ее возможные переходы из состояния в состояние (изображаются стрелками). Стрелками изображаются только непосредственные переходы из состояния в состояние. В некоторых случаях над стрелками ставят вероятности переходов из состояния в состояние. В (ТМО) над стрелками ставят среднее число переходов в единицу времени. В (ТМО) часто используется граф, изображающий ***процесс гибели и размножения.*** Марковская непрерывная цепь называется ***процессом гибели и размножения,*** если ее граф состояний имеет вид, когда каждое из промежуточных состояний связано прямой и обратной связью с каждым соседним состоянием. В такой схеме существует возможность перехода из предыдущего состояния в последующее и обратно, но не возможен перескок через состояние.

Вероятности состояний и переходные вероятности

Марковские случайные процессы обычно описывают с помощью ***вероятностей состояний***. Для марковских процессов с дискретным временем ***вероятностями состояний*** системы на k-ом шаге (в k-й момент времени) называются вероятности

$$P_1(k) = P\left(S_1^{(k)}\right), P_2(k) = P\left(S_2^{(k)}\right), \dots P_n(k) = P\left(S_n^{(k)}\right),$$

здесь $\sum_{i=1}^{n} P_i(k) = 1$.

Для марковских процессов с непрерывным временем ***вероятностями состояний*** системы называются вероятности

$$P_1(t), P_2(t), \dots P_n(t),$$

т.е. аргументом вероятности состояния является не номер шага, а текущее время.

В случае дискретной марковской цепи для любого шага существуют еще вероятности перехода системы из одного состояния в любое другое. Эти вероятности называют ***переходными вероятностями*** марковской цепи. Если переходные вероятности не зависят от номера шага, то марковская цепь называется ***однородной.*** Если же они зависят от номера шага, то цепь называется ***неоднородной.***

Переходные вероятности представляются квадратными матрицами

$$\{P_{ij}\} = \begin{pmatrix} P_{11} P_{12} \dots P_{1n} \\ \dots \dots \dots \dots \\ P_{n1} P_{n2} \dots P_{nn} \end{pmatrix},$$

где сумма членов, стоящих в каждой строке, равна единице. Зная матрицу переходных вероятностей, можно построить граф состояний с отмеченными на нем переходными вероятностями. Граф, на котором отмечены переходные вероятности, называется ***размеченным графом.***

Имея матрицу переходных вероятностей или размеченный граф и зная начальное состояние, можно найти вероятности состояний на любом k-ом шаге для дискретной марковской цепи.

Исследование случайных процессов зависит от вида рассматриваемых состояний. Состояние S_i называется ***несущественным,*** если существует такое S_j и такое n, что $P_{ij}(n) > 0$, но $P_{ij}(m) = 0$ для всех m, где n, m – число тактов перехода. Таким образом, несущественное состояние характеризуется тем, что из него можно попасть в некоторое другое состояние, но вновь вернуться в первоначальное состояние уже нельзя. Все состояния, отличные от несущественных, называются ***существенными.*** Говорят, что система обладает ***эргодическим*** свойством, если ее состояния принадлежат к одному существенному классу состояний, т.е. это свойство заключается в том, что объект, находящийся в момент t, в состоянии i, через достаточно большой промежуток времени возвращается в это состояние. Эргодическое свойство играет большую роль при исследовании стационарных случайных процессов. Эффективным инструментом исследования таких процессов являются методики, опирающиеся на теорему Биркхофа-Хинчина.

Теорема Биркхофа-Хинчина. Если непрерывный стационарный процесс $X(t)$ имеет конечное математическое ожидание, то с единичной вероятностью существует предел

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T X(t)\,dt.$$

Эта теорема позволяет получать такие характеристики, как математическое ожидание и дисперсия на основе обработки информации единственной реализации процесса без проведения многократных испытаний других реализаций этого процесса.

Уравнения Колмогорова для вероятностей состояний

Исследование непрерывной цепи Маркова основывается на уравнениях Колмогорова. Пусть имеется непрерывная цепь Маркова, т.е. система может находиться в n дискретных состояниях $S_1, S_2, .. S_n$, переход в которые осуществляется в любой случайный момент времени. Пусть $P_i(t)$ – вероятность того, что в момент t система находится в состоянии S_i, и пусть требуется найти алгоритм, описывающий изменение всех $P_i(t)$ в любой момент времени. В любой момент времени $\sum_{i=1}^{n} P_i(t) = 1$. Введем понятие плотности вероятности перехода из состояния в состояние.

Плотностью вероятности перехода из состояния S_i в состояние S_j называется величина

$$\lambda_{ij} = \lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t},$$

где $P_{ij}(\Delta t)$ – вероятность того, что система, находящаяся в момент времени t в состоянии S_i, за время Δt перейдет в состояние S_j. С точностью до бесконечно малых высшего порядка

$$P_{ij}(\Delta t) = \lambda_{ij} \Delta t.$$

Будем рассматривать однородные непрерывные марковские цепи (λ_{ij} **не зависят от времени**), характеризующие процессы гибели и размножения. Получим дифференциальные уравнения для вероятностей $P_i(t)$. Прежде всего, найдем дифференциальное уравнение для вероятности $P_0(t)$ начального состояния S_0. Дадим t малое приращение Δt и найдем вероятность того, что в момент $t + \Delta t$ система будет находиться в состоянии S_0. Это событие может произойти двумя способами:

– в момент t система была уже в состоянии S_0 и за время Δt не перешла в состояние S_1;

– в момент t система была в состоянии S_1 и за время Δt перешла в состояние S_0.

Вероятность первого события по теореме умножения равна произведению безусловной вероятности $P_0(t)$ на условную вероятность не перехода из состояния S_0 в состояние S_1 $(1 - \lambda_{01}\Delta t)$, т.е.

$$P_0(t)(1 - \lambda_{01}\Delta t).$$

Аналогично, вероятность второго события равна

$$P_1(t)\lambda_{10}\Delta t.$$

Тогда по теореме сложения

$$P_0(t + \Delta t) = P_0(t)(1 - \lambda_{01}\Delta t) + P_1(t)\lambda_{10}\Delta t,$$

так что искомое дифференциальное уравнение имеет вид

$$\lim_{\Delta t \to 0} \frac{P_0(t+\Delta t) - P_0(t)}{\Delta t} = \frac{dP_0(t)}{dt} = -\lambda_{01}P_0(t) + \lambda_{10}P_1(t). \qquad (1)$$

Найдем дифференциальное уравнение для вероятности $P_i(t)$ промежуточного состояния S_i. Дадим t малое приращение Δt и найдем

вероятность того, что в момент $t+\Delta t$ система будет находиться в состоянии S_i. Это событие может произойти тремя способами:
– в момент t система была в состоянии S_i и за время Δt не перешла ни в состояние S_{i-1}, ни в состояние S_{i+1} ;
– в момент t система была в состоянии S_{i-1} и за время Δt перешла в состояние S_i.
– в момент t система была в состоянии S_{i+1} и за время Δt перешла в состояние S_i.

Как и в предыдущем случае, вероятность первого события определяется как
$P_i(t)\left(1-\lambda_{i,i-1}\Delta t-\lambda_{i,i+1}\Delta t\right)$,
а вероятности второго и третьего событий равны

$$P_{i-1}(t)\lambda_{i-1,i}\Delta t,\quad P_{i+1}(t)\lambda_{i+1,i}\Delta t,$$

так что система линейных дифференциальных уравнений Колмогорова, решение которой при заданных начальных условиях обеспечивает возможность получения функций $P_i(t)$, имеет вид

$$\frac{dP_i(t)}{dt}=P_{i-1}(t)\lambda_{i-1,i}+P_{i+1}(t)\lambda_{i+1,i}-\left(\lambda_{i,i-1}+\lambda_{i,i+1}\right)P_i(t), \qquad (2)$$

где $i=1,2\dots n$.

Входящий поток требований

Исследование (СМО) обычно начинается с изучения входящего потока требований и количественного его описания.

По числу требований, поступающих в систему, входящие потоки бывают ограниченными и неограниченными. По характеру поступления потоки делятся на регулярные (это такие потоки, в которых требования появляются регулярно, строго через заданные периоды) и стахостические (это такие потоки, в которых момент появления очередного требования является случайной величиной). В (ТМО) в основном рассматриваются стахостические потоки. Стахостические потоки могут быть описаны двумя способами: 1) с помощью функции распределения $F(t)=P(\tau<t)$, где τ – случайный интервал времени между двумя требованиями; 2) с помощью вероятности появления 1,2 или более требований на заданном интервале $[0,t]$.

Второй подход в (ТМО) встречается наиболее часто. При этом используется пуассоновский поток, для которого вероятность поступления за промежуток времени Δt ровно k требований описывается формулой Пуассона

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

Пуассоновский поток обладает тремя свойствами: стационарностью, ординарностью и отсутствием последействия. ***Стационарность*** означает, что вероятность поступления определенного числа требований на участок времени длиной Δt зависит только от длины этого участка и не зависит от того, где на оси времени расположен этот участок. ***Ординарность*** означает, что вероятность поступления на элементарный участок 2 и более требований пренебрежимо мала по сравнению с вероятностью поступления 1 требования. ***Отсутствие последействия*** означает, что вероятности поступления определенного числа требований на любых неперекрывающихся участках времени не зависят друг от друга. Важнейшей характеристикой пуассоновского потока является интенсивность, определяемая математическим ожиданием числа требований, поступающих в единицу времени $M(K)=\sum_{k=1}^{\infty} kP_k(1)=\lambda$.

Прежде чем описывать аналитически входящий поток требований, обычно осуществляют выборку из рассматриваемого потока, которую обрабатывают известными методами математической статистики. Затем проверяют согласие полученного экспериментального закона распределение чаще всего с пуассоновским законом распределения по критерию Пирсона.

Обслуживающее устройство

Обслуживающее устройство (СМО) предназначено для обслуживания поступающих в него требований. Если обслуживание требования происходит за один этап, то (СМО) называется ***однофазной***. Когда обслуживание требований происходит в несколько этапов, то (СМО) называется ***многофазной.*** Одной из наиболее важных характеристик обслуживающей системы является длительность времени обслуживания в каждой фазе. Считается, что время обслуживания является случайной величиной $T_{об}$. Обычно предполагается, что функция распределения времени обслуживания имеет показательных вид

$F(t)=1-e^{-\mu t}$,

где параметр распределения $\mu=\frac{1}{M(T_{об})}$, так что этот параметр означает среднее число требований, обслуженных системой в единицу времени.

§5.3. Показатели эффективности (СМО)

Характеристикой работы (СМО) служат показатели эффективности обслуживающей системы, которые можно разделить на 3 группы:

1) показатели, отражающие качество обслуживания (среднее число требований, находящихся в системе, среднее число требований, находящихся в очереди, среднее время пребывания требования в системе, среднее время ожидания требованием начала обслуживания);
2) показатели, описывающие работу обслуживающей системы (среднее число приборов или каналов, занятых обслуживанием требований, коэффициент загрузки обслуживающих устройств, коэффициент постоя обслуживающих приборов) ;
3) показатели, отражающие экономические особенности системы (стоимость (СМО), трудовые затраты обслуживающего песонала, убытки или доходы системы).

Наиболее важным показателем эффективности, относящимся к первой группе, является величина $P_{отк}$ – вероятность того, что поступающее в систему требование отказывается присоединиться к очереди и теряется. Для системы с отказами

$$P_{отк} = P_m,$$

где P_m – вероятность того, что число требований в системе равно числу приборов m. Для системы с ограниченной длиной очереди l

$$P_{отк} = P_{m+l}.$$

Вторым важным показателем эффективности, относящимся к первой группе, является среднее количество требований, ожидающих облуживания,

$$M_{ож} = \sum_{n=m+1}^{m+l} (n-m)P_n,$$

где P_n – вероятность того, что в системе находится n требований, из которых $(n-m)$ находятся в очереди.

Среднее количество требований, находящихся как в очереди, так и на обслуживании, характеризуется величиной

$$M = \sum_{k=1}^{m+l} kP_k.$$

Среднее время ожидания требованием начала обслуживания характеризуется математическим ожиданием

$$T_{ож} = \int_0^{\infty} t dF(t),$$

где $F(t) = P(\bar{T}_{ож} < t)$ – функция распределения времени ожидания.

Среднее время пребывания требования в системе определяется как

$$\bar{T} = \bar{T}_{ож} + \bar{T}_{об},$$

где $\bar{T}_{об} = \frac{1}{\mu}$ – среднее время обслуживания требования.

Важным показателем эффективности системы с отказами, относящимся ко второй группе, является среднее число занятых обслуживанием приборов

$$m_з = \sum_{k=1}^{m} kP_k,$$

где m – общее число приборов обслуживания, P_k – вероятность того, что в системе находится k требований и, соответственно, заняты обслуживанием k приборов.

Среднее число свободных от обслуживания приборов определяется как

$$m_{св} = \sum_{k=0}^{m-1} (m-k)P_k.$$

Для (СМО) с ограниченной длиной очереди

$$m_з = \mathrm{M} - M_{ож},$$

а среднее число свободных приборов

$$m_{св} = m - m_з.$$

С последними 2 характеристиками связаны такие характеристики, как ***коэффициент загрузки*** оборудования

$$k_з = \frac{m_з}{m},$$

и коэффициент простоя оборудования

$$k_п = \frac{m_{св}}{m}.$$

Одной из характеристик 2 группы является вероятность того, что все приборы системы с отказами заняты обслуживанием требований

$$P_з = P_m,$$

где P_m – вероятность того, что на обслуживании находится ровно m требований.

Для (СМО) с ограниченной длиной очереди

$$P_з = \sum_{n=m}^{m+l} P_n.$$

Показатели 3 группы (величина потерь, величина прибыли) используются после того, как проведена оценка показателей 1 и 2 групп.

Величина потерь для (СМО) с отказами, с ожиданием и ограниченной очередью находится по формулам:

1) (СМО) с отказами

$$G_п = [q_k m_з + q_y P_{отк}\lambda + q_{пk} m_{св}]T,$$

где q_k – стоимость эксплуатации одного прибора в единицу времени, q_y – стоимость убытков в результате ухода требования из системы, $q_{пk}$ – стоимость простоя одного прибора в единицу времени, T – интервал времени.

2) (СМО) с ожиданием

$$G_п = [q_k m_з + q_{ож} M_{ож} + q_{пk} m_{св}]T,$$

где $q_{ож}$ – стоимость потерь, связанных с ожиданием требования в очереди в единицу времени.

3) (СМО) с ограниченной длиной очереди

$$G_п = [q_k m_з + q_y P_{отк}\lambda + q_{пk} m_{св} + q_{ож} M_{ож}]T.$$

Величина прибыли подсчитывается по формуле

$$П = P_з \lambda C T - G_п,$$

где С – доход, получаемый от обслуживания одного прибора.

§5.4. Модели (СМО)

1. (СМО) с отказами

На практике через некоторое время после начала работы системы устанавливается стационарный режим, характеризующийся тем, что вероятности состояний не меняются во времени $\left(\frac{dP_i(t)}{dt} = 0\right)$. Тогда система дифференциальных уравнений Колмогорова для стационарного режима переходит в систему алгебраических уравнений

$$\begin{cases} -\lambda_{01}P_0 + \lambda_{10}P_1 = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ P_{i-1}\lambda_{i-1,i} + P_{i+1}\lambda_{i+1,i} - (\lambda_{i,i-1} + \lambda_{i,i+1})P_i = 0, i = 1,2,..,n, \\ \dots\dots\dots\dots\dots\dots\dots\dots \end{cases}$$

Учет условия $\sum_{i=0}^{n} P_i = 1$ позволяет исключить одно из уравнений этой системы.

Рассмотрим вначале одноканальную систему с отказами. Для такой системы существует два состояния: S_0 – отсутствие требования в системе; S_1 – обслуживание требования обслуживающим устройством.

Вероятность состояния S_0 на малом интервале Δt равна

$$P_0(\Delta t) = \frac{\lambda^0}{0!}e^{-\lambda\Delta t} = e^{-\lambda\Delta t},$$

так что вероятность состояния S_1 в силу ординарности пуассоновского потока будет равна

$$P_1(\Delta t) = 1 - P_0(\Delta t) \approx 1 - (1 - \lambda\Delta t) = \lambda\Delta t.$$

С другой стороны, переходная вероятность $P_{01} = \lambda_{01}\Delta t$. Так как очевидно, что $P_1(\Delta t) = P_{01}$, то интенсивность λ является плотностью вероятности переходы λ_{01}. Обратный переход происходит под действием потока обслуженных требований, описываемый экспоненциальным распределением, так что

$P_0(\Delta t) = 1 - e^{-\mu\Delta t} = \mu\Delta t. P_{10} = \lambda_{10}\Delta t = P_0(\Delta t)$

и, следовательно, $\lambda_{10} = \mu$. Таким образом, имеем два уравнения

$$\begin{cases} -\lambda P_0 + \mu P_1 = 0, \\ P_0 + P_1 = 1, \end{cases}$$

откуда находим

$$P_0 = \frac{\mu}{\lambda+\mu}, P_1 = \frac{\lambda}{\lambda+\mu}.$$

Так как событие обслуживания требования соответствует событию отказа для вновь поступившего требования, то

$$P_{\text{отк}} = \frac{\lambda}{\lambda + \mu}.$$

Рассмотрим теперь многоканальную однофазную (СМО) с отказами. Состояния этой системы будут следующие:

S_0 – все каналы свободны,

S_1 – занят один канал, остальные свободны,

S_2 – заняты 2 канала, остальные свободны,

………………………………………………,

S_m – заняты все m каналов.

Прямой переход системы из состояния S_i в состояние S_{i+1}, происходит под действием входящего потока требований с постоянной интенсивностью λ. Обратный же переход происходит под действием потока обслуженных требований с переменной интенсивностью $i\mu$. Тогда система линейных алгебраических уравнений примет вид

$$\begin{cases} -\lambda P_0 + \mu P_1 = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ P_{i-1}\lambda + (i+1)\mu P_{i+1} - (\lambda + i\mu)P_i = 0, i = 1,2,..,m-1 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ -\lambda P_{m-1} + m\mu P_m = 0. \end{cases}$$

Из этих уравнений можно выразить вероятности любого состояния через интенсивность потока входящих требований λ, интенсивность потока обслуженных требований μ и вероятность P_0:

$$P_i = \frac{P_0}{i!}\left(\frac{\lambda}{\mu}\right)^i.$$

Из условия $\sum_{i=0}^{n} P_i = 1$, следует, что

$$P_0 = \left\{\sum_{i=1}^{m}\left(\frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i\right)\right\}^{-1}.$$

Таким образом, для многоканальной (СМО) с отказами вероятность отказа равна

$P_{отк} = P_m = \dfrac{P_0}{m!}\left(\dfrac{\lambda}{\mu}\right)^m.$

Пример. Рассмотрим трехканальную телефонную линию: $m = 3,\quad \lambda = 0.8, \bar{T}_{об} = 1.5$ мин. Найдем $P_{отк}$ и $m_з$. Так как $\bar{T}_{об} = 1.5$ мин, то $\mu = \frac{1}{1.5} = 0.667\ \frac{1}{мин}$, и $\frac{\lambda}{\mu} = \frac{0.8}{0.667} = 1.2$. Тогда

$$P_{отк} = \frac{1.2^3}{3!}\left[1 + 1.2 + \frac{1.2^2}{2} + \frac{1.2^3}{2\cdot 3}\right] = 0.09,$$

а среднее число занятых каналов

$$m_з = P_0\sum_{k=1}^{3}\frac{k}{k!}\left(\frac{\lambda}{\mu}\right)^k = \frac{1}{3.208}\left(1.2 + 2\frac{1.2^2}{2} + 3\frac{1.2^3}{2\cdot 3}\right) = 1.09.$$

Таким образом, показатель эффективности обслуживания высок и равен

$$P_{об} = (1 - P_{отк})100\% = 91,$$

но коэффициент загрузки не высок

$k_з = \dfrac{m_з}{m} = \dfrac{1.09}{3} = 0.36.$

2. (СМО) с ограниченной длинной очереди

Рассмотрим работу однофазной (СМО) вначале с одним, а затем с m каналами, в которой допускается наличие ограниченной очереди ожидающих обслуживания требований.

Пусть в систему поступает пуассоновский поток входящих требований с интенсивностью λ, а интенсивность потока обслуженных требований одним каналом равна μ, количество мест в очереди ограничено величиной l. В одноканальной системе заявка, поступившая в момент, когда канал занят, становится в очередь и ожидает обслуживания при условии, что длина очереди в этот момент менее l. Обозначим состояния данной системы:

S_0 – отсутствие требований в системе;

S_1 – одно требование обслуживается, очереди нет;

S_2 – одно требование обслуживается, одно стоит в очереди;

…………………………………………………………………….,

S_{l+1} – одно требование обслуживается, l стоят в очереди.

Так как интенсивность перехода из состояния S_{i-1} в состояние S_i для одноканальной (СМО) с ограниченной длиной очереди равна λ, а интенсивность обратных переходов равна μ, то уравнения для вероятностей состояний примут вид

$$\begin{cases} -\lambda P_0 + \mu P_1 = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ P_{i-1}\lambda + \mu P_{i+1} - (\lambda + \mu) P_i = 0, i = 1,2,\dots, l, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ -\lambda P_l + \mu P_{l+1} = 0. \end{cases}$$

Решая систему уравнений получим

$$P_i = P_0 \left(\frac{\lambda}{\mu}\right)^i,$$

где с учетом условия $\sum_{i=0}^{n} P_i = 1$

$$P_0 = \left\{\sum_{i=0}^{l+1}\left(\left(\frac{\lambda}{\mu}\right)^i\right)\right\}^{-1} = \begin{cases} \dfrac{1-q}{1-q^{l+2}}, & q = \dfrac{\lambda}{\mu} \neq 1, \\ \dfrac{1}{l+2}, & q = 1. \end{cases}$$

Вероятность отказа в рассматриваемой системе равна вероятности состояния S_{l+1}, т.е.

$$P_{\text{отк}} = P_{l+1} = \begin{cases} q^{l+1}\dfrac{1-q}{1-q^{l+2}}, q \neq 1 \\ \dfrac{q^{l+1}}{l+2}, \quad q = 1. \end{cases}$$

Теперь рассмотрим многоканальную (СМО) с m каналами и ограниченной длиной очереди. При наличии нескольких каналов обслуживания интенсивности потока обслуженных требований меняется до состояния, когда все каналы окажутся занятыми, а для последующих состояний интенсивности обслуживания не меняются. Для рассматриваемой системы в состояниях $S_1, S_2, \dots S_m$ обслуживаются, соответственно, 1,2, …m требований при отсутствии очереди. Следующие состояния соответствуют наличию очереди. Тогда формулы для определения вероятностей состояний $S_1, S_2, \dots S_m$ будут иметь вид

$$P_i = \frac{P_0}{i!}\left(\frac{\lambda}{\mu}\right)^i, i = 1,2 \dots m.$$

Выражения для вероятностей состояний с S_{m+1} до S_{m+l} найти из системы уравнений

$$\begin{cases} P_{m-1}\lambda + m\mu P_{m+1} - (\lambda + m\mu)P_m = 0, \\ P_m\lambda + m\mu P_{i+2} - (\lambda + m\mu)P_{m+1} = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots, \\ P_{m+l-1}\lambda - \mu P_{m+l} = 0. \end{cases}$$

Решение этой системы имеет вид

$$P_i = \frac{P_0}{m!\,m^{i-m}}\left(\frac{\lambda}{\mu}\right)^i, i = m+1, m+2 \dots m+l,$$

где с учетом условия $\sum_{i=0}^{m+l} P_i = 1$

$$P_0 = \left[\sum_{i=0}^{m}\frac{q^i}{i!} + \frac{q^{m+1}}{mm!}\frac{1-\left(\frac{q}{m}\right)^l}{1-\frac{q}{m}}\right]^{-1}.$$

Тогда вероятность отказа для многоканальной (СМО) с ограниченной длиной очереди определяется выражением

$$P_{отк} = P_{m+l} = \frac{q^{m+l}}{m^l m!}P_0.$$

Пример. Исследуем эффективность автозаправочной станции, на которой установлено три колонки для отпуска бензина, а на площадке для ожидания могут находиться не более 3 автомобилей. Предположим, что на станцию прибывают в среднем 3 машины в минуту, т.е. λ=3, а среднее время заправки 1 машины 2 минуты, т.е. $\mu = \frac{1}{2}$. Найдем вероятность отказа, среднее число машин на обслуживании, среднюю длину очереди.

Согласно исходным данным $q = \frac{3}{0.5} = 6$, $\frac{q}{m} = 2$. Тогда

$$P_0 = \left[1 + \frac{6}{1!} + \frac{6^2}{2!} + \frac{6^3}{3!} + \frac{6^4}{3!\,3}\frac{1-2^3}{1-2}\right]^{-1} = 0.00177,$$

$$P_{отк} = \frac{q^{m+l}}{m^l m!}P_0 = 2^3\frac{6^3}{3!}0.00177 = 0.51.$$

Это означает, что 51% прибывающих на станцию машин получают отказ в обслуживании.

Среднее число машин, находящихся на станции, равно

$$M = \sum_{i=0}^{m+l} iP_i = \left\{\sum_{i=1}^{m}\frac{q^i}{(i-1)!} + \frac{q^m}{m!}\sum_{n=1}^{l}(m+n)\left(\frac{q}{m}\right)^n\right\}P_0.$$

Подставив в последнее выражение исходные данные, получим

$$M = \left\{6 + \frac{6^2}{1!} + \frac{6^3}{2!} + \frac{6^4}{3!}[4\cdot 2 + 5\cdot 2^2 + 6\cdot 2^3]\right\}0.00177 = 5.11.$$

Поскольку

$$M_{ож} = \sum_{i=m+1}^{m+l} (i-m)P_i = \frac{P_0 q^m}{m!} \sum_{n=1}^{l} \left(\frac{q}{m}\right)^n,$$

то

$$M_{ож} = \frac{0.00177 \cdot 6^3}{3!} [2 + 2 \cdot 2^2 + 3 \cdot 2^3] = 2.16.$$

Среднее число занятых колонок для данной системы
$m_з = M - M_{ож} = 2.95$.
а коэффициент загрузки $k_з = \frac{2.95}{3} = 0.98.$

3.(СМО) с ожиданием (с неограниченной длиной очереди)

(СМО) с ожиданием характеризуются тем, что поступившие в систему требования после того, как все каналы окажутся занятыми, остаются в ней независимо от длины очереди, так что формулы для определения вероятностей состояний совпадают с формулами для системы (СМО) с ограниченной длиной очереди

$$P_i = \frac{P_0}{i!}\left(\frac{\lambda}{\mu}\right)^i, i = 1,2 \dots m,$$

$$P_i = \frac{P_0}{m!\, m^{i-m}}\left(\frac{\lambda}{\mu}\right)^i, i = m+1, m+2 \dots m+l, \dots.$$

При этом при $q < 1$ согласно формуле геометрической прогрессии

$$P_0 = \lim_{l \to \infty}\left[\sum_{i=0}^{m} \frac{q^i}{i!} + \frac{q^m}{m!} \sum_{i=m+1}^{m+l} \left(\frac{q}{m}\right)^{i-m}\right]^{-1} = \left[\sum_{i=0}^{m} \frac{q^i}{i!} + \frac{q^{m+1}}{m!\,(m-q)}\right]^{-1}.$$

Когда же $q > 1$, то ряд расходится, и найти P_0 из приведенного выше соотношения невозможно. Такой случай возникает тогда, когда $\lambda > m\mu$ и режим работы системы является нестационарным, так что использование уравнений стационарного состояния оказывается недопустимым. Но обычно такие системы не проектируются.

4.(СМО) с ограниченным временем ожиданием

Особенностью (СМО) с ограниченным временем ожидания является то, что требование может покидать систему до начала обслуживания. Рассмотрим систему с ограниченным временем ожидания, содержащую m каналов. Пусть время ожидания ограничено случайным временем $T_{ож}$, среднее значение которого равно $\bar{T}_{ож}$. Тогда интенсивность потока требований, покидающих очередь, вызываемого наличием одного требования в очереди, определяется величиной

$$\nu = \frac{1}{\bar{T}_{ож}},$$

а если в очереди одновременно находится k требований, то интенсивность равна $k\nu$, так что с учетом потока обслуженных требований интенсивность потока требований, покидающих систему, равна сумме интенсивностей этих двух потоков

$$m\mu + k\nu.$$

Так как режим перехода системы из состояния в состояние до S_m у рассматриваемой системы аналогичен режиму перехода системы с отказами и с ожиданием, то вероятности состояний $S_1, S_2, \dots S_m$ находятся по прежним формулам

$$P_i = \frac{P_0}{i!}\left(\frac{\lambda}{\mu}\right)^i, i = 1,2 \dots m.$$

Формулы, определяющие вероятности состояний (СМО) с ограниченным временем, следуют из соответствующих формул для системы с ожиданием путем замены отношения интенсивностей $\frac{\lambda}{m\mu}$ на отношение $\frac{\lambda}{m\mu+k\nu}$, учитывающее длины очереди, так что при $i > m$

$$P_{m+k} = \frac{q^m}{m!}\frac{\lambda^k}{\prod_{j=1}^{k}(m\mu + j\nu)}.$$

Используя формулу

$$\sum_{i=0}^{m+k} P_i = 1, \quad k \to \infty,$$

и учитывая выражения для вероятностей, получим

$$P_0 = \left[\sum_{i=0}^{m}\frac{q^i}{i!} + \frac{q^m}{m!}\sum_{k=1}^{\infty}\frac{\lambda^k}{\prod_{j=1}^{k}(m\mu + j\nu)}\right]^{-1},$$

где в практических расчетах ограничиваются одним-двумя членами входящего в формулу ряда.

5. Замкнутые (СМО)

В замкнутых (СМО) интенсивность потока поступающих на обслуживание требований зависит от самой системы. Для замкнутых систем переходные плотности вероятностей равны

$$\lambda_{01} = m\lambda, \lambda_{12} = (m-1)\lambda, \ldots \lambda_{m-1,m} = \lambda,$$
$$\lambda_{10} = \lambda_{21} = \cdots = \lambda_{m,m-1} = \mu.$$

Тогда из колмогоровских уравнений вероятностей состояний в стационарном режиме, можно получить зависимости для вероятностей состояний замкнутой системы

$$P_i = m(m-1)\ldots(m-i+1)q^i P_0 \; (i = 1,2..m).$$

Откуда, используя условие

$$\sum_{i=0}^{m} P_i = 1,$$

для начальной вероятности можно получить формулу