СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

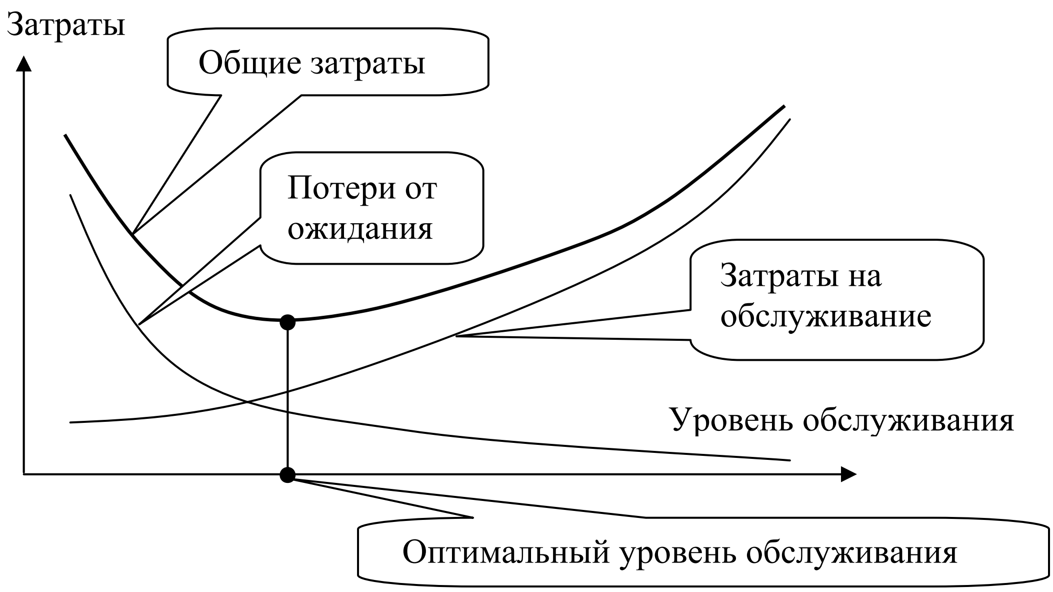
Элементы и классификация

Ожидание того или иного вида обслуживания является частью нашей повседневной жизни. Люди стоят в очередях к различным кассам, кабинетам, ждут на остановках транспорта. Но в очередях ждут не только люди – товары лежат на складе, пока на них не появится спрос; документы, подлежащие обработке, образуют очередь, также как и транспортные средства, которые стоят и перед светофорами и на заправке и при прохождении технического осмотра. Поэтому возникает необходимость в разработке систем массового обслуживания (СМО).

Цель изучения СМО состоит в том, чтобы взять под контроль некоторые характеристики обслуживающих систем, установить зависимость между числом обслуживаемых единиц и качеством обслуживания. Качество выше, чем больше число обслуживаемых единиц. Но экономически невыгодно иметь лишние обслуживающие единицы.

В промышленности СМО применяются при поступлении сырья, материалов, комплектующих изделий на склад и выдаче их со склада; обработке широкой номенклатуры деталей на одном и том же оборудовании; организации наладки и ремонта оборудования; определении оптимальной численности обслуживающих отделов и служб предприятий и т.д.

Разумеется, можно так увеличить пропускную способность любой системы обслуживания, что очереди если не исчезнут, то станут практически незаметны, но стоимость создания и эксплуатации такой системы существенно возрастет. Следовательно, необходимо уметь находить разумный компромисс между затратами на обеспечение услуг и потерями из-за задержек в их предоставлении, как показано на рис.



Такой расчет требует построения стоимостной модели, что на практике ограничивается трудностями расчета величины потерь от ожидания обслуживания, которые особенно сложно определить, если услуги предоставляются индивидууму, чьи интересы обычно не совпадают с интересами системы обслуживания. Поэтому ограничимся рассмотрением прямых задач оптимизации систем массового обслуживания, когда рассчитываются основные показатели функционирования системы. Наиболее значимые среди них –это среднее время ожидания в очереди и средняя длина очереди.

Основными элементами модели массового обслуживания являются заявка на обслуживание (клиент) и обслуживающая система (сервис). Если сервис свободен, то клиент сразу попадает на обслуживание, иначе возникает очередь. Появление клиентов (заявок на обслуживание) характеризуется интервалом между их последовательными поступлениями, а функционирование сервиса – временем обслуживания. Как правило, эти параметры являются случайными. Невозможно согласовать время покупки со всей массой покупателей или время поступления на ремонт различной бытовой техники. Даже предварительная запись (например, к врачу) не исключает образование очереди, потому что продолжительность обслуживания также случайная величина.

Поэтому в системах массового обслуживания выделяют два потока событий: входной поток заявок на обслуживание и выходной поток обслуженных заявок. Эти потоки характеризуются определенными законами распределения вероятностей, а в результате их взаимодействия система оказывается в том или ином состоянии. Расчет вероятностных характеристик состояния системы (длины очереди, времени ожидания и т. д.) – это одна из главных задач теории массового обслуживания.

Иногда, при незнании теории массового обслуживания, возникает иллюзия, что максимум эффективности будет достигаться при совпадении интенсивностей этих потоков. Но это ошибка, потому что возникающая в этом случае очередь будет стремиться к неограниченному росту. Следовательно, клиенты будут искать другую систему обслуживания, что означает их потерю, а значит, и соответствующие убытки.

Чтобы анализировать работу систем обслуживания, надо определить правила формирования и обслуживания очереди, которые называют дисциплиной очереди. Наиболее распространенный принцип ее построения основан на правиле «первым пришел – первым обслуживаешься» (часто обозначается аббревиатурой *FIFO* – от английского *First-In-First-Out*). Второе правило (например, при обработке документов) – «последним пришел – первым обслуживаешься» (обозначается аббревиатурой *LIFO* – от англ. *Last-In-First-Out*). Могут использоваться случайный отбор заявок, учет определенных приоритетов, вводиться ограничение на время пребывания заявки в очереди.

Обычно люди, принимая решение ждать обслуживание или отказаться от него, ориентируются по длине очереди. Поэтому при моделировании учитывают максимально допустимое количество заявок в очереди, которое обозначают m. Возможны случаи: $m = 0$ – без очереди, $m > 0$ – с очередью,

$m = \infty$ – очередь не ограничена. Считается, что заявки, которым не оказалось места в очереди, навсегда теряются. Первый случай ($m = 0$) - системы с отказами, в которых при занятости всех каналов обслуживания заявка не встает в очередь и покидает систему необслуженной. Третий случай ($m = \infty$) - системы с неограниченным ожиданием, в которых заявка встает в очередь, если в момент ее поступления все каналы были заняты. Второй случай ($m > 0$) - системы смешанного типа с ожиданием и ограниченной длиной очереди: заявка получает отказ, если приходит в момент, когда все места в очереди заняты. Заявка, попавшая в очередь, обслуживается обязательно.

Часто система обслуживания содержит несколько обслуживающих каналов, между которыми можно выбирать. По числу каналов обслуживания СМО делятся *на одноканальные* и *многоканальные*. Примером могут служить рабочие места кассиров в супермаркетах или отдельные транспортные средства при транспортном обслуживании.

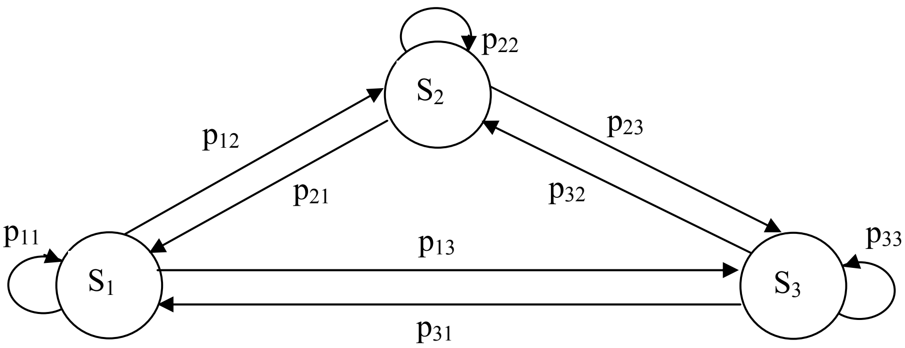
В зависимости от расположения источника требований системы могут быть *разомкнутыми* (источник заявок находится вне системы) и *замкнутыми* (источник находится в самой системе).

Обслуживание как марковский случайный процесс

Случайный процесс называется марковским, если для любого момента времени вероятностные характеристики процесса зависят только от его состояния в данный момент, но не зависят от того, когда и как пришла система в это состояние. Тогда при прогнозировании будущего не требуется учитывать прошлое. Заметим, что любой случайный процесс можно рассматривать как марковский, но число учитываемых при этом состояний может оказаться слишком большим. Например, выделим два состояния автомобиля: исправен или неисправен. Пусть нужно определить вероятность того, что удастся проехать еще некоторое расстояние. Очевидно, что процесс не является марковским, так как вероятность отказа зависит от величины пробега после выпуска или капитального ремонта. Если этот пробег считать одним из параметров состояния, то процесс можно рассматривать как марковский. Таким образом, любой процесс можно представить как марковский, но число состояний может получиться весьма большим.

При исследовании операций часто используются марковские процессы с дискретными состояниями, т. е. предполагается, что все возможные состояния системы можно перечислить, переход системы из одного состояния в другое происходит практически мгновенно, а вероятности переходов известны. Рассмотрим процесс с дискретным шагом по времени, учитывая, что при малом шаге можно достаточно точно аппроксимировать непрерывный процесс. Пример такого процесса с тремя состояниями приведен на рис. Для аналитического описания используется матрица переходов (матрица вероятностей):

$$\mathrm{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}.$$



Рассмотрим в отдельности элементы СМО.

Входящий поток: на практике наиболее распространенным является *простейший* поток заявок, обладающий свойствами стационарности, ординарности и отсутствия последействия.

Стационарность характеризуется тем, что вероятность поступления определенного количества требований (заявок) в течение некоторого промежутка времени зависит только от длины этого промежутка. Вероятность хотя бы одного события за малый промежуток времени Δt в пропорциональна длине этого промежутка: $p(\Delta t) \approx \lambda \Delta t$.

Ординарность потока определяется невозможностью одновременного появления двух или более заявок.

Отсутствие последействия характеризуется тем, что поступление заявки не зависит от того, когда и сколько заявок поступило до этого момента. В этом случае вероятность того, что число заявок, поступивших на обслуживание за промежуток времени t, равно k, определяется по закону Пуассона

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

где λ — *интенсивность потока заявок*, т.е. среднее число заявок в единицу времени:

$$\lambda = 1/\bar{\tau} \text{ (чел./мин, р./ч, автом./дн., квт/ч)},$$

где r — среднее значение интервала времени между двумя соседними заявками.

В большинстве систем массового обслуживания время между последовательными поступлениями заявок и время их обслуживания, будучи случайными, описываются показательным распределением:

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0,$$

где λ – параметр распределения имеет смысл интенсивности потока (среднее число заявок, приходящееся на единицу времени), а средний интервал времени между последовательными событиями равен

$$\int_0^\infty t \lambda e^{-\lambda t} dt = \frac{1}{\lambda}.$$

Случайное время ожидания в очереди начала обслуживания также считают распределенным экспоненциально:

$$f(t) = \nu e^{-\nu t},$$

где v — *интенсивность движения очереди*, т.е. среднее число заявок, приходящих на обслуживание в единицу времени:

$$\nu = 1/\bar{t}_{оч},$$

где tоч — среднее значение времени ожидания в очереди.

Выходящий поток заявок связан с потоком обслуживания в канале, где длительность обслуживания tобс является случайной величиной и часто подчиняется показательному закону распределения с плотностью

$$f(t_{обс}) = \mu e^{-\mu t},$$

где μ — интенсивность потока обслуживания, т.е. среднее число заявок, обслуживаемых в единицу времени:

$$\mu = 1/\bar{t}_{обс} \text{ (чел./мин, р./дн., кг/ч, докум./дн.)},$$

где tобс — среднее время обслуживания.

Важной характеристикой СМО, объединяющей λ и μ, является *интенсивность нагрузки*

$$\rho = \lambda/\mu.$$

Аналитический расчет СМО. Уравнения Колмогорова

Одной из важнейших характеристик СМО является длина очереди. Если она велика, то в условиях рыночной экономики это потеря потенциальных заявок (клиентов), а следовательно, и снижение конкурентоспособности. Длина очереди – это случайная величина, но ее среднюю длину можно рассчитать также, как и другие важные характеристики, а именно:

- вероятность отклонения заявки;
- вероятность обслуживания заявки;
- среднее число заявок, обслуживаемых за единицу времени;
- среднее время пребывания заявки в системе.

Вывод необходимых аналитических формул основывается на уравнениях Колмогорова. В этих дифференциальных уравнениях неизвестными функциями являются вероятности состояний системы. Для примера ограничимся рассмотрением системы с тремя возможными состояниями, а поток событий будем предполагать простейшим. Тогда вероятность перехода системы из *i*-го состояния в *j*-е состояние за малый промежуток времени Δt будет $p_{ij}(\Delta t) \approx \lambda_{ij}\Delta t$, где λ_{ij} – плотность вероятности перехода. Пусть $p_i(t),\ i = (\overline{1,3})$ – вероятность нахождения системы в *i*-м состоянии в момент *t*. Найдем вероятность того, что в момент $t + \Delta t$ система будет в состоянии S_1. Вероятность того, что система уже была в этом состоянии и за время Δt оно не изменилось, равна $p_1(t)p_{11}$. Все построчные суммы в матрице переходов равны единице (система всегда находится в каком-то состоянии), значит, $p_{11} = 1 - p_{12} - p_{13}$. Тогда вероятность сохранения состояния S_1 запишется как $p_1(t)(1 - p_{12} - p_{13})$.

Но система может оказаться в состоянии S_1 в результате перехода в него из любого иного состояния, что может произойти с вероятностью

$p_2(t)p_{21} + p_3(t)p_{31}$. Суммируя (по правилу сложения вероятностей) вероятности этих двух вариантов, получим:

$$p_1(t+\Delta t) = p_1(t) + p_2(t)p_{21} + p_3(t)p_{31} - p_1(t)(p_{12} + p_{13}).$$

Теперь можно найти прирост вероятности первого состояния системы за время Δt, выражая вероятности переходов через интенсивности потоков событий:

$$p_1(t+\Delta t) - p_1(t) = \Delta t\left(p_2(t)\lambda_{21} + p_3(t)\lambda_{31} - p_1(t)(\lambda_{12} + \lambda_{13})\right)$$

.

После деления на Δt и перехода к пределу при $\Delta t \to 0$ получим первое дифференциальное уравнение системы:

$$\frac{dp_1}{dt} = \lambda_{21}p_2 + \lambda_{31}p_3 - p_1(\lambda_{12} + \lambda_{13}).$$

Два других уравнения запишутся аналогично. Если эти же вычисления выполнить для системы с числом возможных состояний *N,* то получим соответствующее число дифференциальных уравнений вида:

$$\frac{dp_j}{dt} = \left[\sum_{i\neq j}\lambda_{ij}p_i - p_j\sum_{i\neq j}\lambda_{ij}\right],\ j = \overline{1,N}.$$

Эти уравнения выражают производные от вероятности *j*-го состояния системы как разности двух сумм. Первая сумма – это скалярное произведение вектора состояний системы на вектор интенсивностей потоков, переводящих систему в *j*-е состояние. Из нее вычитается сумма интенсивностей всех потоков, выводящих систему из *j*-го состояния, умноженная на вероятность этого состояния. Полученную систему уравнений можно дополнить очевидным соотношением $\sum_{j=1}^{N} p_j = 1$, и одно из уравнений можно опустить. Полученная система дифференциальных уравнений носит имя **Колмогорова.** Решение задачи Коши для указанной системы позволяет определить функции вероятностей состояний системы в произвольный момент времени. Но для этого надо задать начальное условие – исходное состояние системы.

Финальные вероятности состояний СМО

Логично ожидать, что если за процессом наблюдать достаточно долго, то можно будет определить вероятность нахождения системы в каждом из ее возможных состояний. Они называются *финальными вероятностями состояний.* В теории случайных процессов доказывается, что если число состояний конечно и за конечное число шагов можно перейти из любого состояния в любое, то финальные вероятности не только существуют, но и не зависят от начального состояния. Простым способом приближенного численного расчета таких систем является метод статистического моделирования. Вероятности перехода из текущего состояния в любое возможное можно воспроизвести, используя датчик случайных чисел.

Отношение числа шагов, в течение которых система находится в состоянии S_i, к общему числу шагов будет стремиться к финальной вероятности p_i.

Решение задачи Коши для системы уравнений Колмогорова при заданных начальных условиях приводит к получению системы функций времени:

$$P_s(t), \quad s = 0, R.$$

Это позволяет получить дискретное распределение вероятностей на множестве состояний СМО для любого значения t >=0. Анализ решения показывает, что для достаточно больших значений t, независимо от начальных условий, это распределение стабилизируется и практически не зависит от времени. Указанное асимптотическое свойство решения означает, что с течением времени функционирование СМО переходит в стационарный режим, соответствующий постоянству значений функций:

$$\lim p_s(t) = \text{const} = p_s, \quad s = \overline{0, R}.$$

Система значений вероятностей состояний, соответствующих стационарному режиму работы СМО и будет *финальными вероятностями*.

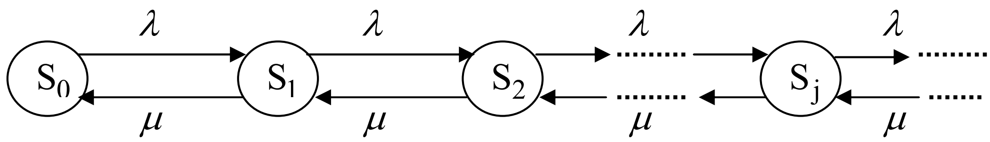
Полагая в системе дифференциальных уравнений

$$\frac{dp_s(t)}{dt} = 0,$$

получаем для определения финальных вероятностей систему алгебраических уравнений.

Пример расчета СМО

Очень часто все состояния системы СМО можно представить в виде цепочки, в которой каждое промежуточное состояние связано прямой и обратной стрелками только с двумя соседними, а крайние – только с одним соседним. Если имеется только один канал обслуживания и на длину очереди не накладывается никаких ограничений, то интенсивность входного потока λ не зависит от числа заявок в очереди. Интенсивность обслуживания μ также можно считать постоянной величиной. На рис. индекс *j* состояния системы S_j соответствует числу заявок в очереди, например, состояние S_0 означает, что канал свободен.



При расчете СМО удобно ввести обозначение для величины относительной нагрузки на систему $\rho = \dfrac{\lambda}{\mu}$, которая показывает отношение интенсивностей входного и выходного потоков. Чтобы все заявки могли быть выполнены, должно выполняться очевидное условие $\rho < n$, где n – число каналов обслуживания. Для одноканальной СМО $\rho < 1$, иначе при $t \to \infty$ очередь будет неограниченно возрастать. Обозначим вероятности состояний системы $p_0, p_1, p_2, ..., p_j$ и запишем для рассматриваемого случая уравнения Колмогорова:

$$\begin{cases} p_1\mu - p_0\lambda = 0 \\ p_2\mu - p_1\lambda = 0 \\ \dots \\ p_j\mu - p_{j-1}\lambda = 0 \\ \dots \end{cases}$$

Используя выражение для относительной нагрузки ρ, перепишем эти уравнения как

$$\begin{cases} p_1 = \rho p_0 \\ p_2 = \rho p_1 \\ \dots \\ p_j = \rho p_{j-1} \\ \dots \end{cases} \Rightarrow \begin{cases} p_1 = \rho p_0 \\ p_2 = \rho^2 p_0 \\ \dots \\ p_j = \rho^j p_0 \\ \dots \end{cases}$$

Поскольку сумма вероятностей всех состояний системы равна единице, то после суммирования получим:

$$1 - p_0 = \rho p_0 + \rho^2 p_0 + \dots + \rho^j p_0 + \dots \text{ или}$$
$$p_0 + \rho p_0 + \rho^2 p_0 + \dots + \rho^j p_0 + \dots = 1,$$

откуда $$p_0\left(1 + \rho + \rho^2 + \dots + \rho^j + \dots\right) = 1.$$

Используя формулу для геометрической прогрессии, имеем $p_0 \dfrac{1}{1-\rho} = 1$, что позволяет записать вероятности состояний системы в виде:

$$p_0 = 1 - \rho,\ p_1 = \rho(1-\rho),\ p_2 = \rho^2(1-\rho),\ \dots,\ p_j = \rho^j(1-\rho),\ \dots .$$

Если вероятность свободного состояния системы $p_0 = 1 - \rho$ можно было предвидеть, то утверждать, что из всех возможных состояний системы оно самое вероятное даже при ρ, близком к единице, до проведения вычислений не было оснований. Чтобы вычислить математическое ожидание числа заявок в системе $L_{сист}$, умножим вероятность каждого состояния на число заявок для этого состояния и найдем предел суммы бесконечного ряда:

$$L_{сист} = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + \ldots + j \cdot p_j + \ldots = \sum_{j=1}^{\infty} j \cdot p_j$$

Подставляя значения вероятностей, последовательными преобразованиями получим:

$$L_{сист} = \sum_{j=1}^{\infty} j \cdot p_j = (1-\rho)\sum_{j=1}^{\infty} j \cdot \rho^j = \rho(1-\rho)\sum_{j=1}^{\infty} j \cdot \rho^{j-1} = \rho(1-\rho)\sum_{j=1}^{\infty} \frac{d\rho^j}{d\rho}$$

.

Изменив порядок выполнения операций дифференцирования и суммирования, можем записать:

$$L_{сист} = \rho(1-\rho)\frac{d}{d\rho}\sum_{j=1}^{\infty}\rho^j = \rho(1-\rho)\frac{d}{d\rho}\frac{\rho}{1-\rho} = \rho(1-\rho)\frac{1}{(1-\rho)^2} = \frac{\rho}{(1-\rho)}$$

.

Те заявки, которые находятся в системе, но не обслуживаются, образуют очередь. Вероятность простоя системы $p_0 = 1-\rho$, следовательно, математическое ожидание числа заявок под обслуживанием $L_{обсл} = \rho$, а длина очереди будет

$$L_{очер} = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho} \quad .$$

СМО с отказами

Для большого числа задач по расчету СМО известны конечные формулы, полученные путем, аналогичным приведенному выше. Например, случай, когда имеется *n* каналов обслуживания, но очередь не допускается (*n*-канальная СМО с отказами), называют задачей Эрланга. Для такой системы число состояний $n+1$ соответствует числу занятых каналов, а состояние S_0 – отсутствию заявок. В этом случае интенсивность перехода к состоянию с меньшим числом занятых каналов пропорциональна числу используемых каналов. В расчетные формулы входит интенсивность потока заявок $\rho = \dfrac{\lambda}{\mu}$, которая показывает среднее число заявок, поступающих за время обслуживания одной заявки. Вероятность отсутствия заявок (свободного состояния системы) будет:

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

Вероятность того, что заявка будет обслужена (применительно к задаче телефонии – это вероятность получить соединение с АТС): $q = 1 - p_0 \dfrac{\rho^n}{n!}$, среднее число занятых каналов: $k = \rho q$.

Если определить величину затрат на увеличение числа каналов обслуживания и прирост дохода от увеличения числа обслуженных заявок, то получим возможность оптимального решения задачи. Такую задачу приходится решать, например, провайдеру, обслуживающему Интернет.

Основные понятия

Заявка, поступившая в систему с отказами и нашедшая все каналы занятыми, получает отказ и покидает систему необслуженной. Показателем качества обслуживания выступает вероятность получения отказа. Предполагается, что все каналы доступны в равной степени всем заявкам, входящий поток является простейшим, длительность (время) обслуживания одной заявки (tобс) распределена по показательному закону.

Формулы для расчета установившегося режима

1. Вероятность простоя каналов обслуживания, когда нет заявок (k = 0):

$$P_0 = 1/\sum_{k=0}^{n} \rho^k/k!$$

2. Вероятность отказа в обслуживании, когда поступившая на обслуживание заявка найдет все каналы занятыми (k = n):

$$P_{отк} = P_n = P_0\rho^n/n!$$

3. Вероятность обслуживания:

$$P_{обс} = 1 - P_{отк}.$$

4. Среднее число занятых обслуживанием каналов:

$$\bar{n}_з = \rho P_{обс}.$$

5. Доля каналов, занятых обслуживанием:

$$k_з{}' = \bar{n}_з/n$$

6. Абсолютная пропускная способность СМО:

$$A = \lambda P_{обс}.$$

СМО с неограниченным ожиданием

Для случая неограниченной длины очереди $(m = \infty)$ вероятность свободного состояния:

$$p_{0(m=\infty)} = \left[\sum_{j=0}^{n} \frac{\rho^j}{j!} + \frac{\rho^{n+1}}{n!(n-\rho)} \right]^{-1}.$$

Средняя длина очереди:

$$L_{очер} = p_0 \frac{\rho^{n+1}}{n!n\left(1-\dfrac{\rho}{n}\right)^2}.$$

Основные понятия

Заявка, поступившая в систему с неограниченным ожиданием и нашедшая все каналы занятыми, становится в очередь, ожидая освобождения одного из каналов.

Основной характеристикой качества обслуживания является время ожидания (время пребывания заявки в очереди).

Для таких систем характерно отсутствие отказа в обслуживании, т.е. Ротк = 0 и Робс = 1.

Для систем с ожиданием существует дисциплина очереди:

1) обслуживание в порядке очереди по принципу "первым пришел — первым обслужен" ;

2) случайное неорганизованное обслуживание по принципу "последний пришел — первым обслужен";

3) обслуживание с приоритетами по принципу "генералы и полковники вне очереди".

Формулы для установившегося режима

1. Вероятность простоя каналов, когда нет заявок (k = 0):

$$P_0 = 1 \Big/ \sum_{k=0}^{n} (\rho^k / k!) + \rho^{n+1} / n!(n-\rho).$$

Предполагается, что ρ/n < 1.

2. Вероятность занятости обслуживанием к заявок:

$$P_k = \rho^k P_0 / k!, \quad 1 \leqslant k \leqslant n.$$

3. Вероятность занятости обслуживанием всех каналов:

$$P_n = \rho^n P_0 / n!$$

4. Вероятность того, что заявка окажется в очереди:

$$P_{\text{оч}} = \frac{\rho^{n+1}}{n!(n-\rho)} P_0.$$

5. Среднее число заявок в очереди:

$$\bar{L}_{\text{оч}} = \frac{\rho^{n+1}}{(n-1)!(n-\rho)^2} P_0.$$

6. Среднее время ожидания заявки в очереди:

$$\bar{t}_{\text{оч}} = \bar{L}_{\text{оч}} / \lambda.$$

7. Среднее время пребывания заявки в СМО:

$$\bar{t}_{\text{смо}} = \bar{t}_{\text{оч}} + \bar{t}_{\text{обс}}.$$

8. Среднее число занятых обслуживанием каналов:

$$\bar{n}_{\text{з}} = \rho.$$

9. Среднее число свободных каналов:

$$\bar{n}_{\text{св}} = n - \bar{n}_{\text{з}}.$$

10. Коэффициент занятости каналов обслуживания:

$$k_{\text{з}} = \bar{n}_{\text{з}} / n.$$

11. Среднее число заявок в СМО:

$$\bar{z} = \bar{L}_{\text{оч}} + \bar{n}_{\text{з}}.$$

СМО с ожиданием и с ограниченной длиной очереди

Основные понятия

Заявка, поступившая в систему с ожиданием с ограниченной длиной очереди и нашедшая все каналы и ограниченную очередь занятыми, покидает систему необслуженной.

Основной характеристикой качества системы является отказ заявке в обслуживании.

Ограничения на длину очереди могут быть из-за:

1) ограничения сверху времени пребывания заявки в очереди;

2) ограничения сверху длины очереди;

3) ограничения общего времени пребывания заявки в системе.

Формулы для установившегося режима

1. Вероятность простоя каналов обслуживания, когда нет заявок (k = 0):

$$P_0 = 1 : \left\{ \sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)} \left[1 - \left(\frac{\rho}{n} \right)^m \right] \right\}.$$

2. Вероятность отказа в обслуживании:

$$P_{\text{отк}} = \frac{\rho^{n+m}}{n! n^m} \cdot P_0.$$

3. Вероятность обслуживания:

$$P_{\text{обс}} = 1 - P_{\text{отк}}.$$

4. Абсолютная пропускная способность:

$$A = P_{\text{обс}} \cdot \lambda.$$

5. Среднее число занятых каналов:

$$\bar{n}_{\text{з}} = \frac{A}{\mu}.$$

6. Среднее число заявок в очереди:

$$\bar{L}_{\text{оч}} = \frac{\rho^{n+1}}{n \cdot n!} \frac{1 - (\rho/n)^m (m + 1 - m\rho/n)}{(1 - \rho/n)^2} P_0.$$

7. Среднее время ожидания обслуживания:

$$\bar{t}_{\text{оч}} = \frac{\bar{L}_{\text{оч}}}{\lambda}.$$

8. Среднее число заявок в системе:

$$\bar{z} = \bar{L}_{\text{оч}} + \bar{n}_{\text{з}}.$$

9. Среднее время пребывания в системе:

$$\bar{t}_{\text{смо}} = \frac{\bar{z}}{\lambda}.$$

Определение эффективности использования трудовых и производственных ресурсов в системах массового обслуживания

Рассмотрим задачу с использованием СМО с отказами.

Пример. В ОТК цеха работают три контролера. Если деталь поступает в ОТК, когда все контролеры заняты обслуживанием ранее поступивших деталей, то она проходит непроверенной. Среднее число деталей, поступающих в ОТК в течение часа, равно 24, среднее время, которое затрачивает один контролер на обслуживание одной детали, равно 5 мин. Определить вероятность того, что деталь пройдет ОТК необслуженной, насколько загружены контролеры и сколько их необходимо поставить, чтобы Р*обс>=0,95 (* — заданное значение Робс).

Решение. По условию задачи λ= 24 дет./ч = 0,4 дет./мин, tобс = 5 мин, тогда μ = 0,2, ρ= 2.

1. Вероятность простоя каналов обслуживания:

$$P_0 = \frac{1}{2^0/0! + 2^1/1! + 2^2/2! + 2^3/3!} = \frac{1}{1+2+2+1,3} = 0{,}1587,$$

где 0! = 1.

2. Вероятность отказа в обслуживании:

Ротк = 8• 0,1587/3! = 0,21.

3. Вероятность обслуживания:

Робс = 1 - 0,21 = 0,79.

4. Среднее число занятых обслуживанием каналов:

n3 = 2 • 0,79 = 1,58.

5. Доля каналов, занятых обслуживанием:

k3 = 1,58/3 = 0,526.

6. Абсолютная пропускная способность:

A = 0,4 • 0,79 = 0,316.

При n = 3 Робс = 0,79 <= Р*обс = 0,95. Произведя аналогичные расчеты для n = 4, получим

Ро = 0,14, Ротк = 0,093, Робс = 0,907.

Так как Робс = 0,907 <= Р*обс = 0,95, то, произведя расчеты для n = 5, получим

$$P_0 = 0{,}137, \quad P_{\text{отк}} = 0{,}035, \quad P_{\text{обс}} = 0{,}965 \geqslant P^*_{\text{обс}} = 0{,}95.$$

Ответ. Вероятность того, что при n = 3 деталь пройдет ОТК необслуженной, составляет 21%, и контролеры будут заняты обслуживанием на 53%.

Чтобы обеспечить вероятность обслуживания более 95%, необходимо не менее пяти контролеров.

Рассмотрим задачу с использованием СМО с неограниченным ожиданием.

Пример. Сберкасса имеет трех контролеров-кассиров (n= 3) для обслуживания вкладчиков. Поток вкладчиков поступает в сберкассу с интенсивностью λ= 30 чел./ч. Средняя продолжительность обслуживания контролером-кассиром одного вкладчика tобс = 3 мин.

Определить характеристики сберкассы как объекта СМО.

РЕШЕНИЕ. Интенсивность потока обслуживания μ= 1/tобс = 1/3 = 0,333, интенсивность нагрузки p = 1,5.

1. Вероятность простоя контролеров-кассиров в течение рабочего дня:

$$P_0 = \frac{1}{\frac{1,5^0}{0!} + \frac{1,5^1}{1!} + \frac{1,5^2}{2!} + \frac{1,5^3}{3!} + \frac{1,5^4}{3!(3-1,5)}} = 0{,}210.$$

2. Вероятность застать всех контролеров-кассиров занятыми:

$$P_n = \frac{1,5^3}{3!}\, 0{,}21 = 0{,}118.$$

3. Вероятность очереди:

$$P_{оч} = \frac{1{,}5^4}{3!(3-1{,}5)}0{,}21 = 0{,}118.$$

4. Среднее число заявок в очереди:

$$\bar{L}_{оч} = \frac{1{,}5^4}{(3-1)!(3-1{,}5)^2}0{,}21 = 0{,}236.$$

5. Среднее время ожидания заявки в очереди:

$$\bar{t}_{оч} = \frac{0{,}236}{0{,}5} = 0{,}472 \text{ мин.}$$

6. Среднее время пребывания заявки в СМО

$t_{смо}$=0,472+3=3,472 мин

7. Среднее число свободных каналов

$n_{св}$=3-1,5=1,5

8. Коэффициент заняточти каналов обслуживания

$k_з$=1,5/3=0,5

9. Среднее число посетителей в сберкассе

z=0.236+1.5=1.736

Ответ. Вероятность простоя контролеров-кассиров равна 21% рабочего времени, вероятность посетителю оказаться в очереди составляет 11,8%, среднее число посетителей в очереди 0,236 чел., среднее время ожидания посетителями обслуживания 0,472 мин.

Заключение

Аналитические зависимости разработаны для большого числа задач, но они всегда предполагают марковский характер процесса и определяют только стационарное состояние системы. Поэтому получил широкое распространение метод имитационного моделирования СМО, который позволяет анализировать системы с любой структурой. Он основан на многократном «проигрывании» функционирования системы. При этом как момент поступления очередной заявки, так и продолжительность ее обслуживания (если она будет обслуживаться) задаются конкретными значениями, полученными на основе датчиков псевдослучайных чисел. С малым шагом по времени рассчитываются последовательные состояния системы. Если накопить достаточное количество таких реализаций поведения системы, то обычной статистической обработкой можно найти искомые характеристики системы с желаемой степенью точности. Для сложных систем метод статистического моделирования, как правило, оказывается проще аналитического подхода, а в ряде случаев ему просто нет альтернативы.

Существуют пакеты прикладных программ, которые позволяют задавать структуру моделируемой системы и устанавливать параметры потоков случайных событий. В случае необходимости можно использовать языки программирования, предназначенные для таких задач. Если ограничиться использованием такой программы, как *Excel*, то посредством ее встроенных средств просто получить на отдельных листах реализации случайных потоков, но расчет их взаимодействия придется выполнять своими силами.