## 1. Предмет теории массового обслуживания. Основные понятия теории.

**Теория массового обслуживания (ТМО)** — научная дисциплина, занимающаяся математическим моделированием постоянно повторяющихся однотипных задач в одних и тех же условиях их появления, анализом способов их решения в целях повышения эффективности этих решений.

**Система массового обслуживания (СМО)** — система, пред назначенная для многократно повторяющегося (многоразового) использования при решении однотипных задач. Процесс обслуживания — решение такой системой возникающей типовой задачи.

**Задача обслуживания** — повторяющаяся типовая задача, возникающая в СМО.

**Заявка (требование)** — объект обслуживания в СМО, для которого решается типовая задача обслуживания.

**Поток заявок** — последовательность объектов, подлежащих обслуживанию в СМО.

**Канал обслуживания** — обслуживающая единица, т. е. элемент системы, предназначенный для решения задачи обслуживания.

**Очередь** — последовательность заявок, ожидающих обслуживания.

**Дисциплина обслуживания** — правила, определяющие порядок выбора заявок из числа поступивших и порядок их распределения между каналами обслуживания.

**Показатели эффективности СМО** — показатели, описывающие способность СМО справляться с потоком заявок.

**Структура СМО**

В структуре СМО выделяют следующие элементы:

* входящий поток заявок;
* очередь (если ее наличие предусмотрено системой);
* каналы обслуживания;
* выходящий поток заявок.

Выходящий поток может делиться на поток обслуженных заявок и поток заявок, оставшихся без обслуживания.

Последний, в свою очередь, может делиться на:

* поток заявок, которым было отказано в обслуживании;
* поток заявок, покинувших систему из‑за превышения лимита времени нахождения в системе.

**Предметом теории массового обслуживания** является построение математических моделей, связывающих заданные условия работы СМО (число каналов, их производительность, характер потока заявок, организация работы и т. п.) с показателями эффективности функционирования СМО.

## 2. Классификация систем массового обслуживания.

**По типу обслуживания выделяют два основных класса СМО:**

* **СМО с отказами** — СМО, в которых заявка, поступившая в тот момент, когда все каналы СМО заняты, получает отказ в обслуживании, покидает СМО и в дальнейшем процессе обслуживания не участвует. Эти СМО называют также системами без очереди;
* **СМО с ожиданием** — СМО, в которых заявка, поступившая в момент, когда все каналы системы заняты, не уходит, а становится в очередь на обслуживание. Эти СМО называют также система ми с очередью.

**СМО с ожиданием подразделяются на:**

**– СМО с неограниченным ожиданием** (с неограниченной очередью);

**– СМО с ограниченным ожиданием двух типов**: с ограничением на длину очереди и с ограничением на время пребывания заявки в очереди (или на общее время пребывания заявки в системе).

**По характеру поступающего потока** заявок СМО делят на:

* немарковские системы;
* системы.

**По числу каналов СМО делят на:**

* одноканальные;
* многоканальные.

**По количеству этапов обслуживания их делят на:**

* **однофазные** — заявка проходит только одну стадию (фазу) обработки, все каналы выполняют одну и ту же операцию;
* **многофазные** — заявка проходит последовательно несколько стадий обработки в системе, каждый канал проводит заявку через все необходимые ей фазы.

По схеме обслуживания заявок из очереди **(по дисциплине очереди)** СМО делят на:

* СМО с упорядоченной очередью (обслуживание в соответствии с определенным порядком);
* СМО с неупорядоченным (случайным) выбором заявок из очереди;
* СМО с обслуживанием заявок в соответствии с приоритетом.

**По ограничению потока заявок выделяют:**

* **открытые СМО**, в них заявки поступают в неограниченном количестве из источников вне самой системы;
* **замкнутые СМО** (количество источников ограничено), в которых источники заявок включены в саму систему; заявка из такого источника, покинувшая систему, через какое‑то время может снова в нее возвратиться.

**По времени обслуживания заявок рассматривают:**

* СМО со случайным временем обслуживания каждой заявки;
* СМО с заданным временем обслуживания заявок, которое может быть одинаковым для всех заявок и разным для разных заявок.

**По способу формирования множества заявок выделяют**:

* СМО с фиксированной очередью. В них очередь формируется до начала работы системы и в процессе ее работы новые заявки в очередь не поступают;
* СМО с потоком заявок. Заявки поступают в такую систему в продолжение всего времени работы системы. Поток заявок в этом случае может быть:
* **регулярным,** когда заявки поступают в систему в заранее фиксированные моменты времени или через определенные промежутки времени
* **случайным,** когда заявки поступают в систему в случайные, заранее не определенные моменты времени

## 3. Задача минимизации штрафа за задержку обслуживания.

Будем рассматривать один пункт обслуживания с поступившим на него определенным количеством заявок (n зая вок), т. е. одноканальную СМО с образовавшейся очередью. Каждая заявка имеет свое определенное время обслуживания. Задержка обслуживания каждой заявки влечет за собой определенный для каждой заявки штраф. Необходимо так выстроить последовательность обслуживания имеющихся заявок, чтобы минимизировать суммарный штраф. Примеры подобных систем: грузовой причал морского порта с пришедшими к нему для разгрузки судами, которые имеют разный объем с различными по сроку хранения и реализации грузами. В силу этого у них отличаются время обслуживания и штраф за задержку обслуживания на единицу времени (час, сутки и т. п.).

Итак, имеется один пункт обслуживания и n ожидающих обслуживания заявок. Каждая заявка имеет свое время обслуживания и штраф за его задержку. Обозначим время обслуживания i‑й заявки как Ti , а штраф за задержку обслуживания этой заявки на единицу времени (минута, час, сутки и т. п.) — Ci ; № — номер заявки. Сведем все данные в табличную форму (см. ниже).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | 1 | 2 | 3 | …. | n |
| T\_{i} | T\_{1} | T\_{2} | T\_{3} | …. | T\_{n} |
| C\_{i} | C\_{1} | C\_{2} | C\_{3} | …. | C\_{n} |

Показатель эффективности рассматриваемой СМО — суммарный штраф за задержку обслуживания. Разные последовательности обслуживания заявок дают в общем случае разный суммарный штраф, т. е. разную эффективность СМО. Решение задачи сводится к определе нию такого порядка обслуживания (установлению такой последова тельности обслуживания заявок), который обеспечивает минимальный суммарный штраф за задержку обслуживания всех этих заявок. Такие задачи называются задачами упорядочения.

Отношение C\_{i}/T\_{i} будем называть относительным штрафом i‑й заявки.

Таким образом, чтобы последовательность обслуживания заявок была оптимальной, требуется выполнение последовательности неравенств C\_{1}/T\_{1} >= C\_{2}/T\_{2} >= … >=C\_{n}/T\_{n}

Отсюда следует правило построения оптимальной последовательности обслуживания заявок, обеспечивающей минимальный суммарный штраф за задержку обслуживания: первыми должны обслуживаться заявки, имеющие больший относительный штраф

## 4. Задача «директора» (задача одного станка).

Если в задаче минимизации штрафа за задержку обслуживания штраф каждой заявки одинаков и равен некоторой постоянной величине С, т. е. для всех заявок Сi = C, то задача называется задачей одного станка, или задачей «директора». В этом случае неравенство C\_{k} T\_{k+1} ≥ C\_{k+1} T\_{k} , выполнение которого для любого k = 1, 2, 3, …, n – 1 определяет оптимальную последовательность обслуживания заявок предыдущей задачи, принимает вид С T\_{k+1} ≥ C T\_{k} . Обе части этого неравенства можно поделить на С > 0, тогда оно примет вид T\_{k+1} ≥ T\_{k} . Следовательно, выполнение цепочки неравенств,

T\_{1} ≤ T\_{2} ≤ T\_{3} ≤ … ≤ T\_{n}, определяет оптимальную последовательность обслуживания таких заявок. Правило построения оптимальной последовательности обслуживания заявок, обеспечивающей минимальный суммарный штраф за задержку обслуживания в данной ситуации, формулируется так: первыми должны обслуживаться заявки, имеющие меньшее время обслуживания.

Суммарный штраф равен произведению суммарного времени ожидания всех заявок очереди на величину С штрафа за задержку обслуживания каждой заявки на единицу времени Если С = 1, то суммарный штраф численно равен суммарному времени ожидания обслуживания всех заявок и его минимизация сводится к минимизации этого времени. В этом случае задачу одного станка (задачу «директора») можно сформулировать таким образом: имеется n заявок, каждая из которых имеет свое время обслуживания Ti . Требуется так организовать их обслуживание, т. е. найти такую последовательность обработки этих заявок, чтобы суммарное время ожидания обслуживания всех заявок было минимальным.

## 5. Задача двух станков. Алгоритм Джонсона.

Рассмотрим обслуживание заявок, проходящих две фазы обработ ки: сначала заявка обрабатывается на пункте А и только после окон чания такой обработки она обрабатывается на пункте В. Такую ситуацию можно интерпретировать как обработку детали на двух станках. Например, сначала деталь обтачивается на токарном станке, а затем идет сверление детали на сверлильном станке. Вторая операция может начаться только по завершении первой. Пусть имеется очередь из n таких заявок. Время обработки каждой детали на каждом из станков свое.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | 1 | 2 | 3 | …. | n |
| A | a\_{1} | a\_{2} | a\_{3} | …. | a\_{n} |
| B | b\_{1} | b\_{2} | b\_{3} | …. | b\_{n} |

Задача двух станков, т. е. задача минимизации общего времени обработки заявок (деталей) на двух пунктах последовательного обслуживания (станках), решается с помощью алгоритма Джонсона.

Алгоритм заключается в следующем: рассматриваем таблицу с исходными данными, где в строку А занесены аi , а в строку В, соответственно, занесены bj :

ai — время обработки i‑й детали на станке A, i = 1, …, n;

bj — время обработки j‑й детали на станке В, j = 1, …, n.

Рассмотрим алгоритм подробнее.

Среди всех аi и bj выбирается наименьшее число m = min {аi ; bj }.

1. Если m находится в строке А, т. е. m = ak , то k‑я деталь обрабатывается первой: m = ak => деталь № k обслуживается первой. Если m находится в строке В, т. е. m = bk , то k‑я деталь обрабатывается последней: m = bk => деталь № k обрабатывается последней.

2. Деталь № k вычеркивается из списка, и тот же алгоритм применяется к оставшимся деталям. Алгоритм повторяется до тех пор, пока не будут вычеркнуты все детали.

В результате получаем оптимальную последовательность обработки деталей.

## 6. Потоки событий. Простейший поток событий.

**Поток событий** — это последовательность однородных событий, следующих одно за другим в какие‑то моменты времени. Эти моменты времени могут быть фиксированными, а могут быть случайными.

Поток событий называется **регулярным**, если события следуют одно за другим через определенные фиксированные промежутки времени (в частности, равные). Например, по ток изделий на конвейере сборочного цеха, движущегося с постоянной скоростью.

Поток событий называется **случайным**, если появления событий происходят в случайные моменты времени. Например, поток вызовов на АТС, поток покупателей в супермаркете, поток отказов ЭВМ и т. п.

Поток характеризуется интенсивностью λ — частотой появления событий, т. е. средним числом событий, происходящих в единицу времени.

Поток называется **стационарным**, если его вероятностные характеристики не зависят от времени (не меняются с течением времени). В частности, интенсивность стационарного потока есть величина постоянная: λ(t) = λ = const. Для стационарного потока вероятность появления определенного числа событий на промежутке времени продолжительностью t не зависит от момента времени t0 , с которого начинается отсчет этого промежутка. Эта вероятность зависит только от длительности t промежутка (t0 , t0 + t), но не от момента времени t0 , т. е. она не зависит от расположения этого промежутка на числовой оси. Неважно, с какого момента времени мы начинаем исследование стационарного потока — все его характеристики будут теми же самыми.

Поток событий называется потоком **без последействия**, если то, что происходило до момента t0 , никак не влияет на характеристики потока после момента t0 для любого значения t0. Иначе говоря, для любых двух непересекающихся промежутков времени число событий, попадающих на второй из них, никак не зависит от типа и числа событий, попавших на первый из них. Это число зависит только от интенсивности потока λ и длительности промежутка t.

Регулярный поток обладает последействием, в нем жестко фиксировано время следования одного события за другим, момент появления следующего события зависит от момента появления предыдущего. Поток прибытия поездов на станцию метро — это регулярный поток.

Поток событий называется **ординарным**, если вероятность появления на весьма малом (элементарном) участке времени Δt двух и более событий пренебрежимо мала по сравнению с вероятностью появления только одного события. Другими словами, поток событий ординарен, если события в нем появляются поодиночке, а не группами. Например, поток поездов, подходящих к станции метро, ординарен, а поток людей, выходящих из такого поезда, не ординарен.

Ординарный поток без последействия называется пуассоновским потоком. Стационарный пуассоновский поток называется простейшим потоком.

Таким образом, поток событий является простейшим, если он одновременно: стационарен; не имеет последействия; ординарен.

Регулярный поток не является простейшим потоком. Имеет место утверждение: При наложении (суперпозиции) достаточно большого числа n независимых, стационарных и ординарных потоков, сравнимых между собой по интенсивностям λi (i = 1, 2, …, n), получается близкий к простейшему поток с интенсивностью λ, равной сумме интенсивностей входящих потоков.

## 7. Случайные процессы. Марковский процесс.

**Случайным** (вероятностным или стохастическим) называется процесс изменения во времени состояния какой‑либо системы, происходящий в соответствии с вероятностными закономерностями, т. е. если эти изменения происходят в случайном порядке в случайные моменты времени.

Процесс называется процессом с дискретными состояниями, если его возможные состояния (S1 , S2 , S3 , …) можно заранее перечислить, а переход системы из одного состояния в другое происходит мгновенно (скачком).

Процесс называется процессом с непрерывным временем, если моменты возможных переходов системы из одного состояния в другое не фиксированы заранее, а случайны.

Случайный процесс называется **марковским процессом** (или случайным процессом без последействия), если для любого момента времени t0 вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент времени t0 и не зависят от того, когда и как система пришла в это состояние.

Процесс работы СМО с простейшим потоком заявок представляет собой марковский случайный процесс с дискретными состояниями и непрерывным временем. Это означает, что состояние СМО меняется скачком в случайные моменты времени, в моменты появления каких‑то событий (например, прихода новой заявки, окончания обслуживания заявки и т. п.).

Пример

Система состоит из двух узлов, каждый из которых в случайный момент времени может выйти из строя, после чего начинается немедленный ремонт этого узла, продолжающийся заранее неизвестное случайное время.

Состояния системы S обозначим S0 , S1 , S2 , S3 .

Состояние S0 — оба узла исправны; S1 — первый узел вышел из строя, второй исправен; S2 — второй узел вышел из строя, первый исправен; S3 — оба узла неисправны.

Здесь рисуем штуку с 4 прямоугольниками, внутри них S0 , S1 , S2 , S3. Прямоугольники соединены каждый к соседу 2 линиями, над линиями лямбда(интенсивность перехода системы в другое состояние). Получается замкнутый круг из состояний.

Вероятность i‑го состояния системы, т. е. вероятность того, что в момент t система будет находиться в состоянии Si , обозначим pi (t). Для любого момента времени t выполняется условие которое означает, что сумма вероятностей всех возможных в момент времени t состояний системы S равна единице, т. е. рассмотрены все возможные варианты состояний системы. Будем называть это условие условием нормировки.

## 8. Уравнения Колмогорова. Предельные вероятности.

**Уравнения Колмогорова** — это система дифференциальных уравнений, описывающих динамику вероятностей состояний однородного марковского процесса с непрерывным временем (непрерывной цепи Маркова).

Вероятность того, что за время Δt система перейдет из состояния Si в состояние Sj , как было показано ранее, приближенно (с точностью до Δt) равна произведению интенсивности перехода на длительность промежутка времени Δt.

Она равна вероятности того, что за время Δt произойдет хотя бы одно событие такого перехода. В итоге мы получим систему дифференциальных уравнений, описывающих работу указанной СМО. Такие уравнения называют уравнениями Колмогорова.

Система из 4 уравнений:

|  |  |
| --- | --- |
| | |  |
| | |  |
| | |  |
| | |  |

Аналогично составляется система уравнений Колмогорова при любом числе n возможных состояний СМО. В полученной системе независимых уравнений на единицу меньше числа имеющихся состояний СМО, т. е. общего числа возможных уравнений, т. к. если мы определим вероятности (n – 1) состояния СМО, то вероятность оставшегося состояния однозначно определится из условия равенства суммы всех вероятностей единице. Поэтому для поиска решения такой системы нужно одно из уравнений исключить, добавив вместо него условие нормировки p0 + p1 + p2 + … + pn = 1.

**Теорема существования предельных вероятностей**: если число состояний системы конечно и из каждого из них можно (за конечное число шагов) перейти в любое другое состояние, то предельные вероятности существуют. Примем эту теорему без доказательства. Предельная вероятность состояния Si имеет четкий смысл: она показывает среднее относительное время пребывания системы в дан ном состоянии. Поскольку предельные вероятности постоянны, то заменяя их производные нулевыми значениями, получим систему линейных алгебраических уравнений, описывающих стационарный режим работы. Для рассмотренного примера такая система имеет вид:

Система из 4 уравнений:

|  |  |
| --- | --- |
| | |  |
| | |  |
| | |  |
| | |  |

В каждом уравнении системы слева стоит предельная вероятность данного состояния pi = const, умноженная на суммарную интенсивность потоков, выходящих из данного состояния, а справа — сумма произведений интенсивностей всех потоков, входящих в данное состояние, на вероятности тех состояний, из которых эти потоки исходят.

## 9. Процесс гибели и размножения. Формулы для предельных вероятностей.

Рассмотрим в качестве объекта исследования биологическую популяцию. В ней происходят процессы гибели ее членов и рождения новых особей. Построим математическую модель изменения численности популяции.

Состояние популяции с некоторой минимальной численностью обозначим через S0. В процессе размножения численность популяции будет увеличиваться, в процессе гибели — уменьшаться. Рождение каждой особи увеличивает численность на единицу, смерть одного члена уменьшает численность популяции также на единицу. Через S1 обозначим состояние популяции при увеличении ее численности по сравнению с численностью в состоянии S0 на единицу, через Sk — увеличение ее численности на k единиц. Пере ход из какого‑либо состояния в соседнее означает рождение одной особи (переход вправо) или гибель одной особи (переход влево). Переходы могут осуществляться из любого состояния только в состояние с соседними номерами, т. е. из состояния Sk только либо в состояние Sk–1 (уменьшение численности на единицу), либо в состояние Sk+1 (увеличение численности на единицу). Будем считать, что все потоки, переводящие систему из одного состояния в другое, — простейшие с интенсивностями соответственно λk, k–1 или λk, k+1 .

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 𝛌01 |  | 𝛌12 |  | 𝛌23 |  |  |  | 𝛌k,k+1 |  | 𝛌n-1,n |  |
| S0 | ⇄ | S1 | ⇄ | S2 | ⇄ | … | ⇄ | Sk | ⇄ | … | ⇄ | Sn |
|  | 𝛌10 |  | 𝛌21 |  | 𝛌32 |  |  |  | 𝛌k-1,k |  | 𝛌n,n-1 |  |

p - Вероятность нахождения системы в каком-либо положении. Ниже формула для нахождения p0

Для нахождения остальных вероятностей

=

## 10. Одноканальная СМО с отказами. Расчет показателей эффективности.

Показатели эффективности:

А — абсолютная пропускная способность СМО, т. е. среднее число заявок, обслуживаемых в единицу времени;

Q — относительная пропускная способность СМО, т. е. средняя доля пришедших заявок, обслуживаемых системой;

Р\_{отк} — вероятность отказа, т. е. вероятность того, что заявка покинет СМО не обслуженной.

Система состоит из одного канала обслуживания. Очередь не допускается. Заявка, пришедшая в тот момент, когда канал занят обслуживанием, уходит из системы не обслуженной. Поток заявок и поток обслуживаний — простейшие.

Система может находиться только в двух состояниях: S0 — канал свободен и S1 — канал занят. Размеченный граф системы имеет следующий вид.

|  |  |  |
| --- | --- | --- |
|  | λ |  |
| S0 | ⇄ | S1 |
|  | μ |  |

Обозначим, что λ — интенсивность потока заявок, т. е. среднее число поступающих заявок в единицу времени; μ — интенсивность потока обслуживаний.

**Поток обслуживаний** — это поток заявок, обслуживаемых одним непрерывно занятым каналом. События этого потока — покидание обслуженной заявкой канала СМО. Момент такого покидания — это момент появления данного события. Поскольку продолжительность обслуживания разных заявок считается случайной величиной, то поток обслуживаний является случайным потоком. Будем считать, что это простейший поток.

Обозначим через μ интенсивность потока обслуживаний, т. е. среднее число заявок, обслуживаемых в единицу времени одним непрерывно работающим каналом.

Среднее время обслуживания t\_{об} = 1/μ — это среднее время, затрачиваемое на обслуживание одной заявки.

Найдем предельные вероятности состояний системы, существование которых очевидно, т. к. система имеет всего два состояния. По схеме процесса гибели и размножения

|  |  |
| --- | --- |
| | |  |
| | |  |

;

p0 — предельная вероятность того, что система свободна, т. е. это средняя доля времени, когда система может обслужить пришедшую заявку (заявке не будет отказано в обслуживании);

p1 — предельная вероятность того, что система занята обслуживанием имеющейся заявки, а значит пришедшей в это время очередной заявке будет отказано в обслуживании, она покинет СМО не обслуженной;

Q — относительная пропускная способность СМО, это доля пришедших на СМО заявок, которая будет обслужена; другими словами, это вероятность того, что пришедшая на СМО заявка будет обслужена. Она равна вероятности того, что ей не будет отказано в обслуживании. Вероятность отказа в обслуживании равна вероятности того, что система занята:

P\_{отк} = p\_{1} = ;

Q = 1 - P\_{отк} = 1 - p\_{1} = p\_{0}=

Теперь найдем абсолютную пропускную способность СМО. Это среднее число заявок, обслуживаемых СМО за единицу времени, т. е. то число заявок, которое составляет долю, определяемую относитель ной пропускной способностью, от общего числа поступающих в единицу времени заявок (от интенсивности потока заявок). Поэтому абсолютная пропускная способность А СМО равна

Выразим основные характеристики эффективности СМО через среднее время ожидания прихода очередной заявки и среднее время обслуживания заявки:

μ = 1/t\_{об} = t\_{ож}

λ = 1/t\_{ож} = t\_{об}

Q = t\_{ож}/(t\_{ож}+t\_{об})

P\_{отк} = t\_{об}/(t\_{ож}+t\_{об})

A = λQ = 1/(t\_{ож}+t\_{об})

## 11. Многоканальная СМО с отказами. Расчет показателей эффективности.

Показатели эффективности:

А — абсолютная пропускная способность СМО, т. е. среднее число заявок, обслуживаемых в единицу времени;

Q — относительная пропускная способность СМО, т. е. средняя доля пришедших заявок, обслуживаемых системой;

Р\_{отк} — вероятность отказа, т. е. вероятность того, что заявка покинет СМО не обслуженной.

k — среднее число занятых каналов

λ — интенсивность потока заявок, т. е. среднее число поступающих заявок в единицу времени;

μ — интенсивность потока обслуживаний.

Будем рассматривать систему, работающую без очереди, т. е. с отказами в обслуживании заявкам, пришедшим в тот момент, когда все каналы системы заняты обслуживанием ранее пришедших заявок.

Состояния системы будем нумеровать по числу заявок, находящихся на обслуживании в системе, т. е. Sk — состояние системы, когда в ней находится k заявок (занято k каналов).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | λ |  | λ |  | λ |  | λ |  | λ |  | λ |  |
| S0 | ⇄ | S1 | ⇄ | S2 | ⇄ | … | ⇄ | Sk | ⇄ | … | ⇄ | Sn |
|  | μ |  | 2μ |  | 3μ |  | kμ |  | (k+1)μ |  | nμ |  |

Предельные вероятности состояний системы существуют, т. к. система имеет конечное число состояний и из каждого из них можно за конечное число шагов перейти в любое другое состояние. Поскольку система работает по схеме гибели и размножения, найдем предельные.

Интенсивностью нагрузки каналов(трафик):

Для нахождения остальных вероятностей

=

Среднее число занятых каналов можно найти как отношение абсолютной пропускной способности к интенсивности потока обслуживаний, т. к. каждый занятый канал в единицу времени может обслужить в среднем μ заявок

В качестве дополнительных показателей эффективности работы СМО можно рассмотреть среднее число свободных каналов,

Коэффициент загрузки (использования, занятости) каналов

Коэффициент загрузки каналов определяет долю среднего числа занятых каналов в общем числе имеющихся каналов. Можно определить и коэффициент простоя каналов.

## 12. Одноканальная СМО с неограниченной очередью. Формулы Литтла.

Рассматриваются системы, в которых для заявок предусмотрена возможность ожидания обслуживания, находясь в организованной очереди. В дальнейшем будем предполагать, что дисциплина очереди простейшая, т. е. заявки обслуживаются в порядке их последовательного поступления в очередь.

Р\_{отк} — вероятность отказа заявке в обслуживании;

Q — относительная пропускная способность системы;

A — абсолютная пропускная способность системы;

k — среднее число занятых каналов;

k\_{заг} — коэффициент загрузки каналов.

Р\_{зан} — вероятность того, что канал занят обслуживанием заявки;

Р\_{оч} — вероятность наличия очереди в системе;

L\_{сист} — среднее число заявок, находящихся в системе;

T\_{сист} — среднее время пребывания заявки в системе;

L\_{оч} — среднее число заявок, находящихся в очереди;

T\_{оч} — среднее время пребывания заявки в очереди;

L\_{об} — среднее число заявок, находящихся под обслуживанием;

T\_{об} — среднее время обслуживания заявки в системе по отношению ко всем заявкам, поступившим в систему в единицу времени.

Состояния системы будем нумеровать по суммарному числу занятых каналов и заявок, находящихся в очереди, т. е. по числу заявок, находящихся в системе:

S0 — система свободна, т. е. канал не занят, очереди нет;

S1 — канал занят, обслуживается одна заявка, очереди нет;

S2 — канал занят, обслуживается одна заявка, другая заявка стоит в очереди;

S3 — канал занят, две заявки стоят в очереди; …;

Sk — канал занят, (k — 1) заявок стоит в очереди; и т. д.

Поскольку очередь не ограничена по длине, система имеет бесконечное множество состояний.

Интенсивность потока заявок (λ) не зависит ни от состояния канала обслуживания (свободен, занят), ни от числа заявок, находящихся в очереди, поэтому переход от любого состояния Sk к состоянию Sk+1 , который происходит под влиянием потока заявок, совершается с постоянной интенсивностью, равной интенсивности потока заявок (с интенсивностью λ).

Переход из любого состояния Sk в состояние Sk−1 происходит с постоянной интенсивностью, равной интенсивности потока обслуживаний, т. е. с интенсивностью μ.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | λ |  | λ |  | λ |  | λ |  | λ |  |
| S0 | ⇄ | S1 | ⇄ | S2 | ⇄ | … | ⇄ | Sk | ⇄ | … |
|  | μ |  | μ |  | μ |  | μ |  | μ |  |

Система работает по схеме процесса гибели и размножения. Полагая, что предельные вероятности состояний системы существуют, воспользуемся формулами этой схемы. Запишем выражение для p0

По формуле суммы бесконечно убывающей геометрической прогрессии имеем

= 1/(1-)

Теорема. Если 1, то предельные вероятности состояний данной СМО существуют.

Это означает, что для стабильной работы системы без отказа заявкам в обслуживании необходимо, чтобы интенсивность потока заявок была меньше интенсивности потока обслуживаний.

Если же , т. е. интенсивность потока заявок превышает (в крайнем случае равна) интенсивность потока обслуживаний, то система не успевает справляться с заявками, канал становиться постоянно загруженным (вероятность его свободного состояния p0 уже при ρ = 1 становится равной нулю), число заявок в очереди быстро стремится к бесконечности и система за конечное время не может обслужить все пришедшие и стоящие в очереди заявки.

Итак, при 1 существуют предельные вероятности состояний системы:

Найдем показатели эффективности работы системы. Для СМО с неограниченной очередью при 1 любая заявка, пришедшая в систему, будет обслужена.

Канал будет занят при всех состояниях системы, кроме S0 , следовательно, вероятность того, что канал занят, равна

а среднее число занятых каналов, равное в данном случае коэффициенту загрузки канала (n = 1), равно

Вероятность наличия очереди

Вероятность того, что пришедшая заявка будет обслужена сразу (без ожидания в очереди) равна вероятности того, что система в это момент свободна:

Вероятность того, что пришедшая заявка попадет в очередь, равна

Среднее число заявок в системе равно

Среднее число заявок, находящихся под обслуживанием в системе

Найдем теперь среднее число заявок, находящихся в очереди. Эту величину можно найти как разность между средним числом заявок в системе и средним числом заявок, находящихся под обслуживанием

Теорема. Для любой СМО, при любом характере потока заявок, при любом распределении времени обслуживания, любой дисциплине обслуживания имеют место формулы. Эти формулы называются формулами Литтла.

Среднее время пребывания заявки в системе

Среднее время пребывания заявки в очереди

Среднее время пребывания заявки под обслуживанием, отнесенное к среднему общему количеству заявок, поступающих в систему в единицу времени

В силу того, что в данной СМО все поступающие заявки будут обслужены (вероятность отказа в обслуживании равна нулю)

## 13. Многоканальная СМО с неограниченной очередью.

Рассматриваются системы, в которых для заявок предусмотрена возможность ожидания обслуживания, находясь в организованной очереди. В дальнейшем будем предполагать, что дисциплина очереди простейшая, т. е. заявки обслуживаются в порядке их последовательного поступления в очередь.

Р\_{отк} — вероятность отказа заявке в обслуживании;

Q — относительная пропускная способность системы;

A — абсолютная пропускная способность системы;

k — среднее число занятых каналов;

k\_{заг} — коэффициент загрузки каналов.

Р\_{зан} — вероятность того, что канал занят обслуживанием заявки;

Р\_{оч} — вероятность наличия очереди в системе;

L\_{сист} — среднее число заявок, находящихся в системе;

T\_{сист} — среднее время пребывания заявки в системе;

L\_{оч} — среднее число заявок, находящихся в очереди;

T\_{оч} — среднее время пребывания заявки в очереди;

L\_{об} — среднее число заявок, находящихся под обслуживанием;

T\_{об} — среднее время обслуживания заявки в системе по отношению ко всем заявкам, поступившим в систему в единицу времени.

Рассмотрим систему, которая имеет n каналов обслуживания и возможность накопления заявок в очереди. На очередь не наложено никаких ограничений: ни на длину очереди, ни на время пребывания заявки в очереди.

Состояния системы будем нумеровать по числу занятых каналов и числу заявок, находящихся в очереди, т. е. по числу заявок, находящихся в системе:

S\_{0} — система свободна, т. е. каналы не заняты, очереди нет;

S\_{1} — один канал занят, обслуживается одна заявка, очереди нет;

S\_{2} — заняты два канала, обслуживаются две заявки, очереди нет;

S\_{3} — три канала заняты, система обслуживает три заявки, очереди нет; …;

S\_{n} — все n каналов заняты обслуживанием заявок, очереди нет;

S\_{n+1}– все n каналов заняты обслуживанием, в очереди стоит одна заявка; …;

S\_{n+r} — n каналов заняты обслуживанием, в очереди стоят r заявок; и т. д

Как только один из n каналов освободится, ближайшая заявка из очереди поступает в освободившийся канал на обслуживание и очередь уменьшается на одну заявку, т. е. система переходит из состояния S\_{n+r} в состояние S\_{n+r-1} . Следовательно, переход из любого состояния S\_{n+r} в состояние S\_{n+r-1} происходит с постоянной интенсивностью, равной максимально возможной суммарной интенсивности потока обслуживаний си стемы, т. е. с интенсивностью nμ, независимо от числа r заявок, находящихся в очереди.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | λ |  | λ |  | λ |  | λ |  | λ |  | λ |  | λ |  |
| S0 | ⇄ | S1 | ⇄ | S2 | ⇄ | … | ⇄ | Sn | ⇄ | … | ⇄ | Sn+1 | ⇄ | … |
|  | μ |  | 2μ |  | 3μ |  | nμ |  | nμ |  | nμ |  | nμ |  |

Система работает по схеме процесса гибели и размножения. Полагая, что предельные вероятности состояний системы существуют, воспользуемся формулами этой схемы. Запишем выражение для p0

Теорема. Если ρ/n < 1, то предельные вероятности состояний рассматриваемой системы СМО существуют.

= (

Вероятность наличия в системе очереди из заявок, т. е. вероятность того, что хотя бы одна заявка стоит в очереди на обслуживание, равна сумме вероятностей всех состояний системы, начиная с S\_{n+1} .

Ни одна из пришедших в систему заявок не получит отказа в обслуживании.

Откуда следует, что относительная пропускная способность системы равна

Абсолютная пропускная способность

Среднее число занятых каналов найдем как отношение абсолютной пропускной способности системы к интенсивности потока обслуживания. В данном случае оно равно приведенной интенсивности потока заявок (трафику).

Можно найти среднее число свободных каналов

Коэффициент загрузки каналов, соответственно, равен отношению среднего числа занятых каналов к общему числу каналов системы

Коэффициент загрузки каналов отражает среднюю долю времени, в течение которого каждый канал бывает занят обслуживанием заявки. Следовательно, он равен вероятности того, что произвольно выбранный канал окажется занят.

Найдем среднее число заявок, ожидающих обслуживания в очереди (среднюю длину очереди), как математическое ожидание случайной величины.

Среднее число заявок, находящихся под обслуживанием, очевидно, равно среднему числу занятых каналов

Тогда среднее число заявок в системе найдем как сумму среднего числа заявок под обслуживанием и среднего числа заявок в очереди

По формуле Литтла найдем среднее время пребывания заявки под обслуживанием

Формулы Литтла позволят найти также среднее время пребывания заявки в очереди

И в системе

## 14. Одноканальная СМО с ограниченной очередью.

Р\_{отк} — вероятность отказа заявке в обслуживании;

Q — относительная пропускная способность системы;

A — абсолютная пропускная способность системы;

k — среднее число занятых каналов;

k\_{заг} — коэффициент загрузки каналов.

Р\_{зан} — вероятность того, что канал занят обслуживанием заявки;

Р\_{оч} — вероятность наличия очереди в системе;

L\_{сист} — среднее число заявок, находящихся в системе;

T\_{сист} — среднее время пребывания заявки в системе;

L\_{оч} — среднее число заявок, находящихся в очереди;

T\_{оч} — среднее время пребывания заявки в очереди;

L\_{об} — среднее число заявок, находящихся под обслуживанием;

T\_{об} — среднее время обслуживания заявки в системе по отношению ко всем заявкам, поступившим в систему в единицу времени.

Рассмотрим систему с одним каналом обслуживания и очередью, допустимой, если она не превышает числа m > 0. Состояния системы будем нумеровать по числу заявок, находящихся в системе, т. е. Sk — состояние системы, когда в ней находится k заявок

S1 — канал занят, обслуживается одна заявка, очереди нет;

S2 — канал занят, обслуживается одна заявка, одна заявка стоит в очереди;

S3 — канал занят, две заявки стоят в очереди; …;

Sm+1 — канал занят, m заявок стоят в очереди, т. е. все допустимые места в очереди заняты, очередной заявке, поступившей в систему, будет отказано в обслуживании. Общее число состояний системы в рассматриваемом случае равно m + 2. Потоки заявок и обслуживания предполагаются простейшие, их интенсивности равны, соответственно, λ и μ.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | λ |  | λ |  | λ |  | λ |  | λ |  | λ |  |
| S0 | ⇄ | S1 | ⇄ | S2 | ⇄ | … | ⇄ | Sk | ⇄ | … | ⇄ | Sm+1 |
|  | μ |  | μ |  | μ |  | μ |  | μ |  | μ |  |

Поток заявок в систему при любом ее состоянии сохраняет свою интенсивность, которая равна λ, следовательно, для любого состояния системы переход из состояния в состояние происходит с интенсивностью λ. Интенсивность потока обслуживаний равна μ, это интенсивность освобождения одного работающего в системе канала, следовательно, переход из любого состояния в состояние происходит с интенсивностью μ. Система работает по схеме гибели и размножения и имеет конечное число состояний. Из каждого состояния за конечное число шагов можно перейти в любое другое состояние. Значит, предельные вероятности всех состояний системы существуют. Составим выражение для вычисления предельной вероятности состояния :

, где

При ρ ≠ 1 в скобках этого выражения стоит сумма членов геометрической прогрессии, но только теперь конечного числа ее членов. Эту сумму найдем по формуле

где - первый член прогрессии, - ee заменитель, - число членов суммы.

В нашем случае

Отсюда,

Значит, при ρ ≠ 1

Таким образом, существует и равно полученному выражению при или при . Случай разберем особо. Из этого следует, что при ρ ≠ 1 предельные вероятности остальных состояний системы соответственно равны

, …,

Найдем показатели эффективности для рассматриваемого случая. Вероятность отказа заявке в обслуживании равна вероятности того, что система находится в состоянии :

Относительная пропускная способность системы:

Абсолютная пропускная способность системы:

Вероятность наличия очереди найдем как сумму вероятностей состояний системы, начиная с состояния и заканчивая состоянием :

Здесь мы снова воспользовались формулой суммы конечного числа членов геометрической прогрессии с первым членом, равным ρ2 , знаменателем, равным ρ, и числом членов, равным m.

Среднее число заявок, находящихся под обслуживанием, найдем как математическое ожидание случайной величины, принимающей значение, равное нулю, если система свободна от заявок, и равное единице при всех остальных состояниях системы.

Среднее число заявок, стоящих в очереди, также найдем как математическое ожидание случайной величины, принимающей возможные значения этого числа, т. е. через сумму произведений значений этой случайной величины на соответствующие им вероятности состояний системы. При этом воспользуемся тем обстоятельством, что производная суммы конечного числа дифференцируемых слагаемых равна сумме их производных.

Среднее число заявок в системе найдем как сумму среднего числа заявок под обслуживанием и среднего числа заявок в очереди:

Среднее время пребывания заявки в очереди, под обслуживанием и в системе найдем по формулам Литтла:

,

В результате получены формулы, характеризующие работу системы при условии .

Предельные вероятности состояний системы:

, …,

Вероятность отказа заявке в обслуживании:

Вероятность наличия очереди:

Относительная пропускная способность системы:

Абсолютная пропускная способность системы:

Среднее число заявок, находящихся под обслуживанием, или средняя загрузка канала:

=

Среднее число заявок, стоящих в очереди:

Среднее число заявок в системе:

Теперь рассмотрим случай, когда и, следовательно, λ = μ. В этом случае в формуле для предельной вероятности свободного состояния системы все слагаемые в скобках равны единице:

Следовательно, в скобках стоит сумма не геометрической, а ариф‑ метической прогрессии с конечным числом членов и (в данном случае) с разностью d = 0. Число единиц в скобках равно числу состояний системы, т. е. равно m + 2. Таким образом, для этого случая

Переходя к нахождению предельных вероятностей остальных состояний системы, видим, что вероятности всех без исключения состояний системы равны в этом случае друг другу и равны вероятности свободного состояния системы:

Очевидно, что сумма всех этих вероятностей (т.к. их число равно m + 2) равна единице.

Вероятность отказа заявке в обслуживании равна вероятности состояния системы, когда все места в очереди заняты:

Вероятность наличия очереди равна сумме вероятностей состояний системы от до , таких состояний у системы m:

Относительная пропускная способность системы:

Абсолютная пропускная способность системы:

)

Среднее число заявок, находящихся под обслуживанием, или среднюю загрузку канала снова найдем по формуле:

=

Найдем среднее число заявок, стоящих в очереди:

В скобках сумма m членов арифметической прогрессии с разностью d = 1.

Формула для суммы n членов арифметической прогрессии:

где — первый и — последний члены прогрессии.

В нашем случае

Значит,

Среднее число заявок в системе:

Итак,

Среднее время пребывания заявки под обслуживанием, в очереди и в системе можно получить в следующем виде:

т.к. мы рассматриваем случай, когда λ = μ, то в данном случае

Среднее время обработки заявки в системе , рассчитанное по от‑ ношению ко всему потоку заявок (т. е. с учетом как обслуженных заявок, так и тех, которым было отказано в обслуживании), не совпадает со средним временем обработки одной заявки при непрерывном потоке обслуживания. Оно зависит от числа m — имеющихся мест в очереди — и меньше, чем , т.к. общее время, затраченное на обслуживание заявок, делится на общее количество пришедших в систему заявок, включая и ушедшие необслуженными. Проведенный анализ показывает, что системы рассмотренного типа сохраняют работоспособность при интенсивности потока заявок, равном или большем интенсивности обслуживания, за счет возможности отказа неко‑ торой части заявок в обслуживании.

## 15. Многоканальная СМО с ограниченной очередью.

Р\_{отк} — вероятность отказа заявке в обслуживании;

Q — относительная пропускная способность системы;

A — абсолютная пропускная способность системы;

k — среднее число занятых каналов;

k\_{заг} — коэффициент загрузки каналов.

Р\_{зан} — вероятность того, что канал занят обслуживанием заявки;

Р\_{оч} — вероятность наличия очереди в системе;

L\_{сист} — среднее число заявок, находящихся в системе;

T\_{сист} — среднее время пребывания заявки в системе;

L\_{оч} — среднее число заявок, находящихся в очереди;

T\_{оч} — среднее время пребывания заявки в очереди;

L\_{об} — среднее число заявок, находящихся под обслуживанием;

T\_{об} — среднее время обслуживания заявки в системе по отношению ко всем заявкам, поступившим в систему в единицу времени.

Рассмотрим систему, которая имеет n каналов обслуживания и может обладать очередью, не превышающей m заявок (m > 0). Состояния системы будем нумеровать по числу заявок, находящихся в системе, т. е. Sk — состояние системы, когда в ней находится k заявок:

S0 — система свободна: заявок в системе нет;

S1 — занят один канал, обслуживается одна заявка, очереди нет;

S2 — два канала заняты, обслуживаются две заявки, очереди нет; …;

Sn — заняты все n каналов, очереди нет;

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

Sn+m — все каналы заняты, m заявок стоят в очереди, т. е. все допустимые места в очереди заняты, очередной поступившей в систему заявке будет отказано в обслуживании.

Общее число состояний системы в рассматриваемом случае равно n + m + 1. Потоки заявок и обслуживания предполагаются простейшие. Интенсивность входящего потока заявок равна λ, интенсивность потока обслуживаний равна μ.

Граф системы:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | λ |  | λ |  | λ |  | λ |  | λ |  | λ |
| S0 | ⇄ | S1 | ⇄ | S2 | ⇄ | … | ⇄ | Sn | ⇄ | … | Sn+m |
|  | μ |  | 2μ |  | 3μ |  | nμ |  | nμ |  | nμ |

Поток заявок в систему при любом ее состоянии сохраняет свою интенсивность λ, следовательно, для любого состояния системы при приходе очередной заявки переход из состояния Sk в состояние Sk+1 происходит с интенсивностью λ. Интенсивность потока обслуживаний равна μ — интенсивности освобождения одного канала. Следовательно, если заняты k каналов, интенсивность освобождения одного (любого) из них происходит с интенсивностью kμ, переход из состояния Sk в состояние Sk−1 происходит с интенсивностью kμ. Если заняты все n каналов, то интенсивность освобождения одного из них равна nμ. Это максимально возможная интенсивность освобождения каналов, она не зависит от числа заявок, находящихся в очереди. Поэтому переход из состояния Sn+k в состояние Sn+k­-1 происходит с постоянной интенсивностью, равной nμ, независимо от числа k заявок, ожидающих обслуживания в очереди (1≤ k ≤ m).

Система работает по схеме гибели и размножения и имеет конечное число состояний. Из каждого состояния за конечное число шагов можно перейти в любое другое состояние. Значит, предельные вероятности всех состояний системы существуют. Составим выражение для вычисления предельной вероятности состояния S0:

Сумма последних m слагаемых, стоящих в скобках этого выражения, является при (ρ/n ≠ 1) суммой конечного числа членов геометрической прогрессии с первым членом прогрессии , знаменателем и числом членов k = m. Эту сумму найдем по формуле:

*,* получим

Значит, при ρ/n ≠ 1

Из этого следует, что при ρ/n ≠ 1 предельные вероятности остальных состояний системы соответственно равны:

Найдем показатели эффективности для рассматриваемого случая.

Вероятность отказа заявке в обслуживании равна вероятности того, что система находится в состоянии :

Относительная пропускная способность системы:

Абсолютная пропускная способность системы:

Вероятность наличия очереди найдем как сумму вероятностей состояний системы, начиная с состояния и заканчивая состоянием

Здесь воспользовались ранее найденной суммой конечного числа членов геометрической прогрессии.

Среднее число заявок, находящихся под обслуживанием, найдем как среднее число занятых каналов, т. е. как среднее число обслуживаемых заявок за среднее время обслуживания одной заявки:

Среднее число заявок, стоящих в очереди, найдем как математическое ожидание случайной величины, принимающей возможные значения этого числа заявок, т. е. через сумму произведений значений этой случайной величины на соответствующие им вероятности состояний системы. Суммируем по состояниям системы, соответствующим наличию очереди, т. е. от состояния до состояния . При этом опять воспользуемся тем обстоятельством, что производная суммы конечного числа дифференцируемых слагаемых равна сумме их производных:

Итак, среднее число заявок, находящихся в очереди, можно найти по формулам:

Среднее число заявок в системе найдем как сумму среднего числа заявок под обслуживанием и среднего числа заявок в очереди:

Среднее время пребывания заявки в очереди, под обслуживанием и в системе найдем по формулам Литтла:

Коэффициент загрузки каналов найдем по формулам:

В результате получены формулы, характеризующие работу системы при условии

Теперь рассмотрим случай, когда и, следовательно,

В этом случае в формуле для предельной вероятности p0 (в скобках) последние m слагаемые, соответствующие наличию очереди, равны друг другу:

Значит, они образуют сумму не геометрической, а арифметической прогрессии с разностью d = 0. Число этих слагаемых равно числу со‑ стояний системы, при которых в системе есть очередь, т. е. равно m. Следовательно, для этого случая

Предельные вероятности остальных состояний системы:

Можно доказать, что сумма всех этих вероятностей, включая p0, равна единице. Вероятность отказа заявке в обслуживании равна вероятности состояния системы, когда все места в очереди заняты:

Вероятность наличия очереди равна сумме вероятностей состояний системы от до состояния . Таких состояний у системы — (m), вероятность каждого из них равна pn,

Относительная пропускная способность системы:

Абсолютная пропускная способность системы

Среднее число заявок, находящихся под обслуживанием, или среднее число занятых каналов, найдем как отношение абсолютной пропускной способности к интенсивности потока обслуживания:

Коэффициент загрузки каналов:

Найдем среднее число заявок, стоящих в очереди:

В скобках сумма m членов арифметической прогрессии с разностью d = 1. Формула для суммы n членов арифметической прогрессии:

где а1 — первый и an — последний члены прогрессии.

В нашем случае

Значит,

Среднее число заявок в системе:

Среднее время пребывания заявки под обслуживанием:

Среднее время пребывания заявки в очереди:

Среднее время пребывания заявки в системе:

Мы рассматривали случай, когда λ = nμ. В данном случае

т. е. в этом случае tож — среднее время ожидания очередной заявки — в n раз меньше tоб — среднего времени обслуживания одной заявки. Иными словами, за среднее время обслуживания одной заявки в этом случае в систему приходит в среднем n новых заявок.