теория массового обслуживания

При разработке аналитических моделей информационных систем, как правило, используют методы теории массового обслуживания (ТМО), учитывающие вероятностный характер информационных процессов, протекающих в этих системах.

Предметом изучения ТМО являются системы массового обслуживания (СМО) и стохастические сети массового обслуживания (СеМО).

теория массового обслуживания

Случайный процесс (вероятностный, стохастический) – это процесс изменения во времени состояния какой-либо системы в соответствии с вероятностными закономерностями.

Процесс работы СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем. Это означает, что состояние СМО меняется скачком в случайные моменты появления каких-либо событий (например, прихода новой заявки, окончания обслуживания, сбоя в работе и т.п.).

теория массового обслуживания

Для упрощения математического анализа работы СМО предполагают, что процесс этой работы является марковским (случайным процессом без последствий), т.е. для любого момента времени *t0* вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент *t0* и не зависят от того, когда и как система пришла в это состояние.

теория массового обслуживания

СМО - система обслуживания потока заявок, поступающих на ее вход при заданном законе и порядке обслуживания этих заявок с учетом параметров структуры самой обслуживающей системы.

СеМО - совокупность взаимосвязанных СМО. Структуру СеМО можно представить в виде графа, вершины которого соответствуют отдельным СМО, а дуги - вероятностям переходов заявок между этими СМО.

Модели СМО и СеМО удобны как для анализа сложных систем и организационных структур, так и для анализа отдельных подсистем, входящих в их состав.

теория массового обслуживания

Классификация СМО проводится по следующим основным признакам:

По числу каналов обслуживания:

- одноканальные;
- многоканальные.

По характеру поступления заявок:

- с отказами;
- с ожиданием (очередью).

По дисциплине обслуживания:

- с приоритетом (абсолютным или относительным);
- без приоритета.

По организации очереди:

- с ограниченной очередью;
- с неограниченной очередью.

По времени ожидания заявок в очереди:

- с ограниченным временем ожидания;
- с неограниченным временем ожидания.

Компоненты СМО

1. Источники заявок, определяющие входящие в СМО потоки заявок.

   Количество источников заявок определяет количество входящих в СМО потоков заявок. В зависимости от характера источника заявок различают источники с бесконечным числом и с конечным числом заявок.

В первом случае источник генерирует неограниченное число заявок в соответствии с заданной функцией распределения интервала времени между поступающими заявками.

Во втором случае в системе циркулирует конечное (м.б. постоянное) число заявок.

СМО с бесконечным числом заявок называются *разомкнутыми*, а с конечным числом заявок - *замкнутыми*.

Компоненты СМО

2. Входящий поток заявок - это совокупность всех типов заявок, поступающих на вход СМО. Обычно его задают функцией распределения интервалов времени между моментами поступления двух соседних заявок для каждого потока, т.е. средним значением интервала времени между заявками, и коэффициентом вариации этого времени.

Одной из важнейших характеристик входящего потока является его интенсивность, равная среднему числу заявок поступающих в единицу времени.

Величина, обратная интенсивности определяет средний интервал времени между двумя последовательными заявками.

Компоненты СМО

Наиболее часто в качестве входящего потока при анализе СМО используют простейший поток, который обладает следующими свойствами:

- стационарность, когда вероятность поступления заявок в систему в интервале [*t, t+h*] зависит лишь от величины *h* ;
- ординарность, когда на бесконечно малом промежутке времени *h* поступает не более одной заявки;
- отсутствие последействия, когда вероятность поступления заявок в систему в интервале [*t, t+h]* не зависит от количества заявок поступивших в систему до момента времени *t*.

Для простейшего потока интервалы времени между двумя последовательными заявками – это непрерывные случайные величины

с экспоненциальной функцией распределения $F(\tau)=1-e^{-\lambda\tau}$

и плотностью распределения $f(\tau)=\lambda e^{-\lambda\tau}$

Компоненты СМО

3. Механизм обслуживания заявок, который включает:

- длительность обслуживания заявок (среднее время и коэффициент вариации времени обслуживания заявок).

На практике наиболее часто используют экспоненциальное распределение длительности обслуживания заявок в обслуживающем аппарате (ОА), которое описывается функцией распределения времени обслуживания

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

и плотностью распределения

$$f(\tau) = \mu e^{-\mu\tau}$$

где μ - интенсивность обслуживания заявок в СМО в единицу времени.

Среднее время обслуживания заявки *M(t)* и дисперсию времени обслуживания заявки *D(t)* определяют по формуле:

$$M(\tau) = 1/\mu$$

$$D(\tau) = 1/\mu^2$$

Компоненты СМО

- количество обслуживающих аппаратов (ОА) в обслуживающей системе равное *с*, которые могут одновременно обслуживать поступающие заявки.

По количеству ОА в обслуживающей системе различают: одноканальные СМО, если в системе один ОА и многоканальные СМО, если в системе несколько ОА.

- пропускная способность обслуживающей системы - $\mu_{сис}$ которая зависит от числа обслуживающих аппаратов *с* и средней интенсивности обслуживания заявок μ_i каждым ОА.

$$\mu_{сис} = \sum_{i=1}^{c} \mu_i$$

- Количество мест в очереди на обслуживание, т.е. число заявок которое может находиться в очереди и ожидать начала обслуживания, если все ОА заняты обслуживанием.

В зависимости от числа мест в очереди различают СМО *с отказами* и СМО *без отказов*.

Компоненты СМО

- дисциплина *формирования* очереди. По правилу формирования очереди различают СМО с общей очередью и СМО с несколькими очередями. При общей очереди заявки заполняют очередь в порядке поступления. В случае наличия нескольких очередей, как правило, очереди неоднородны по значимости. Более важные заявки поступают в более приоритетные для обслуживания очереди;

- дисциплина *обслуживания* очереди определяет правила выбора заявок из очереди на обслуживание. Имеют место бесприоритетные и приоритетные дисциплины обслуживания.

В общем случае возможны различные комбинации дисциплин формирования и обслуживания очереди, что порождает достаточно широкий спектр вариантов управления процессами обработки заявок в СМО.

Описания СМО

Описания СМО унифицированы и имеют следующий символьный вид: A/ S/ c/ K/ N/ D

где символы соответствуют конкретным и наиболее важным элементам представления процессов обслуживания заявок в СМО и интерпретируются следующим образом:

A - вид распределения интервалов времени между моментами поступления входящих в СМО заявок (arrival);

S - вид распределения времени обслуживания заявок в обслуживающем аппарате СМО (service);

c - количество обслуживающих аппаратов в СМО (channels);

K - максимальное количество заявок, которое может одновременно находиться в очереди СМО (capacity);

D - дисциплина обслуживания заявок из очереди;

N - емкость источника, генерирующего заявки на обслуживание.

Нотация Кендалла (Kendall)

Описания СМО

Для конкретизации символов A и S приняты следующие обозначения:

M - пуассоновское распределение моментов поступления входящего потока заявок на обслуживание или экспоненциальное распределение интервалов времени обслуживания заявок (марковский поток);

D - фиксированный (детерминированный) интервал времени между моментами последовательных поступлений заявок в систему на обслуживание или детерминированная продолжительность обслуживания;

E*k* - распределение Эрланга интервалов времени между моментами последовательных поступлений заявок в систему или продолжительностей обслуживания, при этом К -параметр распределения Эрланга;

GI - распределение произвольного вида моментов поступления/ продолжительностей обслуживания/ заявок в системе;

PH - распределение произвольного вида заявок системы (phase-type distribution).

Описания СМО

Для конкретизации символа c указывается число от 1 до ∞;

Для конкретизации символа K указывается число от 1 до ∞;

Для конкретизации символа D приняты следующие обозначения:
FIFO/FCFS -- LIFO/LCFS – SIRO – PPQ/ NPQ / WFQ -- PS
(ПППО; АБС ; ОТН; КОМ; ДБС; ЗАН)

Для конкретизации символа N используют число 1,2 .N. ∞ которое указывает емкость источника заявок.

Описания СМО

Написать обозначение для СМО с пуассоновским входящим потоком, экспоненциальным распределением времени обслуживания, с одним ОА, с бесконечной емкостью буфера, с дисциплиной выбора заявок /первый пришел - первым обслужен/ и с бесконечной емкостью источника заявок.

Рассматриваемая СМО имеет следующее обозначение:

M / M / 1 / ∞ / ∞ / ПППО

Следует иметь в виду, что СМО, в описании которых есть такие последние три символа, считают базовыми и эти последние три символа обычно опускают, а для обозначения таких СМО используют краткую форму записи, содержащую только первые три символа. Поэтому краткая форма обозначения рассматриваемой СМО

M / M / 1

оценки эффективности функционирования

Показатели оценки эффективности функционирования СМО – количественные показатели, характеризующие уровень выполнения СМО возложенных на нее функций по обслуживанию заявок при определенном наборе рабочих параметров СМО. К основным показателям оценки качества обслуживания заявок СМО относят следующие показатели:

- Вероятность обслуживания заявок $P_{обс}$ - вероятность того, что произвольно выбранная из входящего потока с интенсивностью λ заявка будет обслужена, т.е. окажется в выходящем потоке обслуженных заявок с интенсивностью $\lambda_{вых}$

Обычно эту вероятность обслуживания заявок называют также относительной пропускной способностью СМО, обозначают символом q и определяют так:

$$P_{обс} = q = \lambda_{вых} / \lambda$$

оценки эффективности функционирования

- вероятность потери заявок $P_{отк}$ - вероятность того, что произвольно выбранная из входящего потока с интенсивностью λ заявка окажется в потоке заявок с интенсивностью $\lambda_{отк}$, которым будет отказано в обслуживании: $P_{отк} = 1 – P_{обс}$

- среднее время пребывания заявок в СМО T - время равно среднему промежутку времени от момента поступления заявки на вход СМО до момента появления ее в выходящем потоке. Среднее время пребывания заявки в СМО равно сумме среднего времени нахождения заявки в очереди и среднего времени обслуживания в ОА: $T= W + 1 / \mu$

оценки эффективности функционирования

- среднее время ожидания заявок в очереди W. В общем случае время ожидания - случайная величина, равная сумме длительностей интервалов времени, в течение которых заявка находится в очереди, начиная с момента поступления заявки на вход СМО и кончая моментом, когда заявка последний раз покидает очередь, уходя из очереди на обслуживание, т.е. является суммой двух составляющих:

А) среднего времени ожидания начала обслуживания и среднего времени ожидания в прерванном состоянии, когда обслуживается более приоритетная заявка. Первая составляющая равна промежутку времени между моментом поступления заявки на вход СМО и моментом первого назначения заявки на обслуживание.

Б) вторая составляющая равна сумме промежутков времени между моментами поступления заявки, обслуживание которой было прервано, снова в очередь, и моментами поступление этой заявки на дообслуживание.

оценки эффективности функционирования

- Средняя длина очереди Q - среднее число заявок, находящихся в очереди. Для систем без потерь средняя длина очереди связана со средним временем ожидания следующим соотношением:

$$Q = W * \lambda$$

- Среднее число заявок в СМО L - среднее суммарное число заявок, которое находится в очереди и в обслуживающих аппаратах.

  Для систем без потерь среднее число заявок в системе связано со средним временем пребывания следующим соотношением:

$$L = T * \lambda$$ *(формула Литтла)*

оценки эффективности функционирования

- загрузка обслуживающего аппарата СМО ρ - коэффициент использования ОА, вычисляют по формуле

$$\rho = \lambda / (\mu * c)$$

- загрузка обслуживающей системы φ - среднее число занятых обслуживающих аппаратов. Если обслуживающие аппараты однородны, то φ определяется по формуле

$$\varphi = c * \rho = \lambda / \mu$$

Описания СМО

Пространство состояний СМО

M / M / 1

Число означает количество заявок в системе
λ - интенсивность поступления заявок
μ – параметр показательно распределенного времени обслуживания в ОА

т.н. процесс «размножения-гибели»

показатели эффективности СМО

M / M / 1 $\quad (\lambda < \mu)$

загрузка (коэф.использования) ОА $\rho = {}^{\lambda}/_{\mu}$

вероятность занятости (в СМО *i* заявок) $\pi_i = (1 - \rho)\rho^i$

оценка количества заявок в СМО $L = \frac{\rho}{1-\rho}$

дисперсия количества заявок в СМО $Dq = \frac{\rho}{(1-\rho)^2}$

среднее время пребывания заявки в системе $T = \frac{1}{\mu-\lambda}$

среднее время ожидания в очереди $W = \frac{\rho}{\mu-\lambda}$

показатели эффективности СМО

M / M / 2 $(\lambda < \mu)$

загрузка (коэф.использования) ОА $\rho = \lambda / (2\mu)$

вероятность простоя ОА $P_0 = (1-\rho) / (1+\rho)$

оценка количества заявок в СМО $L = \frac{2 * \rho}{1-\rho^2}$

дисперсия количества заявок в СМО $Dq = \frac{2\rho^3 (1+2\rho-\rho^3)}{(1-\rho^2)^2}$

среднее время пребывания заявки в системе $T = \frac{1}{(1-\rho^2)\mu}$

среднее время ожидания в очереди $W = \frac{\rho^2}{(1-\rho^2)\mu}$

Критерий эффективности функционирования

В качестве критерия эффективности функционирования СМО обычно выбирают один наиболее важный показатель, или несколько показателей, и в этом случае критерий является некоторой функцией свертки набора этих показателей. Критерий эффективности функционирования СМО является средством оценки соответствия СМО возложенным на нее функциям.

В качестве простого и широко используемого на практике критерия, как правило, используют критерий, который минимизирует общую стоимость обработки заявок в СМО. Его можно представить в следующем виде:

$$E_l = \min_j E_j = \min_j \left[c \cdot \rho \cdot \sum_{i=1}^{n} e_{1ij} \cdot (1/\mu_{ij}) + \sum_{i=1}^{n} e_{2ij} \cdot T_{ij} \right]$$

$$E_l = \min_j E_j = \min_j \left[\; c \cdot \rho \cdot \sum_{i=1}^{n} e_{1ij} \cdot (1/\mu_{ij}) + \sum_{i=1}^{n} e_{2ij} \cdot T_{ij} \; \right]$$

E_l - стоимость обработки заявок в СМО, имеющей вариант структурной организации l , который является наилучшим с точки зрения выбранного критерия эффективности ;

E_j - стоимость обработки заявок в СМО имеющей вариант структурной организации j ;

В - количество возможных вариантов построения СМО ;

e_{1ij} - стоимость единицы времени обслуживания заявки *i*-го типа в СМО, имеющей вариант структурной организации j ;

e_{2ij} - стоимость единицы времени пребывания заявки *i*-го типа в СМО, имеющей вариант структурной организации j ;

$1/\mu_{ij}$ - среднее время обслуживания заявки *i*-го типа в СМО, имеющей вариант структурной организации j ;

T_{ij} - среднее время пребывания заявки *i*-го типа в СМО, имеющей вариант структурной организации j ;

ρ - загрузка ОА ;

c - число ОА ;

n - количество типов заявок, поступающих на вход СМО ;

E_l - значение критерия эффективности наилучшего варианта

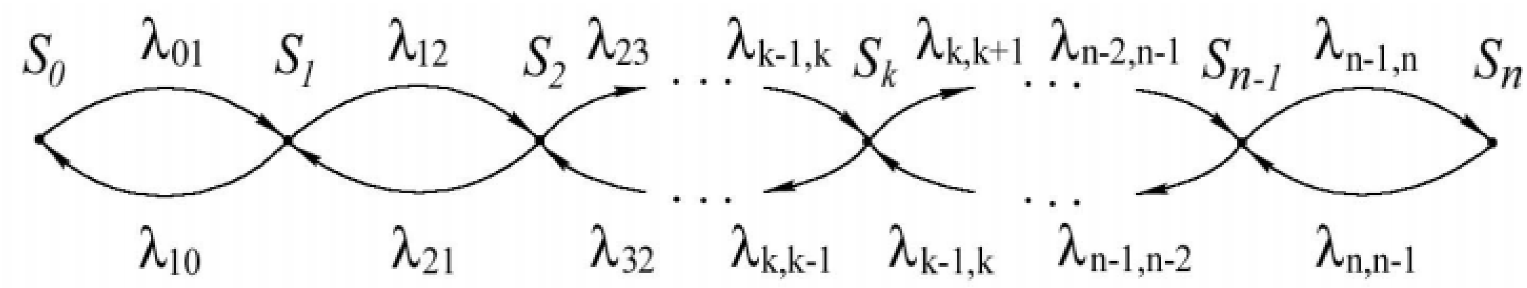
Граф состояний СМО M / M / 1

Если система находится в состоянии S1, то как только закончится обслуживание заявки, занимающей этот канал, система перейдет в состояние S0.
Этот переход будет определяться интенсивностью μ обслуживания каналов.
Если система находится в состоянии S2, то есть в системе обслуживанием занято 2 канала, то интенсивность перехода из S2 в S1 вырастает вдвое.
Очевидно, если обслуживанием занято К каналов, то интенсивность этого обслуживания будет в К раз выше, чем при обслуживании одним каналом.

Схема "размножения-гибели" встречается во многих случаях, поэтому целесообразно определить аналитические выражения предельных вероятностей состояний системы, и пользоваться готовыми формулами.

Согласно правилу «что втекает то и вытекает», записывается система уравнений для нахождения предельных вероятностей состояний процесса "размножения-гибели".

Граф состояний и переходов



Для любого К получаем

$$p_k = \frac{\lambda_{01}\ \lambda_{12}\dots\lambda_{k-2,k-1}\ \lambda_{k-1,k}}{\lambda_{k,k-1}\ \lambda_{k-1,k-2}\dots\lambda_{21}\ \lambda_{10}}\ p_0, \qquad k=\overline{1,n}.$$

Структура формулы следующая -
в числителе находится произведение всех лямбд стоящих у стрелок, направленных слева направо от S0 и до той которая идет в Sk,
а в знаменателе произведение всех лямбд, идущих в обратном направлении справа налево от Sk до S0.

Если все *pi* подставить в выражение, то можно получить выражение для *p0*

$$p_0 = \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{01}\lambda_{12}}{\lambda_{21}\lambda_{10}} + \dots + \frac{\lambda_{01}\lambda_{12}\dots\lambda_{n-1,n}}{\lambda_{n,n-1}\lambda_{n-1,n-2}\dots\lambda_{10}}\right)^{-1}$$

Формулы *pi* и *p0* дают общее решение
задачи "размножения-гибели".

Граф состояний СМО M / M / n / m

Для правильного построения графа состояний необходимо перечислinь возможные состояния СМО:

S0 – все каналы свободны, очереди нет;

S1 – занят один канал, очереди нет;

S2 – два канала заняты, очереди нет;

. . . .

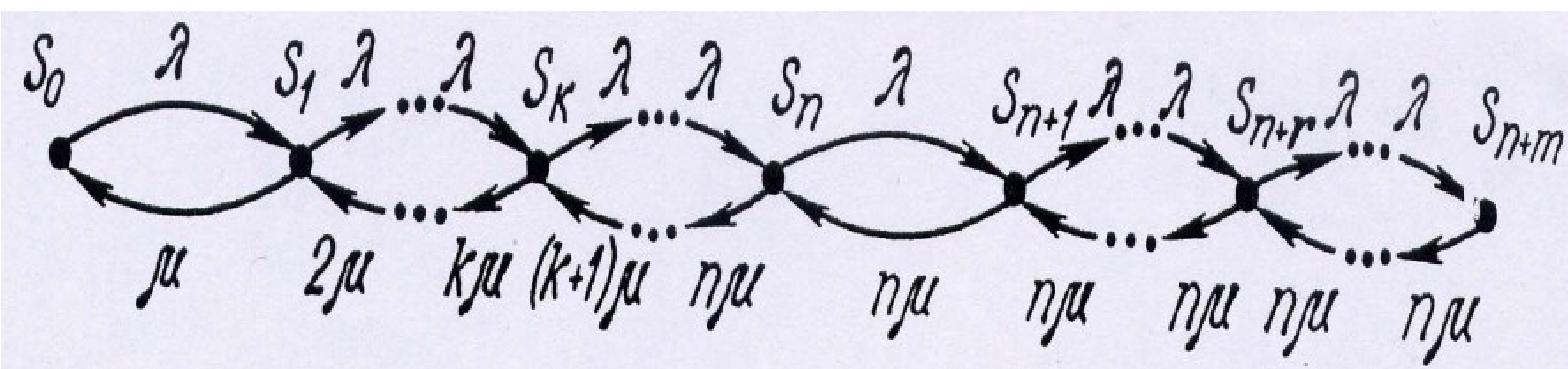
Sn – заняты все n каналов, очереди нет;

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

Sn+2 – заняты все n каналов, две заявки находится в очереди;

. . . .

Sn+m – заняты все n каналов, заняты все m мест в очереди



Граф состояний СМО M / M / n / m

Выражения расчета вероятностей состояний СМО можно записать компактно, используя $\rho = \dfrac{\lambda}{\mu}$

$$p_\alpha = \frac{\rho^k}{\alpha!} p_0, \quad \alpha = \overline{1,n},$$

$$p_{n+\beta} = \frac{\rho^{n+\beta}}{n^\beta n!} p_0, \quad \beta = \overline{1,m},$$

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

Последние *m* слагаемых в выражении являются членами геометрической прогрессии со знаменателем ρ / n поэтому

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

Для задач большой размерности использование процедуры анализа на основе формирования матрицы коэффициентов, решение СЛАУ и расчет по формулам для одной точки пространства параметров (анализа одного варианта) в многовариантных задачах анализа может оказаться затруднительным.

При N = 30 имеем систему с числом уравнений порядка 500. Формирование коэффициентов для такой системы уравнений требует порядка 10 000 операций.

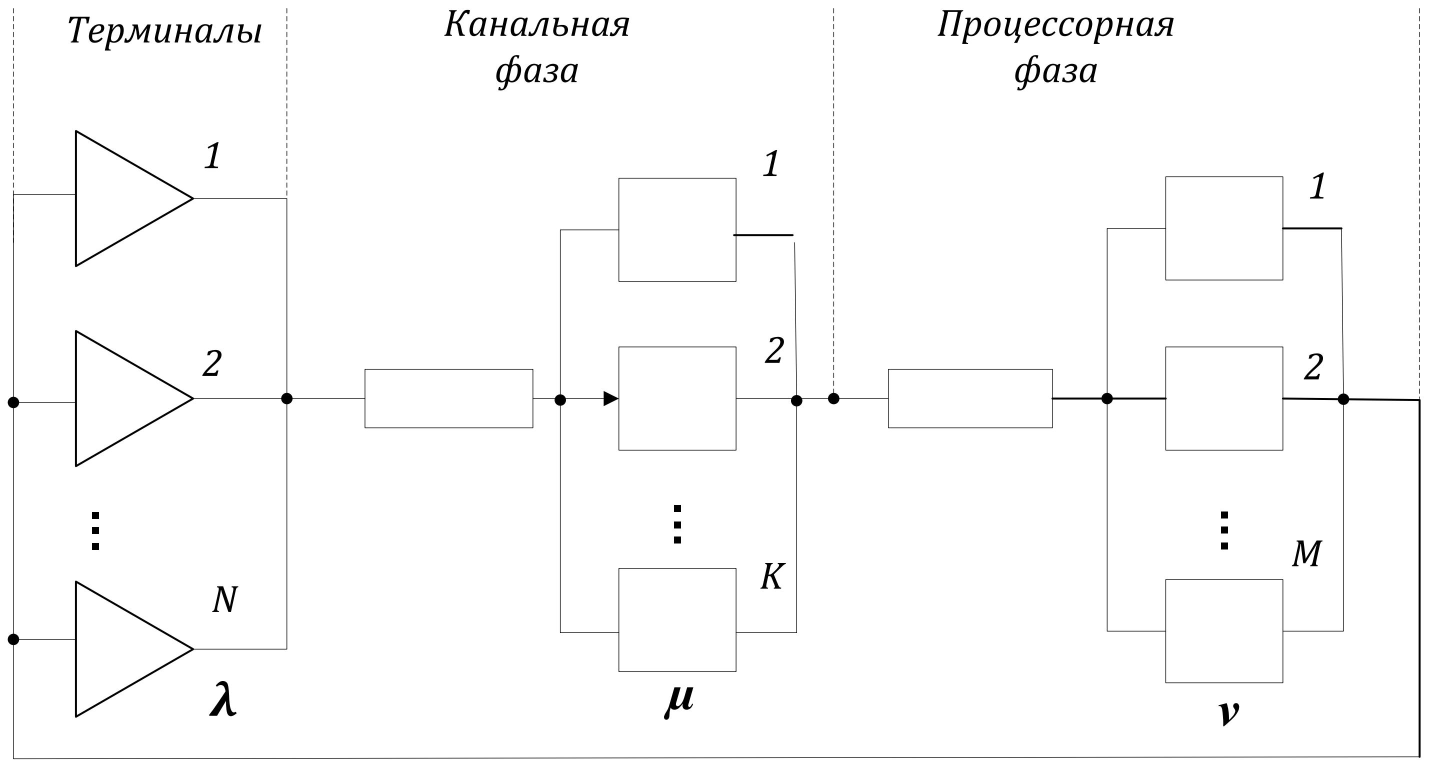
Алгоритм Гаусса для такой СЛАУ имеет вычислительную сложность порядка 10^8 операций. Для построения кривых, характеризующих зависимости выходных параметров от конструктивных параметров системы с учетом интерфейсных моделей надо проделать анализ для числа вариантов ~10 000.

Идея ***метода декомпозиции*** состоит в том, чтобы анализ сложной по структуре модели производить по частям с помощью совокупности частично укрупненных («агрегированных») моделей.

В каждой из таких моделей подробно представлена некоторая часть системы, а влияние остальных частей отражается некоторым обобщенным параметром (параметром связи).

В итоге получается система уравнений, описывающих изменение состояния каждой из агрегированных моделей, которые решают совместно, так как в качестве переменных они содержат и параметры связи.

Метод агрегирования



Например, система содержит N клиентов, K каналов связи и M серверов.

Изменение состояния системы во времени может быть описано марковским случайным процессом.

Среднее время реакции $M[t_p]$ является функцией параметров N, K, M, Tk, Tc, Ts.

Метод агрегирования

Среднее время реакции $M[t_p]$ является функцией параметров

N, K, M, Tk, Tc, Ts

Мат.ожидания функций распределения $Tk = 1/\lambda$; $Tc = 1/\mu$; $Ts = 1/v$

λ – интенсивность поступления заявок
μ – параметр времени передачи в ОА
v – параметр времени обслуживания в ОА

В качестве состояния системы в момент *t* можно взять вектор

$$z = (f(t), y(t))$$

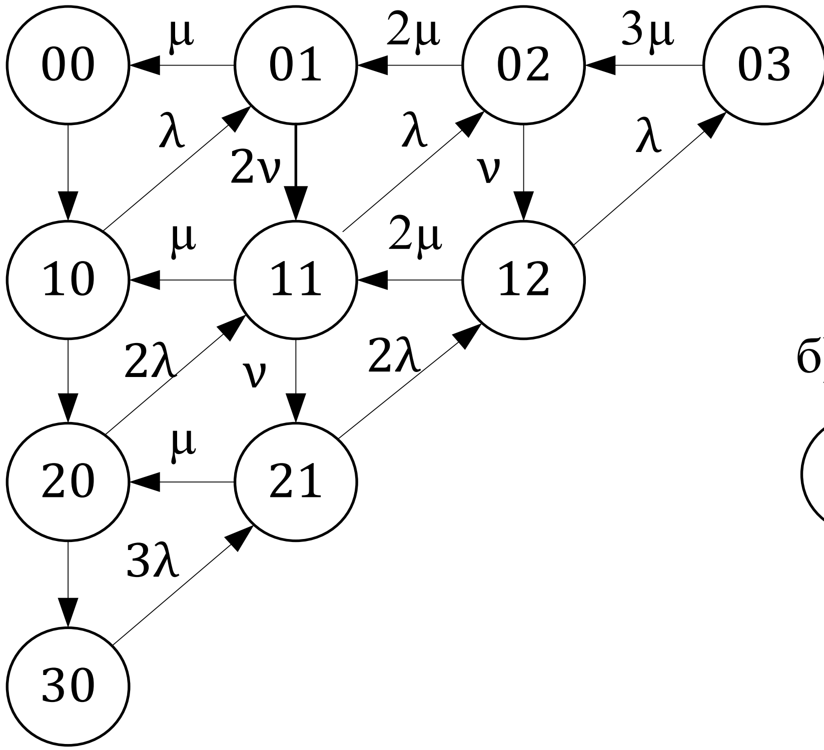
f(t) -число пользователей в момент t, осуществляющих подготовку своего запроса (число заявок в фазе клиента)

y(t) -число запросов в момент t, осуществляющих передачу (ожидающих передачу) между клиентом и сервером (число заявок в фазе каналов связи)

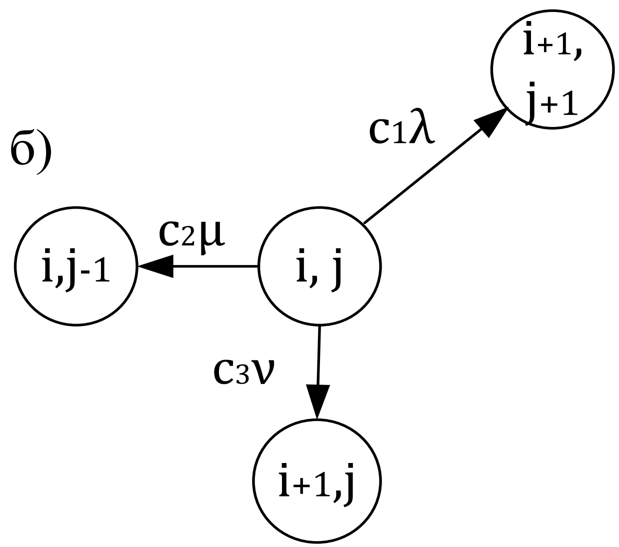
число заявок в фазе процессинга (в сервере) $N – (f(t) + y(t))$

Входящие в эту модель параметры λ, μ, v являются функциями технических характеристик серверов, каналов связи, типов запросов клиентов.

Описание СеМО



б)



Граф переходов модели для случая N = 3, K = 3, M = 2

i= f(t)   j= y(t)   c1=i   c2= min{K, j}   c3=min{M, (N –i –j)}

Метод агрегирования

Принцип эквивалентности потоков

При его использовании часть системы заменяется агрегированным узлом – обслуживающим аппаратом (ОА) с очередью. Интенсивность обслуживания в таком ОА (параметр связи) зависит от числа заявок в узле.

Принцип эквивалентности потоков состоит в том, что агрегированный узел при любом числе заявок в нем должен обеспечивать такой же поток во внешнюю сеть (т.е. такую же пропускную способность), как заменяемая им подсистема.

Этот принцип в анализе марковских стохастических моделей напоминает теорему об эквивалентном генераторе тока (теорему Нортона) в анализе электрических цепей.

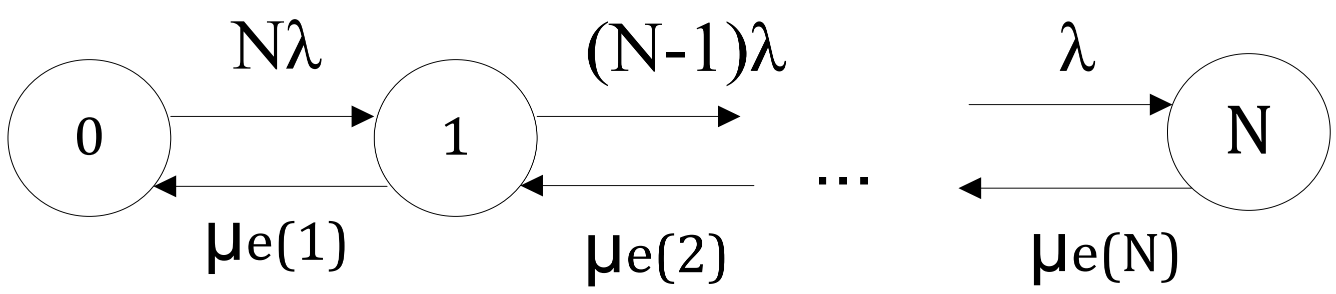
Метод агрегирования

Рассмотрим модель из двух агрегированных моделей АЭМ1 и АЭМ2.

Структура модели АЭМ1 - укрупненный один агрегированный узел

ОА с очередью, который описывается обобщенным параметром μ_{ef}.

Особенность состоит в том, что интенсивность обслуживания в агрегированном узле (параметр связи) не постоянная (μ_{ef} , зависящая от среднего количества заявок N_{cp} в узле), а вектор { $\mu_e(n)$ } каждая компонента которого зависит от количества заявок *n* в агрегированном узле



Метод агрегирования

Решение системы уравнений относительно стационарных вероятностей состояний записывается в явном виде:

$$P_0 = \left(1 + \frac{N\lambda}{\mu_e(1)} + \frac{N(N-1)\lambda^2}{\mu_e(1)\mu_e(2)} + \ldots + \frac{N!\lambda^N}{\prod_{n=1}^{N}\mu_e(n)} \right)^{-1};$$

$$P_1 = P_0 \frac{N\lambda}{\mu_e(1)};$$

$$P_2 = P_1 \frac{(N-1)\lambda}{\mu_e(2)};$$

$$\ldots$$

$$P_N = P_{N-1} \frac{\lambda}{\mu_e(N)}.$$

$$N_{cp} = \sum_{n=1}^{N} nP_n$$

$$M[t_p] = \frac{N_{cp}}{\lambda(N - N_{cp})}$$

$$\mu_{cp} = \sum_{n=1}^{N} \mu_e(n) P_n$$

Метод агрегирования

Для расчета *Nср, M[tр], $\mu_{ср}$* кроме N и λ, необходимо иметь значения компонент вектора параметра связи $\mu_e(n)$ (n=1,2,…N).

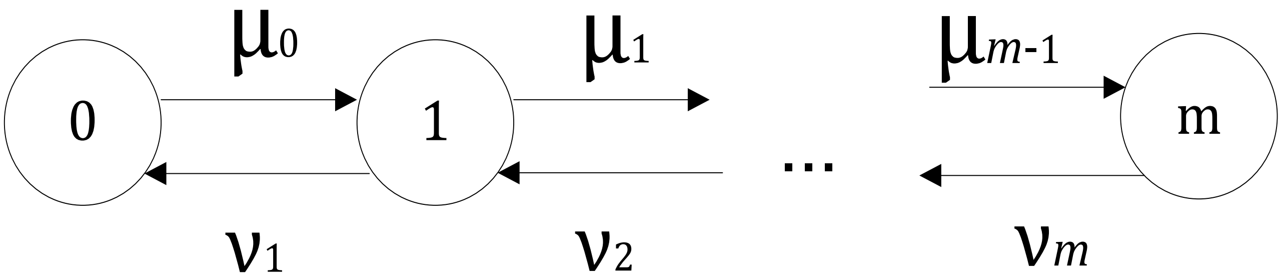
Для его нахождения используется вторая агрегированная модель АЭМ2.

В модели АЭМ2 укрупненно представлена фаза клиентов. Ее влияние отражается с помощью обобщенного параметра ***m***, представляющего собой число заявок в системе «серверы-каналы». В качестве состояния в момент ***t*** берется число заявок в серверной фазе.

Структура модели АЭМ2 совпадает со структурой типовой модели, но она рассматривается не с фиксированным в ней числом заявок $N_{ср}$, а анализируется последовательно N раз, с циркулирующим в ней числом заявок *m,* где m = 1,2,…,N.

Метод агрегирования

Граф переходов, когда в АЭМ2 циркулирует ***m*** заявок



Для каждого значения m=(1,N) рассчитываются последовательно:

$$\mu_n = \mu \min\{K, m-n\}, n = 0, 1, \ldots, m-1$$

$$\nu_n = \nu \min\{M, n\}, n = 1, 2, \ldots, m$$

Метод агрегирования

Если обозначить через π компоненту стационарного распределения вероятностей числа заявок на процессорной фазе, то решение системы уравнений относительно стационарных вероятностей запишется в виде:

$$\pi_0 = \left(1 + \frac{\mu_0}{\nu_1} + \frac{\mu_0 \mu_1}{\nu_1 \nu_2} + \ldots\right)^{-1};$$

$$\mu_e(m) = \sum_{n=0}^{m-1} \mu_n \pi_n$$

$$\pi_1 = \pi_0 \frac{\mu_0}{\nu_1}$$

$$\pi_2 = \pi_1 \frac{\mu_1}{\nu_2}$$

$$\ldots$$

$$\pi_m = \pi_{m-1} \frac{\mu_{m-1}}{\nu_m}$$

После N расчетов значений $\mu_e(m)$ по этим формулам полученные значения подставляются в формулы расчета стационарных вероятностей состояний модели АЭМ1. После этого рассчитываются выходные параметры системы $N_{ср}$; $\mu_{ср}$; $t_р$

Расчет показателей в многофазной СеМО

Согласно теореме Джексона последовательно соединенные разомкнутые СМО можно рассматривать независимо друг от друга. При этом следует помнить, что параметры выходного потока предыдущей СМО являются параметрами входного потока последующей СМО.

Поэтому при исследовании многофазной СеМО ее следует разбить на отдельные фазы (отдельные СМО) и для каждой из них последовательно, начиная с первой СМО, определять параметры входного и выходного потоков заявок.

Расчет показателей в многофазной СеМО

Исходные данные для оценки показателей функционирования многофазной СеМО:

- состав СеМО (n - количество СМО, входящих в ее состав, номера СМО в порядке их последовательного соединения);
- архитектура СеМО (последовательность соединения отдельных СМО, при этом в каждой СМО отсутствуют обратные связи);
- параметры входного потока заявок, входящих в первую СМО (λ интенсивность потока заявок, v2 квадрат коэффициента вариации интервалов времени между заявками);
- параметры потоков заявок, входящих в каждую СМО из других СеМО (интенсивности входящих потоков заявок и квадраты коэффициентов вариации интервалов времени между заявками каждого потока);
- вероятности выхода заявок в другие СеМО после их обслуживания каждой СМО в составе рассматриваемой СеМО;
- параметры обслуживания заявок в каждой СМО в составе рассматриваемой СеМО (интенсивности обслуживания заявок и квадраты коэффициентов вариации интервалов времени обслуживания)

Расчет показателей в многофазной СеМО

Этап 1. Определяют параметры входного потока заявок в первую СМО (СМО1), при необходимости используя метод композиции потоков.

Этап 2. Определяют показатели функционирования СМО1:

- загрузку обслуживающего аппарата СМО $\rho_i = \lambda_i / \mu_i$
- среднее количество заявок в очереди, используя формулу Файнберга $Q_i = \dfrac{\rho_i^2(\upsilon_{вхi}^2 + \upsilon_{oi}^2)}{2(1-\rho_i)}$
- среднее количество заявок в СМО (в очереди и на обслуживании) $L_i = Q_i + \rho_i$
- среднее время ожидания заявок в очереди $W_i = \dfrac{Q_i}{\lambda_i}$
- среднее время пребывания заявок в СМО1 $T_i = \dfrac{L_i}{\lambda_i}$
- интенсивность и квадрат коэффициента вариации интервалов времени выходного потока, поступающего на вход следующей СМО, в данном случае СМО2

$$\lambda_j = \lambda \cdot q_j$$

$$\upsilon_{выхi}^2 = \upsilon_{вхi}^2 + \rho_i^2 \cdot (\upsilon_{oi}^2 - \upsilon_{вхi}^2)$$

Расчет показателей в многофазной СеМО

Этап 3. Последовательно повторяют этап 2 и этап 3 для каждой СМО, входящей в состав рассматриваемой СеМО.

Этап 4. Определяют показатели функционирования СеМО, на основе показателей функционирования отдельных СМО, входящих в ее состав

$$Q = \sum_{i=1}^{n} Q_i$$

среднее количество заявок в очереди

$$L = \sum_{i=1}^{n} L_i$$

среднее количество заявок в СМО

$$W = \sum_{i=1}^{n} W_i$$

среднее время ожидания заявок в очереди

$$T = \sum_{i=1}^{n} T_i$$

среднее время пребывания заявок в СМО

Для СМО типа Ек/ Ек/ 1 имеем, что параметр К потока Эрланга и квадрат коэффициента вариации интервалов времени между поступающими заявками связаны соотношением

$$K = 1/v^2$$

Расчет показателей СеМО с обратными связями

Исходным данными для анализа СеМО с обратными связями являются:

- n количество СМО, входящих в состав исследуемой СеМО;
- λ интенсивность (интенсивности) потоков поступающих в СеМО;
- c_i (i=1..n) количество идентичных ОА в составе i -той СМО;
- $\mu_i = 1 / t_i$ (i=l..n) интенсивность обслуживания заявок в каждом ОА в составе *i*-ой СМО;
- P_{ij} - вероятность перехода заявки после обслуживания в *i* -ой СМО на вход *j* -ой СМО.

Порядок расчёта характеристик функционирования СеМО с обратными связями

1. Составляем систему линейных уравнений потоков заявок, поступающих на вход каждой СМО (i=l..n) в составе исследуемой СеМО. Количество уравнений равно количеству СМО в составе СеМО.

Каждое уравнение имеет вид:

$$\lambda_j = \sum_{j1}^{n} \lambda_i Pij$$

2. Проводим во всех уравнениях замену обозначения λi на обозначения

$\lambda i = \lambda\ \alpha_i$ где α_i - количество проходов заявки через *i*-ю СМО за время пребывания этой заявки в исследуемой СеМО.

Сокращаем во всех уравнениях на λ и получаем следующую систему уравнений:

$$\alpha_j = \sum_{i1}^{n} \alpha_i Pij$$

Расчет показателей СеМО с обратными связями

3. Решаем полученную систему линейных уравнений и определяем численные значение а, . (i=l..n).

4. Определяем p_i - загрузку каждого ОА СМО:

$$p_i = \lambda_i / \mu_i c_i$$

Проверяем условия наличия стационарного режима для каждой *i*-*той* СМО в составе исследуемой СеМО

$$p_i < 1 \ (i=1..n)$$

Если это условие выполняется, то проводим расчёт характеристик функционирования каждой i-той СМО в составе СеМО. При этом, согласно теореме Джексона, все СМО рассматриваются независимо друг от друга и каждая представляется в виде М/М/с

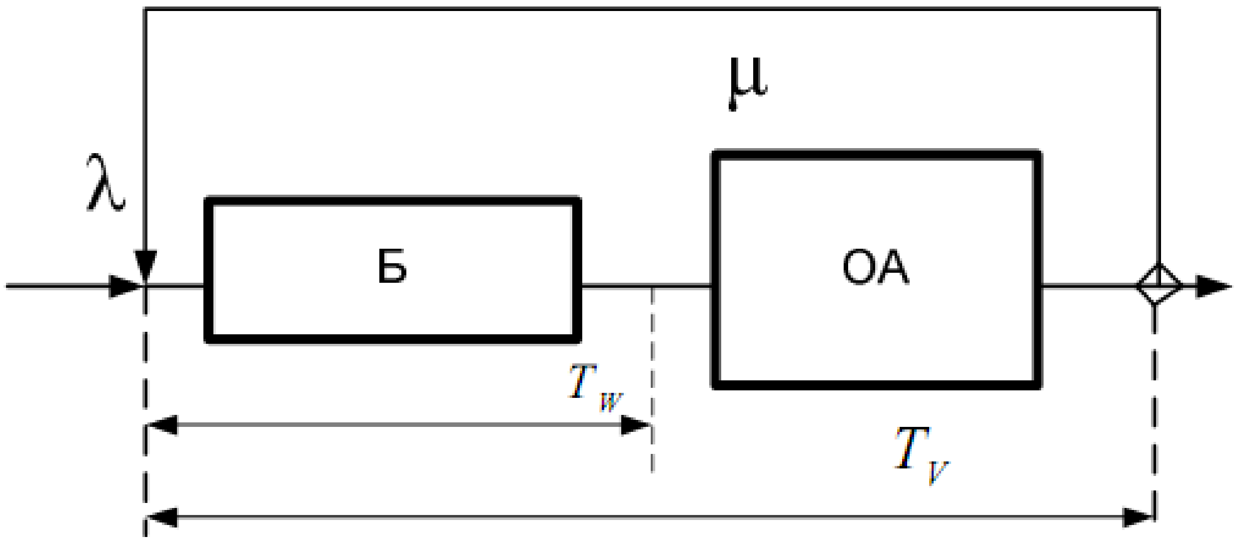
Расчет показателей СеМО с обратными связями

Если условия стационарности не выполняется, то для соответствующих СМО необходимо выполнить одно из следующих действий:

- увеличить количество ОА в составе СМО;
- увеличить производительность каждого ОА в составе СМО;
- уменьшить вероятности потоков заявок, поступающих в СМО.

5. Определяем характеристики функционирования каждой *i*-той СМО в составе СеМО.
6. Определяем характеристики функционирования СеМО

Расчет показателей СеМО с обратными связями



Пример СМО М/М/1 с обратной связью

λ = 4 заявки/сек

μ = 10 заявки/сек

p=0.5 – вероятность повторного обслуживания

Расчет показателей СеМО с обратными связями

Расчет характеристик СМО:

1. Определяем входной поток, поступающий в буфер

$$\lambda_{вх} = \lambda + \lambda_{вх} \cdot p \qquad \lambda_{вх} = \frac{\lambda}{1-p} = \frac{4}{1-0,5} = 8$$

2. Определяем количество проходов заявкой буфера и ОА за время пребывания в системе

$$\alpha = \frac{1}{1-p} = \frac{1}{1-0,5} = 2$$

3. Определяем загрузку ОА

$$\rho = \frac{\lambda_{вх}}{\mu} = \alpha \cdot \lambda \cdot t_0, = 0,8$$

Расчет показателей СеМО с обратными связями

-Определяем среднее количество заявок в очереди

$$Q = L_W = \frac{\rho^2}{1-\rho} = \frac{0,64}{1-0,8} = 3,2.$$

-Определяем количество заявок в системе

$$L = L_V = L_W + \rho \cdot \alpha = 4\,\textit{заявки}$$

-Среднее время пребывания заявок в очереди

По формуле Литтла: $L_W = \lambda_{вх} \cdot T_W$,

$$W = T_W = \frac{L_W}{\lambda_{вх}} = \frac{3,2}{8} = 0,4\ c.$$

$$W_{сист} = T_{Wсист} = T_W \cdot \alpha = 0,4 \cdot 2 = 0,8$$

проверим по формуле Колмогорова

$$T_W = \frac{\rho \cdot t_0 \cdot \alpha}{1-\rho} = \frac{0,8 \cdot 0,1 \cdot 2}{0,2} = 0,8$$

-Среднее время пребывания заявок в СМО

$$T = T_V = \frac{t_0 \cdot \alpha}{1-\rho} = \frac{0.1 \cdot 2}{1-0.8} = 1$$

Расчет показателей СМО с отказами

Показатели оценки качества функционирования СМО M/M/C/0

Загрузка СМО $\varphi = \frac{\lambda}{\mu}$

Загрузка ОА СМО $\rho = \frac{\varphi}{c} = \frac{\lambda}{c \cdot \mu}$

Вероятность простоя СМО $P_0 = [1 + \varphi + \frac{\varphi^2}{2!} + ... + \frac{\varphi^c}{c!}]^{-1} = [\sum_{i=0}^{c} \frac{\varphi^i}{i!}]^{-1}$

Вероятность что в СМО i заявки $P_i = \frac{\varphi^i}{i!} P_0$

Вероятность отказа в обслуживании $P_{отк} = P_c = \frac{\varphi^c}{c!} P_0$

Интенсивность потока заявок поступающих на обслуживание $\lambda_c = (1 - P_{отк}) \cdot \lambda$

Коэффициент использования ОА $U = \lambda_c / (c \cdot \mu) = (1 - P_{отк}) \cdot \rho$

Среднее число занятых каналов $k_{зан} = U \cdot c = (1 - P_{отк}) \cdot \varphi$

Среднее число заявок в очереди $Q = 0$

Среднее число заявок в СМО $L = Q + c \cdot U = c \cdot U$

Среднее время в очереди $W = 0$

Пропускная способность СМО $A = q \cdot \lambda$ $\quad q = (1 - P_{отк})$

Расчет показателей СМО с отказами

Показатели оценки качества функционирования СМО M/M/C/m

Загрузка СМО $\varphi = \dfrac{\lambda}{\mu}$

Загрузка ОА СМО $\rho = \dfrac{\varphi}{c} = \dfrac{\lambda}{c \cdot \mu}$

Вероятность простоя СМО $P_0 = \left[\sum_{i=0}^{c-1} \dfrac{\varphi^i}{i!} + \dfrac{\varphi^c (1-\rho^{m+1})}{c! \cdot (1-\rho)} \right]^{-1}$

Вероятность что в СМО i заявки $P_i = \dfrac{\varphi^i}{i!} P_0$

Вероятность отказа в обслуживании $P_{отк} = P_{m+c} = \dfrac{\varphi^{c+m}}{c! \cdot c^m} P_0$

Интенсивность потока заявок

поступающих на обслуживание $\lambda_c = (1 - P_{отк}) \cdot \lambda$

Коэффициент использования ОА $U = \lambda_c / (c \cdot \mu) = (1 - P_{отк}) \cdot \rho$

Среднее число занятых каналов $k_{зан} = U \cdot c = (1 - P_{отк}) \cdot \varphi$

Среднее число заявок в очереди $Q = \sum_{i=c+1}^{c+m} (i-c) \cdot P_i = \dfrac{\varphi^{c+m}}{c! \cdot c} \left[\dfrac{1 - \rho^m [(m+1) - m\rho]}{(1-\rho)^2} \right] P_0$

Среднее число заявок в СМО $L = \sum_{i=1}^{c+m} i \cdot P_i = Q + c \cdot U$

Среднее время в очереди $W = \dfrac{Q}{\lambda_c} = \dfrac{\varphi^c}{c! \cdot c \cdot \mu} \left[\dfrac{1 - \rho^m [(m+1) - m\rho]}{(1-\rho)^2} \right] P_0$

Пропускная способность СМО $A = q \cdot \lambda \qquad q = (1 - P_{отк})$

Пример M/M/1 в SimPy

Состав

*Packet*s: объекты типа Packet

packet_generator : функция для генерации Packets

packet_consumer : функция моделирующая связь с потребителем паектов из очереди

Store : модель очереди FIFO

Environment : среда моделирования

Пример M/M/1 в SimPy

```
class Packet(object):
    """A very simple class that represents a packet...."""
    def __init__(self, time, id, src="a", dst="z"):
        self.time = time
        self.id = id
        self.src = src
        self.dst = dst
```

```
def packet_generator(numPackets, env, out_pipe):
    """A generator function for creating packets...."""
    global queue_size
    for i in xrange(numPackets):
        # wait for next transmission
        yield env.timeout(random.expovariate(1/ARRIVAL))
        #print "Sending packet {} at time {}".format(i, 
        p = Packet(env.now, i)
        # Measuring queue statistics here is only valid 
        queue_size += len(out_pipe.items)
        yield out_pipe.put(p)
```

```
def packet_consumer(env, in_pipe):
    """A generator function which consumes packets...."""
    global queue_wait, total_wait
    while True:
        # Get event for message pipe
        msg = yield in_pipe.get()
        queue_wait += env.now - msg.time
        yield env.timeout(random.expovariate(1/SERVICE))
        total_wait += env.now - msg.time
        #print "at time {} processed packet: {} ".format(e
```

Пример M/M/1 в SimPy

```
## The simulation environment.
env = simpy.Environment()
## The switch output port object based on the SimPy Store class
pipe = simpy.Store(env)
# Turns our generator functions into SimPy Processes
env.process(packet_generator(NUM_PACKETS, env, pipe))
env.process(packet_consumer(env, pipe))
print 'A simple M/M/1 queueing simulation'
env.run()
print "Ending simulation time: {}".format(env.now)
```

Ресурс Store в SimPy

-- моделирует производство и потребление требуемых объектов.

Объекты, поступающие в Store, могут быть любого типа. По умолчанию они помещаются и извлекаются из хранилища в порядке FIFO.

Емкость (capacity) определяет размер хранилища и должна быть положительным числом. По умолчанию хранилище имеет неограниченный размер.

Операция Put(событие)

Используется для размещения объектов в хранилище, если емкость будет исчерпана, вызывающий процесс будет “заблокирован”, и его запрос на размещение будет добавлен в “очередь размещения”.

Операция Get(событие)

Используется для получения объектов из хранилища. Если хранилище пусто, вызывающий процесс будет “заблокирован” и помещен в “очередь получения” до тех пор, пока не появится элемент, подлежащий извлечению.

Пример M/M/1/n в SimPy

```
def packet_generator(numPackets, env, out_pipe):
    """A generator function for creating packets...."""
    global queue_size
    for i in xrange(numPackets):
        # wait for next transmission
        yield env.timeout(random.expovariate(1/ARRIVAL))
        #print "Sending packet {} at time {}".format(i, e
        p = Packet(env.now, i)
        # Measuring queue statistics here is only valid :
        queue_size += len(out_pipe.items)
        yield out_pipe.put(p)
```

Генератор без ограничений

Для моделирования ограниченной очереди добавим проверку состояния очереди Store в функцию –генератор.

Получаем размер очереди

Если буфер заполнен, то

не размещаем в очередь и считаем потери

```
def packet_generator(numPackets, env, out_pipe):
    """A generator function for creating packets...."""
    global queue_size, dropped
    for i in xrange(numPackets):
        # wait for next transmission
        yield env.timeout(random.expovariate(1/ARRIVAL))
        #print "Sending packet {} at time {}".format(i, e
        p = Packet(env.now, i)
        tmp = len(out_pipe.items)
        queue_size += tmp
        if tmp == BUFF_SIZE:
            dropped += 1
        else:
            yield out_pipe.put(p)
```