Параметры и характеристики СМО

Параметры СМО - величины, описывающие поток заявок СМО и каналы обслуживания.

Структурные параметры:

- K – количество каналов;
- n – количество накопителей и их емкости E_j $(j=1,2\ldots n)$;
- способ взаимосвязи накопителей с приборами (в случае многоканальных СМО), например в виде матрицы связей.

Нагрузочные параметры:

- количество поступающих в систему классов заявок Н;
- закон распределения интервалов времени между поступающими в систему заявками каждого класса (λ_i – интенсивность, среднее количество заявок i-класса , поступающих в СМО в единицу времени);
- закон распределения длительности обслуживания заявок каждого класса ($\mu_i = 1/b_i$ - интенсивность обслуживания заявок i-класса, среднее количество заявок, которое может быть обслужено одним каналом СМО в единицу времени при условии, что канал никогда не простаивает из-за отсутствия заявок; b_i - среднее время обслуживания заявки в канале).

Функциональные параметры (параметры управления):

- дисциплина буферизации;
- дисциплина обслуживания.

Обозначения СМО (символика Кендалла)

Для компактного описания систем массового обслуживания часто используются обозначения вида:

A/B/N/L ,

где **A** и **B** – задают законы распределений соответственно интервалов времени между моментами поступления заявок в систему и длительности обслуживания заявок в приборе; **N** – число обслуживающих приборов в системе (N = 1, 2, ...) ; **L** – число мест в накопителе (L= 0, 1, 2, …).

Для задания законов распределений **A** и **B** используются следующие обозначения:

G (General) – произвольное распределение общего вида;

M (Markovian) – экспоненциальное (показательное) распределение;

D (Deterministik) – детерминированное распределение;

U (Uniform) – равномерное распределение;

Ek (Erlangian) – распределение Эрланга k-го порядка (с k последовательными одинаковыми экспоненциальными фазами);

hk (hipoexponential) – гипоэкспоненциальное распределение k-го порядка (с k последовательными разными экспоненциальными фазами);

Hr (Hiperexponential) – гиперэкпоненциальное распределение порядка r (с r параллельными экспоненциальными фазами);

и т.д.

Примеры:
M/M/1 – одноканальная СМО с накопителем неограниченной ёмкости, в которую поступает однородный поток заявок с экспоненциальным распределением интервалов времени между последовательными заявками (простейший поток) и экспоненциальной длительностью обслуживания заявок в приборе.
M/G/3/10 – трёхканальная СМО с накопителем ограниченной ёмкости, равной 10, в которую поступает однородный поток заявок с экспоненциальным распределением интервалов времени между последовательными заявками (простейший поток) и длительностью обслуживания заявок, распределённой по закону общего вида.
D/E2/7/0 – семиканальная СМО без накопителя (ёмкость накопителя равна 0), в которую поступает однородный поток заявок с детерминированными интервалами времени между последовательными заявками (детерминированный поток) и длительностью обслуживания заявок в приборе, распределённой по закону Эрланга 2-го порядка.

Для обозначения более сложных СМО дополнительно могут использоваться обозначения, описывающие неоднородный поток заявок и приоритеты между заявками разных классов.

Режимы функционирования СМО

- установившийся или стационарный (вероятностные характеристики системы не изменяются со временем);
- неустановившийся (например, переходной режим, режим перегрузки, нестационарный режим, связанный с зависимостью от времени входящего потока)

Характеристики СМО с однородным потоком заявок

Характеристики СМО - величины, по которым можно оценивать эффективность работы СМО и выбирать лучший из нескольких вариантов СМО. Характеристики систем со стохастическим характером функционирования являются случайными величинами и полностью описываются соответствующими законами распределений. На практике при моделировании часто ограничиваются определением только средних значений (математических ожиданий), реже – определением двух первых моментов этих характеристик.

В качестве *основных характеристик СМО с однородным потоком заявок* используются следующие величины:

- *нагрузка системы* $y = \lambda / \mu = \lambda b$;
- *коэффициент загрузки* ρ или просто *загрузка* системы

$$\rho = \lim_{T \to \infty} \frac{1}{KT} \sum_{i=1}^{K} T_i \; ;$$

T - время наблюдения, K – количество обслуживающих приборов, T_i - время работы i – го прибора. Очевидно, что $0 \le \rho \le 1$;

- *коэффициент простоя системы*: $\eta = 1 - \rho$;

- *вероятность потери заявок*: $$\pi_n = \lim_{T\to\infty} \frac{N_n(T)}{N(T)}$$

$N(T)$ – число заявок, поступивших в систему за время T; $N_n(T)$ – число потерянных заявок за время T;

- *вероятность обслуживания заявки* (относительная пропускная способность СМО) :

$$\pi_0 = (1-\pi_n) = \lim_{T\to\infty} \frac{N_0(T)}{N(T)},$$

где $N_0(T)$ – число обслуженных в системе заявок за время T;

- *производительность системы (абсолютная пропускная способность* СМО), представляющая собой интенсивность потока обслуженных заявок, выходящих из системы: $\lambda' = \pi_0 \lambda = (1 - \pi_n)\lambda$;
- *интенсивность потока потерянных* (не обслуженных) *заявок*: $\lambda'' = \pi_n\lambda = (1 - \pi_0)\lambda$;

очевидно, что сумма интенсивностей потоков обслуженных и потерянных заявок должна быть равна интенсивности входящего в систему потока заявок: $\lambda' + \lambda'' = \lambda$;

- *среднее время ожидания заявок в очереди w*;
- *среднее время пребывания заявок в системе*, складывающееся из времени ожидания w и *среднего времени обслуживания* b: $u = w + b$;
- *средняя длина очереди заявок*: $\boldsymbol{l = \lambda' w}$;
- *среднее число заявок в системе* (в очереди и на обслуживании в приборе): $\boldsymbol{m = \lambda' u}$.

Формулы Литтла

Нагрузка $y = \lambda / \mu = \lambda b$ и загрузка ρ являются важнейшими характеристиками СМО, определяющими качество функционирования системы.

Если значение нагрузки $y < 1$, то заданная нагрузка может быть выполнена одним обслуживающим прибором, то есть одноканальная СМО будет работать без перегрузки.

Если $y > 1$, то реализация заданной нагрузки в одноканальной СМО приведет к режиму перегрузки, т.е. с течением времени всё большее число заявок будет оставаться не обслуженным, и в случае накопителя неограниченной емкости очередь заявок будет расти неограниченно.

Для того чтобы система работала без перегрузок необходимо использовать многоканальную СМО, количество приборов которой должно быть больше, чем значение нагрузки: $K > y$.

В общем случае для любой СМО с K каналами (с накопителем ограниченной или неограниченной ёмкости) загрузка системы может быть рассчитана через нагрузку следующим образом:

$\rho = \dfrac{(1-\pi_n)y}{K}$, если СМО работает без перегрузки, и $\rho = 1$, если СМО перегружена.

Таким образом, *загрузка системы, в отличие от нагрузки, определяется через интенсивность только обслуженных заявок, поскольку потерянные заявки не обслуживаются в приборах и, следовательно, не загружают систему*.

Характеристики СМО с неоднородным потоком заявок

Для СМО с неоднородным потоком заявок, в которую поступают H классов заявок с интенсивностями $\lambda_1,\ldots,\lambda_H$ и средними длительностями обслуживания заявок $b_1,\ldots,b_H$ определяются две группы характеристик обслуживания заявок:

- характеристики по каждому классу (потоку) заявок;
- характеристики объединённого (суммарного) потока заявок.

Характеристики по каждому классу заявок (i = 1, H) аналогичны характеристикам СМО с однородным потоком:

- нагрузка, создаваемая заявками класса i: $y_i = \lambda_i / \mu_i = \lambda_i b_i$;
- вероятность потери заявок: π_{n_i};
- вероятность обслуживания заявки: $\pi_{0_i} = (1-\pi_{n_i})$;
- интенсивность потока обслуженных заявок (производительность по i- му классу заявок):

$$\lambda_{0_i} = \pi_{0_i} \lambda_i = (1-\pi_{n_i})\lambda_i;$$

- интенсивность потока потерянных заявок: $\lambda_{n_i} = \pi_{n_i} \lambda_i$
- загрузка системы, создаваемая заявками класса i: $\rho_i = \min\left(\frac{(1-\pi_{n_i})y_i}{K};1\right)$,

где π_{n_i} – вероятность потери заявок класса i из-за ограниченной ёмкости накопителя;

- время ожидания заявок в очереди: w_i ;
- время пребывания заявок в системе: $u_i = w_i + b_i$;
- длина очереди заявок: $l_i = \lambda_i w_i$;
- число заявок в системе (в очереди и на обслуживании): $m_i = \lambda_i u_i$.

Характеристики объединённого (суммарного) потока заявок

позволяют определить усредненные по всем классам заявок показатели эффективности функционирования СМО:

- суммарная интенсивность поступления заявок в систему (интенсивность суммарного потока): $\Lambda = \sum_{i=1}^{H} \lambda_i$;
- суммарная нагрузка Y и суммарная загрузка R системы:

$Y = \sum_{i=1}^{H} y_i$; $R = \min(\sum_{i=1}^{H} \rho_i ;1)$, (условие отсутствия перегрузок в СМО с неоднородным потоком заявок и накопителем неограниченной ёмкости имеет вид: $R < 1$);

- коэффициент простоя системы: $\eta = 1 - R$;
- среднее время ожидания W и среднее время пребывания U заявок объединённого потока в системе:

$W = \sum_{i=1}^{H} \xi_i w_i$; $U = \sum_{i=1}^{H} \xi_i u_i$, где $\xi_i=\lambda_i/\Lambda$ – коэффициент, учитывающий долю заявок класса i в суммарном потоке;

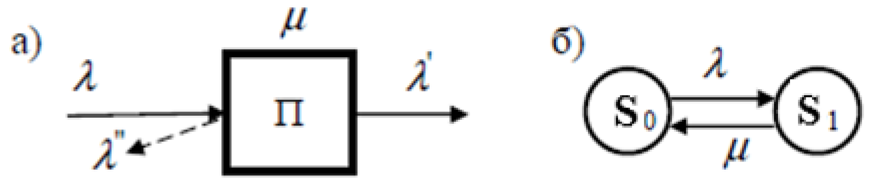
АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
Одноканальные СМО с однородным потоком заявок

- СМО содержит один обслуживающий прибор, в котором в каждый момент времени может обслуживаться только одна заявка;
- заявки поступают в СМО с интенсивностью λ ;
- средняя длительность обслуживания одной заявки в приборе равна b, причем длительности обслуживания разных заявок не зависят друг от друга;
- обслуживающий прибор не простаивает, если в системе (накопителе) имеется хотя бы одна заявка, причем после завершения обслуживания очередной заявки мгновенно из накопителя выбирается следующая заявка;
- в системе существует стационарный режим, предполагающий отсутствие перегрузок.

Одноканальная экспоненциальная СМО без накопителя (M/M/1/0)

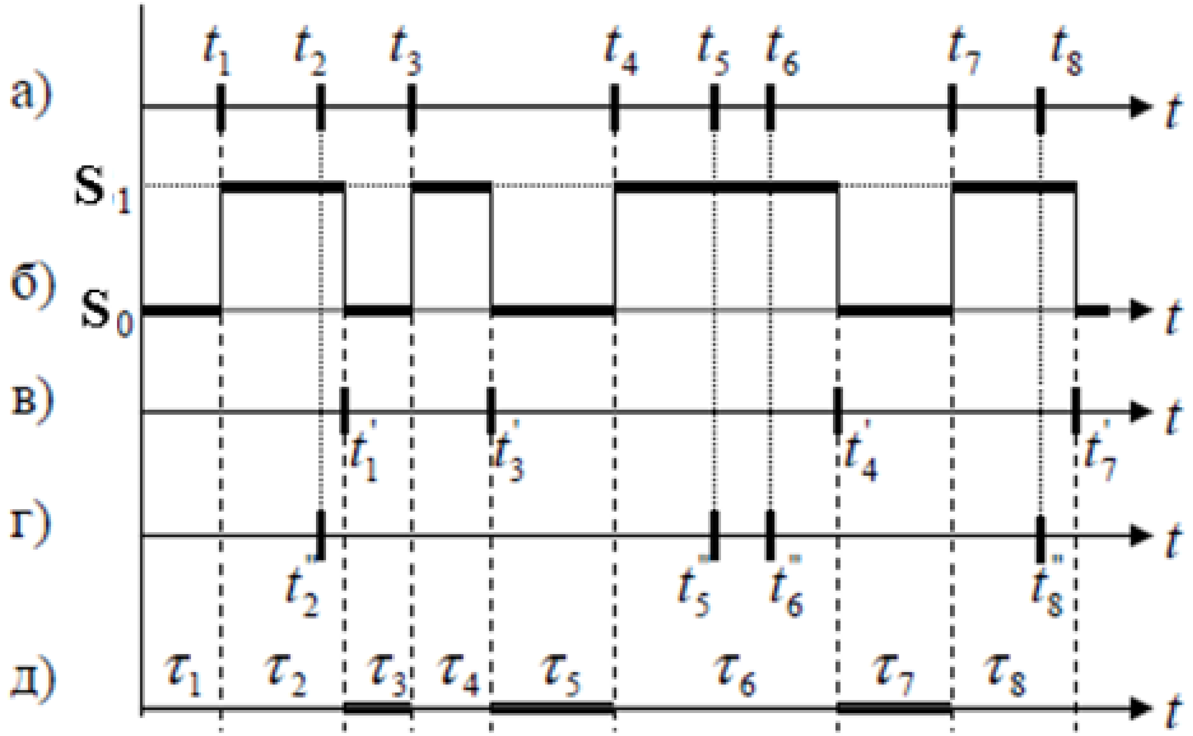
- Система содержит один обслуживающий прибор (П), то есть является одноканальной.
- Перед прибором не предусмотрены места для ожидания заявок, то есть в системе отсутствует накопитель.
- Поступающие в систему заявки образуют простейший поток с интенсивностью λ .
- Длительность обслуживания заявок в приборе распределена по экспоненциальному закону с интенсивностью $\mu = 1/b$, где b – средняя длительность обслуживания.
- Дисциплина буферизации – с отказами: заявка, поступившая в систему и заставшая прибор занятым обслуживанием другой заявки, теряется.
- Дисциплина обслуживания – в естественном порядке: заявка, поступившая в систему и заставшая прибор свободным, принимается на обслуживание.

Замечание: в СМО с отказами всегда будет существовать установившийся режим, поскольку даже при больших значениях нагрузки ($y >> 1$) число заявок в системе не может вырасти до бесконечности (с ростом нагрузки увеличивается доля заявок, получающих отказ в обслуживании).



СМО с отказами (а) и ее граф переходов (б)

Диаграммы процессов системы (M/M/1/0)



а) поступление в СМО заявок, интервалы между которыми в случае простейшего потока распределены по экспоненциальному закону;

б) переход из состояния S_0 в состояние S_1 и обратно, в которых может находиться система; время нахождения случайного процесса в состоянии S_1 равно длительности обслуживания заявки в приборе, которая представляет собой случайную величину, распределенную по экспоненциальному закону;

в) выход из системы обслуженных заявок;

г) выход из системы необслуженных заявок, получивших отказ из-за занятости прибора;

д) формирование интервалов времени между соседними переходами случайного процесса.

Характеристики СМО M/M/1/0

Уравнения Колмогорова (см. семинар): $\frac{dp_0}{dt} = \mu p_1 - \lambda p_0; \quad \frac{dp_1}{dt} = \lambda p_0 - \mu p_1.$

Решение: $p_0(t) = \frac{\mu}{\lambda+\mu}\left[1 + \frac{\lambda}{\mu}e^{-(\lambda+\mu)t}\right], \qquad p_1(t) = \frac{\lambda}{\lambda+\mu}\left[1 - e^{-(\lambda+\mu)t}\right].$

Система уравнений для определения стационарных вероятностей: $\begin{cases} \lambda p_0 = \mu p_1 \\ p_0 + p_1 = 1 \end{cases}.$

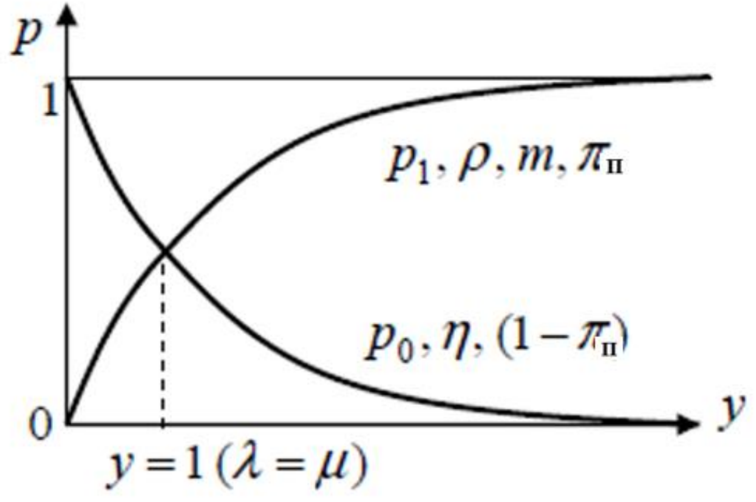
Финальные вероятности: $p_0 = \frac{\mu}{\lambda+\mu} = \frac{1}{1+y}; \qquad p_1 = \frac{\lambda}{\lambda+\mu} = \frac{y}{1+y}$

- нагрузка: $y = \lambda / \mu = \lambda b$ (по определению);
- загрузка (вероятность работы прибора): $\rho = p_1$;
- коэффициент простоя системы: $\eta = p_0 = 1 - \rho$;
- вероятность обслуживания заявки (относительная пропускная способность СМО) : $\pi_0 = p_0$
- вероятность потери заявок в результате отказа в обслуживании: $\pi_n = p_1$
- производительность системы (абсолютная пропускная способность СМО): $\lambda' = \pi_0 \lambda = (1 - \pi_n)\lambda = p_0 \lambda$;
- интенсивность потока потерянных (не обслуженных) заявок: $\lambda'' = \pi_n \lambda = p_1 \lambda$;
- среднее число заявок в системе: $m = p_1 = \rho$, определяемое как математическое ожидание случайной величины X– числа заявок в системе: $m = 0 \times p_0 + 1 \times p_1 = p_1$

x_i	0	1
p_i	p_0	p_1

- среднее время пребывания заявок в системе: $u = m/\lambda' = b$

Анализ полученных зависимостей показывает, что с ростом нагрузки коэффициент простоя системы $\eta = p_0$ уменьшается, а загрузка системы $\rho = p_1$ (а также среднее число заявок в системе и вероятность отказа) увеличиваются, причем их сумма всегда равна единице. При $y \to \infty$ коэффициент простоя $\eta \to 0$, в то время как загрузка $\rho \to 1$. Нагрузка системы также может быть определена через стационарные вероятности как отношение вероятности работы системы к вероятности простоя: $y = p_1 / p_0$, или через загрузку и коэффициент простоя: $y = \rho / \eta$.

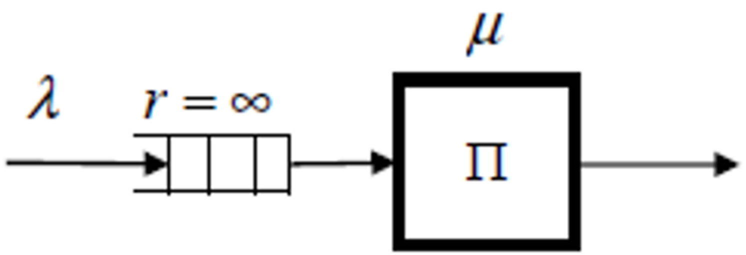


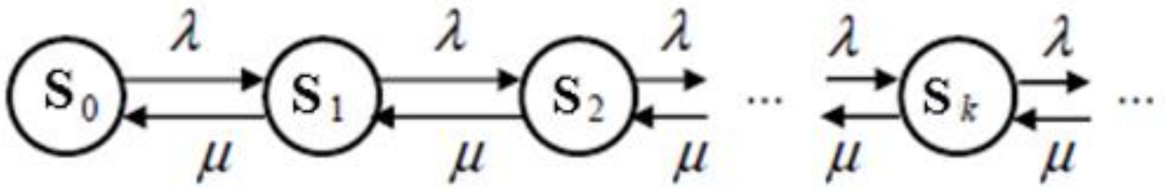
Пример. Узел связи принимает простейший поток сообщений с интенсивностью $\lambda = 5$ сообщений в секунду. Время передачи сообщений по каналу связи распределено по экспоненциальному закону. Среднее время передачи одного сообщения составляет 0,1 секунды. Сообщения, поступающие в моменты времени, когда обслуживающий канал занят передачей ранее поступившего сообщения, получают отказ и не передаются. Определить следующие показатели эффективности СМО при условии ее работы в установившемся режиме:

- вероятность отказа приема сообщения для передачи по каналу связи;
- загрузку СМО;
- относительную пропускную способность Q;
- абсолютную пропускную способность A.

Простейшая одноканальная СМО с неограниченной очередью (M/M/1/∞)

- Система – одноканальная – с одним обслуживающим прибором.
- Поток заявок простейший поток с интенсивностью λ .
- В приборе происходит задержка поступающих в систему заявок на некоторое случайное время. Длительность обслуживания заявок в приборе распределена по экспоненциальному закону с интенсивностью $\mu = 1/ b$, где b – средняя длительность обслуживания заявок в приборе.
- В системе имеется накопитель неограниченной ёмкости: $r = \infty$, то есть любая заявка, поступившая в систему, найдет место для ожидания в очереди и не будет потеряна.
- Дисциплина буферизации отсутствует, поскольку накопитель имеет неограниченную ёмкость.
- Дисциплина обслуживания – в порядке поступления по правилу «первым пришел – первым обслужен» (FIFO).
- Нагрузка системы совпадает с загрузкой, причём выполняется условие: $y = \rho < 1$, то есть система работает в установившемся режиме без перегрузок. При $y > 1$ в СМО устанавливается режим перегрузок.





В качестве параметра, описывающего состояние марковского процесса, будем рассматривать количество заявок k, находящихся в СМО (в приборе и в накопителе).

S_0 : k = 0 – в системе нет ни одной заявки;

S_1: k = 1 – в системе находится 1 заявка (на обслуживании в приборе);

S_2: k = 2 – в системе находятся 2 заявки (одна – на обслуживании в приборе и вторая ожидает в накопителе);

…

S_j: k =j в системе находятся j заявок (одна – на обслуживании в приборе и (j -1) – в накопителе).

…

Система уравнений для определения стационарных вероятностей:

$$\begin{cases} \lambda p_0 = \mu p_1 \\ (\lambda+\mu) p_1 = \lambda p_0 + \mu p_2 \\ (\lambda+\mu) p_2 = \lambda p_1 + \mu p_3 \\ \dots \\ (\lambda+\mu) p_k = \lambda p_{k-1} + \mu p_{k+1} \\ \dots \\ p_0 + p_1 + p_2 + \ldots + p_k + \ldots = 1 \end{cases}$$

решение может быть получено с помощью формул для финальных вероятностей в схеме гибели и размножения (см. пред. лекцию)

$$p_0 = \left\{1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \ldots + \frac{\lambda_0\lambda_1\ldots\lambda_{k-1}}{\mu_1\mu_2\ldots\mu_k} + \ldots + \frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n}\right\}^{-1};$$

$$p_1 = \frac{\lambda_0}{\mu_1} p_0; \quad p_2 = \frac{\lambda_0\lambda_1}{\mu_1\mu_2} p_0; \ \ldots; \quad p_k = \frac{\lambda_0\lambda_1\ldots\lambda_{k-1}}{\mu_1\mu_2\ldots\mu_k} p_0 \ (0 \le k \le n); \ \ldots; \qquad p_n = \frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n} p_0.$$

Финальные вероятности (при $y < 1$):

$$p_k = y^k(1-y) = \rho^k(1-\rho) \quad (k = 0, 1, 2, \ldots)$$

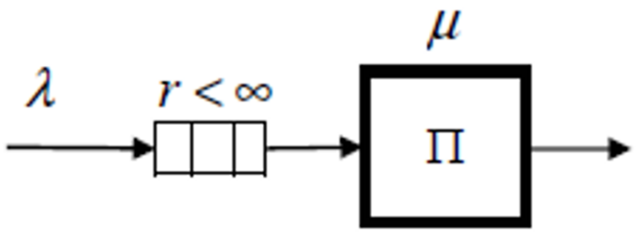
Характеристики СМО $M/M/1/\infty$

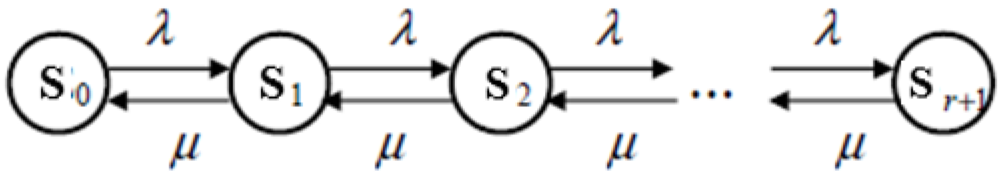
- нагрузка $y = \lambda / \mu = \lambda b$;
- загрузка $\rho = 1 - p_0 = \lambda b$ и совпадает с нагрузкой;
- коэффициент простоя системы $\eta = p_0 = 1 - \rho$;
- среднее число заявок в очереди: $l = \sum_{k=1}^{\infty}(k-1)p_k = \frac{\rho^2}{1-\rho}$;
-
- среднее число заявок в системе: $m = \sum_{k=0}^{\infty} k p_k = \frac{\rho}{1-\rho}$;
- вероятность потери заявок $\pi_n = 0$;
- производительность системы при отсутствии потерь совпадает с интенсивностью поступления заявок в систему: $\lambda' = \lambda$;
- интенсивность потерянных заявок $\lambda'' = 0$;
- среднее время ожидания заявок в очереди: $w = \frac{l}{\lambda} = \frac{\rho b}{1-\rho}$;
-
- среднее время пребывания заявок в системе: $u = w + b$ или $u = \frac{m}{\lambda} = \frac{b}{1-\rho}$

Одноканальная экспоненциальная СМО с накопителем ограниченной емкости (M/M/1/r)

- Система – одноканальная – с одним обслуживающим прибором.
- Поток заявок простейший поток с интенсивностью λ .
- В приборе происходит задержка поступающих в систему заявок на некоторое случайное время. Длительность обслуживания заявок в приборе распределена по экспоненциальному закону с интенсивностью $\mu = 1/ b$, где b – средняя длительность обслуживания заявок в приборе.
- Перед прибором имеется r мест для заявок, ожидающих обслуживания и образующих очередь, то есть в системе имеется накопитель ограниченной ёмкости: $r < \infty$.
- Дисциплина буферизации – с потерями: заявка, поступившая в систему и заставшая накопитель заполненным, теряется.
- Дисциплина обслуживания – в порядке поступления по правилу «первым пришел – первым обслужен» (FIFO).

Замечание: в СМО с накопителем ограниченной ёмкости всегда существует установившийся режим, поскольку длина очереди не будет расти до бесконечности даже при больших значениях нагрузки.





S_0 : k = 0 – в системе нет ни одной заявки;
S_1 : k = 2 – в системе находится 1 заявка на обслуживании в приборе;
S_2 : k = 2 – в системе находятся 2 заявки: одна – на обслуживании в приборе и вторая ожидает в накопителе;
…
S_{r+1} : k = r +1 – в системе находятся (r +1) заявок: одна – на обслуживании в приборе и r – в накопителе.

Система уравнений для определения стационарных вероятностей:

$$\begin{cases} \lambda p_0 = \mu p_1 \\ (\lambda+\mu)p_1 = \lambda p_0 + \mu p_2 \\ (\lambda+\mu)p_2 = \lambda p_1 + \mu p_3 \\ \dots \\ \mu p_{r+1} = \lambda p_r \\ \sum_{k=0}^{r+1} p_k = 1 \end{cases}.$$

Финальные вероятности:

$$p_0 = \frac{1}{\sum_{k=0}^{r+1} y^k} = \begin{cases} \frac{1-y}{1-y^{r+2}}, & y \neq 1 \\ \frac{1}{r+2}, & y = 1 \end{cases}$$

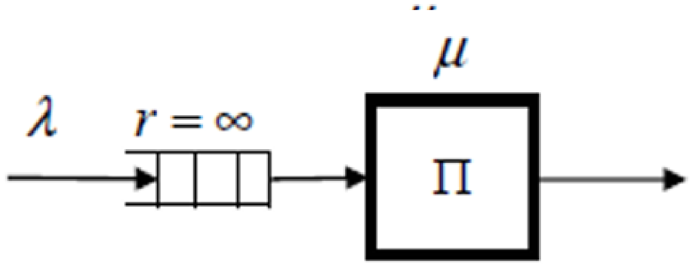
$$p_k = y^k p_0 = \begin{cases} \frac{y^k(1-y)}{1-y^{r+2}}, & y \neq 1 \\ \frac{y^k}{r+2}, & y = 1 \end{cases}$$

Характеристики СМО M/M/1/r

- нагрузка: $y=\lambda/\mu=\lambda b;$
- загрузка: $\rho=\sum_{k=1}^{r+1}p_k=1-p_0;$
- коэффициент простоя системы: $\eta=p_0=1-\rho$;
- среднее число заявок в очереди: $l=\sum_{k=2}^{r+1}(k-1)p_k = \dfrac{y^2(1-y^r(r+1-ry))}{(1-y^{r+2})(1-y)}.$
- среднее число заявок в системе: $m=\sum_{k=1}^{r+1}kp_k=l+\rho;$
- вероятность потери заявок: $\pi_n = p_{r+1}$;
- производительность системы $\lambda'=\lambda(1-\pi_n)$;
- интенсивность потока потерянных заявок $\lambda''=\lambda\pi_n$;
- среднее время ожидания заявок в очереди $w=l/\lambda'$;
- среднее время пребывания заявок в системе $u=m/\lambda'=w+b$

Одноканальная неэкспоненциальная СМО с неограниченной очередью M/G/1/∞

- Система – одноканальная – с одним обслуживающим прибором.
- Поток заявок простейший поток с интенсивностью λ .
- В приборе происходит задержка поступающих в систему заявок на некоторое случайное время. Длительность τ_b обслуживания заявок в приборе распределена по произвольному закону $B(\tau)$ со средним значением b (интенсивностью $\mu = 1/b$) и коэффициентом вариации v_b.
- В системе имеется накопитель неограниченной ёмкости: $r = \infty$, то есть любая заявка, поступившая в систему, найдет место для ожидания в очереди и не будет потеряна.
- Дисциплина буферизации отсутствует, поскольку накопитель имеет неограниченную ёмкость.
- Дисциплина обслуживания – в порядке поступления по правилу «первым пришел – первым обслужен» (FIFO).
- Нагрузка системы совпадает с загрузкой.



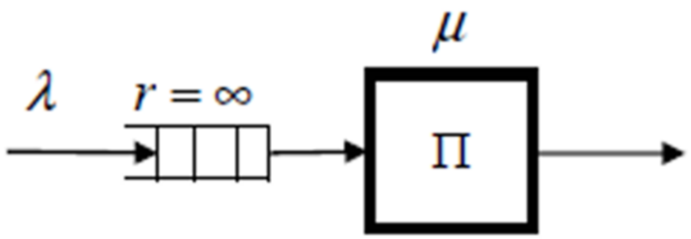
Характеристики СМО M/G/1/ ∞

- нагрузка $y = \lambda / \mu = \lambda b$;
- загрузка $\rho = \lambda b$ совпадает с нагрузкой;
- вероятность потери заявок $\pi_n = 0$;
- производительность системы при отсутствии потерь совпадает с интенсивностью поступления заявок в систему: $\lambda' = \lambda$;
- интенсивность потерянных заявок $\lambda'' = 0$;
- среднее время ожидания заявок в очереди (формула Поллачека- Хинчина): $w = \dfrac{\lambda b^2 (1 + \nu_b{}^2)}{2(1-\rho)}$
- среднее время пребывания заявок в системе: $u = w + b = \dfrac{\lambda b^2 (1 + \nu_b{}^2)}{2(1-\rho)} + b$
- среднее число заявок в очереди: $l = \lambda w$;
- среднее число заявок в системе: $m = \lambda u$.

Замечание: средние значения характеристик обслуживания заявок зависят только от двух первых моментов длительности обслуживания заявок и не зависят от моментов более высокого порядка. Следовательно, для того чтобы рассчитать средние характеристики обслуживания, необязательно знать закон распределения длительности обслуживания заявок – достаточно знать только два первых момента распределения. Можно показать, что для расчета вторых моментов характеристик обслуживания заявок достаточно задать три первых момента длительности обслуживания и т.д. Для СМО с простейшим потоком заявок для расчёта первых k моментов характеристик обслуживания необходимо задать (k+1) моментов длительности обслуживания заявок.

Одноканальная неэкспоненциальная СМО с неограниченной очередью G/M/1/∞

- Система – одноканальная – с одним обслуживающим прибором.
- Поток заявок произвольного вида, задаваемый функцией распределения интервалов между заявками $A(\tau)$ и с интенсивностью λ.
- В приборе происходит задержка поступающих в систему заявок на некоторое случайное время. Длительность τ_b обслуживания заявок в приборе распределена по экспоненциальному закону $B(\tau)$ со средним значением b (интенсивностью $\mu = 1/b$) и коэффициентом вариации v_b.
- В системе имеется накопитель неограниченной ёмкости: $r = \infty$, то есть любая заявка, поступившая в систему, найдет место для ожидания в очереди и не будет потеряна.
- Дисциплина буферизации отсутствует, поскольку накопитель имеет неограниченную ёмкость.
- Дисциплина обслуживания – в порядке поступления по правилу «первым пришел – первым обслужен» (FIFO).
- Нагрузка системы совпадает с загрузкой.



СМО G/M/1 является симметричной по отношению к СМО M/G/1. Однако получение конечного результата в виде аналитического выражения для расчёта среднего времени ожидания, по аналогии с предыдущей моделью, в общем случае, оказывается невозможным. Это обусловлено тем, что среднее время ожидания, как и другие числовые моменты, зависит не только от двух первых моментов интервалов между поступающими заявками, но и от моментов более высокого порядка, т.е. от закона распределения интервалов между заявками.

Характеристики СМО G/M/1/∞

- нагрузка $y = \lambda / \mu = \lambda b$;
- загрузка $\rho = \lambda b$ совпадает с нагрузкой;
- вероятность потери заявок $\pi_n = 0$;
- производительность системы при отсутствии потерь совпадает с интенсивностью поступления заявок в систему: $\lambda' = \lambda$;
- интенсивность потерянных заявок $\lambda'' = 0$;
- среднее время ожидания заявок в очереди: $w = \varsigma b/(1-\varsigma)$, где ζ - единственный в области $0 < \zeta < 1$ корень уравнения $\varsigma = A^*(\mu - \mu\varsigma)$.

  A(s)*-преобразование Лапласа плотности распределения *a(τ)* интервалов между поступающими в систему заявками:

$$A^*(s) = \int_0^\infty e^{-s\tau} a(\tau) d\tau \qquad (s \geq 0).$$

- среднее время пребывания заявок в системе: $u = w + b$
- среднее число заявок в очереди: $l = \lambda w$;
- среднее число заявок в системе: $m = \lambda u$.

Пример. Применение описанного метода расчета к рассмотренной выше СМО М/М/1 с простейшим потоком заявок. В простейшем потоке интервалы времени между последовательными заявками распределены по экспоненциальному закону, преобразование Лапласа которого имеет вид: $A(s)=\dfrac{\lambda}{\lambda+s}$.

Тогда:

$$\varsigma=\frac{\lambda}{\lambda+\mu-\mu\varsigma} \;\rightarrow\; \mu\varsigma^2-(\lambda+\mu)\varsigma+\lambda=0 \;\rightarrow\; \varsigma^2-(1+\rho)\varsigma+\rho=0.$$

Из двух корней $\varsigma_1 = 1$ и $\varsigma_2 = \rho$ последнего уравнения условию $0 < \varsigma < 1$ удовлетворяет только второй корень. Подставляя его в выражение $w=\varsigma b/(1-\varsigma)$, получим выражение для среднего времени ожидания, совпадающее с известным для СМО М/М/1.

Одноканальная неэкспоненциальная СМО с неограниченной очередью G/G/1/∞

- Система – одноканальная – с одним обслуживающим прибором.
- Поток заявок произвольного вида, задаваемый функцией распределения интервалов между заявками $A(\tau)$ и с интенсивностью λ.
- В приборе происходит задержка поступающих в систему заявок на некоторое случайное время. Длительность τ_b обслуживания заявок в приборе распределена по произвольному закону $B(\tau)$ со средним значением b (интенсивностью $\mu = 1/b$) и коэффициентом вариации v_b.
- В системе имеется накопитель неограниченной ёмкости: $r = \infty$, то есть любая заявка, поступившая в систему, найдет место для ожидания в очереди и не будет потеряна.
- Дисциплина буферизации отсутствует, поскольку накопитель имеет неограниченную ёмкость.
- Дисциплина обслуживания – в порядке поступления по правилу «первым пришел – первым обслужен» (FIFO).
- Нагрузка системы совпадает с загрузкой.

Для большинства законов распределений интервалов между поступающими в систему заявками и длительностей их обслуживания в приборе невозможно получить точное решение в аналитической форме. Однако, при исследовании реальных систем редко бывают известны законы распределений указанных величин. Обычно при описании процессов поступления заявок в систему и их обслуживания в приборе ограничиваются несколькими моментами соответствующих распределений, чаще всего – двумя первыми моментами, задаваемыми в виде математического ожидания и среднеквадратического отклонения или коэффициента вариации искомой случайной величины.

- среднее время ожидания заявок в очереди может быть оценено по формуле: $\tilde{w} \approx \dfrac{\rho b(v_a^2+v_b^2)}{2(1-\rho)} f(v_a)$,

где $\rho=\lambda b<1$ - загрузка системы; λ, v_a - интенсивность потока заявок и коэффициент вариации интервалов между поступающими в систему заявками; b, v_b - среднее значение и коэффициент вариации длительности обслуживания заявок; $f(v_a)$ - корректирующая функция, зависящая от значения коэффициента вариации v_a.

$$f(v_a) = \begin{cases} \exp\left[-\dfrac{2(1-\rho)(1-v_a^2)^2}{3\rho(v_a^2+v_b^2)}\right], & v_a<1 \\ \exp\left[-(1-\rho)\dfrac{v_a^2-1}{v_a^2+4v_b^2}\right], & v_a \ge 1. \end{cases}$$

- среднее время пребывания заявок в системе: $u = w+b$
- среднее число заявок в очереди: $l = \lambda w$;
- среднее число заявок в системе: $m = \lambda u$.

Замечание 1. При решении многих практических задач выходящий поток заявок из одной СМО является входящим потоком в другую СМО. В этом случае для расчёта характеристик функционирования второй СМО необходимо знать характер входящего потока, наиболее полно описываемый законом распределения интервалов между последовательными заявками. В то же время, для проведения оценочных расчётов во многих случаях достаточно знание первых двух моментов этих интервалов: математического ожидания и коэффициента вариации.

Очевидно, что в СМО с накопителем *неограниченной* ёмкости, работающей без перегрузок, интенсивность выходящего потока заявок равна интенсивности входящего потока, то есть математические ожидания интервалов между последовательными заявками на выходе и входе СМО совпадают.

Для СМО G/G/1 коэффициент вариации выходящего потока заявок может быть рассчитан по следующей приближённой формуле

$$\nu_c^2 \approx \nu_a^2 + 2\rho\nu_b^2 - 2\rho(1-\rho)\frac{\tilde{w}}{b}.$$

Замечание 2. Для бесприоритетных дисциплин обслуживания в обратном порядке (ООП) и обслуживания в случайном порядке (ОСП) средние времена ожидания заявок будут такими же, как и при обслуживании в порядке поступления, но дисперсии времени ожидания будут больше. Это обусловлено, в частности для дисциплины ООП, тем, что часть заявок, поступивших последними, будут ожидать незначительное время, в то время как заявки, попавшие в начало очереди, могут ожидать обслуживания достаточно долго, то есть увеличивается разброс значений времени ожидания относительно среднего значения.

Пример. *Дублированная СМО с восстановлением* (классическая задача теории надежности).

Некоторое устройство в процессе работы может выходить из строя. Имеется резервное устройство, которое в случае неисправности основного автоматически включается в работу. В этот же момент начинается восстановление основного. Предполагается, что резерв ненагруженный, т. е. во время работы основного устройства резервное не может потерять работоспособность.
Дано: λ – интенсивность потока отказов, μ – интенсивность восстановления.
Тогда $1/\lambda$ - ожидаемая наработка на отказ, т. е. среднее время работы устройства до его отказа, $1/\mu$ - ожидаемое время восстановления неисправного устройства, т. е. среднее время устранения неисправности.
Изначально система находится в состоянии S_0 – работает основное устройство. В случае выхода из строя основного устройства, система переходит в состояние S_1 – работает резервное устройство. Если во время работы резервного устройства было восстановлено основное, система возвращается в S_0. Если же до восстановления основного устройства вышло из строя резервное, система переходит в состояние S_2, что фактически означает прекращение работы системы.