Система массового обслуживания

ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ

это прикладная область теории случайных процессов, занимающаяся исследованием вероятностных моделей реальных систем обслуживания

Основоположник ТМО:
Агнер Эрланг (1878 – 1929)
занимался решением задач телефонии

Термин ТМО ввёл:
А. Я. Хинчин (1894 – 1959)

Основоположник теории массового обслуживания

Ангер Краруп Эрланг (1878—1929)
Датский математик и инженер, один из основателей ТМО.
1909 год – опубликована работа «Теория вероятностей и телефонные разговоры» (The Theory of Probabilities and Telephone Conversations.) , получившая признание во всем мире.

В его честь названа единица измерения трафика в телекоммуникационных системах – эрланг. 1 эрланг (1 Эрл) эквивалентен разговору двух абонентов в течение 1 часа.

Формулой Эрланга пользуются до сих пор.

СМО – это система, в которой, с одной стороны, возникают массовые запросы (требования) на выполнение каких-либо услуг, а с другой происходит удовлетворение этих запросов.

Определение: **Система массового обслуживания (СМО)** – это совокупность приборов, каналов, станков, линий обслуживания, на которые в случайные или детерминированные моменты времени поступают заявки на обслуживание.

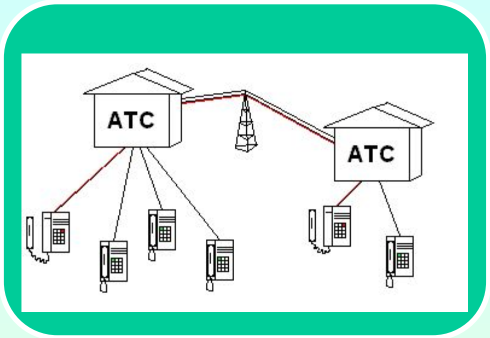
Примеры СМО:

- ✔ вычислительные комплексы,
- ✔ банковские системы
- ✔ торговые терминалы
- ✔ коммутаторы телефонных станций
- ✔ информационные службы
- ✔ комбинаты бытового обслуживания и т.д.

Элементы СМО:

- источник требований (заявка на обслуживание)
- входящий поток требований
- очередь
- обслуживающие устройства (каналы обслуживания)
- выходящий поток требований

Одноканальные



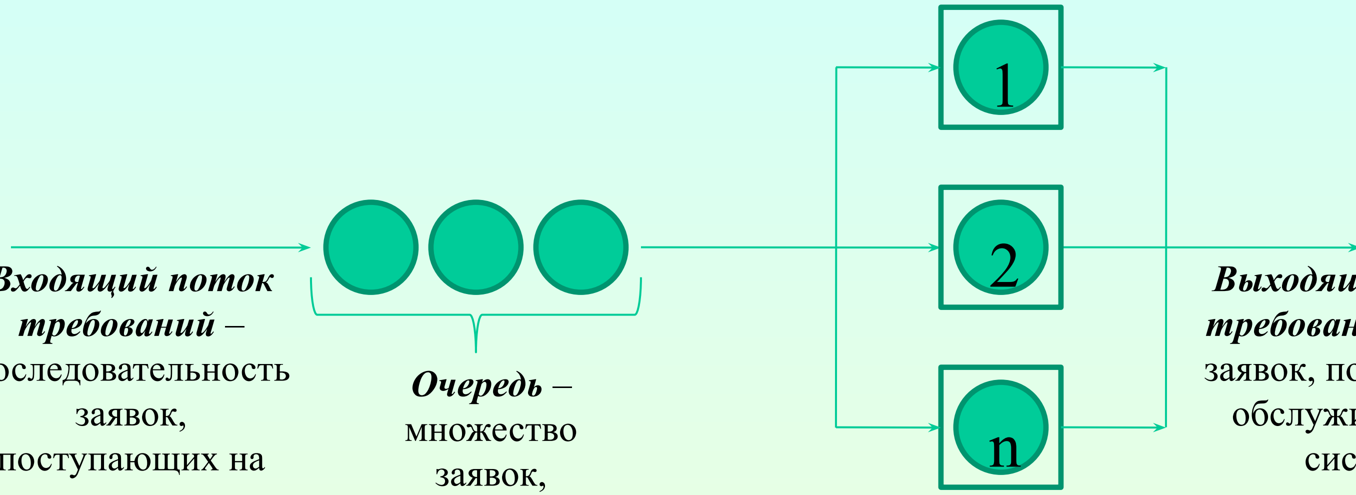
Многоканальные

СМО с очередью

Показатели эффективности СМО:

- среднее число заявок, обслуживаемых в единицу времени;
- среднее время ожидания обслуживания;
- среднее число заявок в очереди;
- вероятность отказа в обслуживании без ожидания;
- вероятность превышения числа заявок в очереди определенного значения и др.

СТРУКТУРА СМО



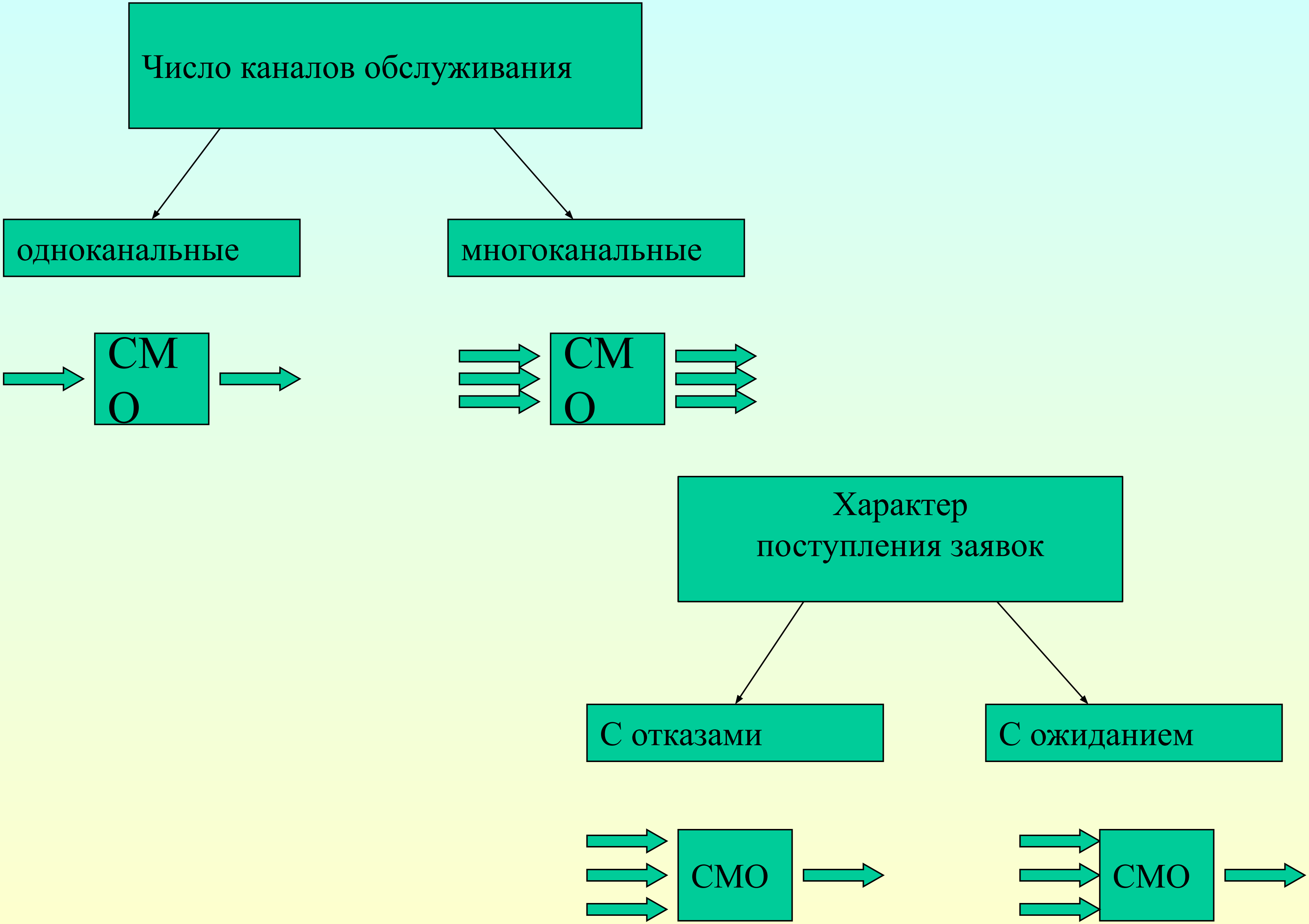
Входящий поток требований – последовательность заявок, поступающих на пункт обслуживания

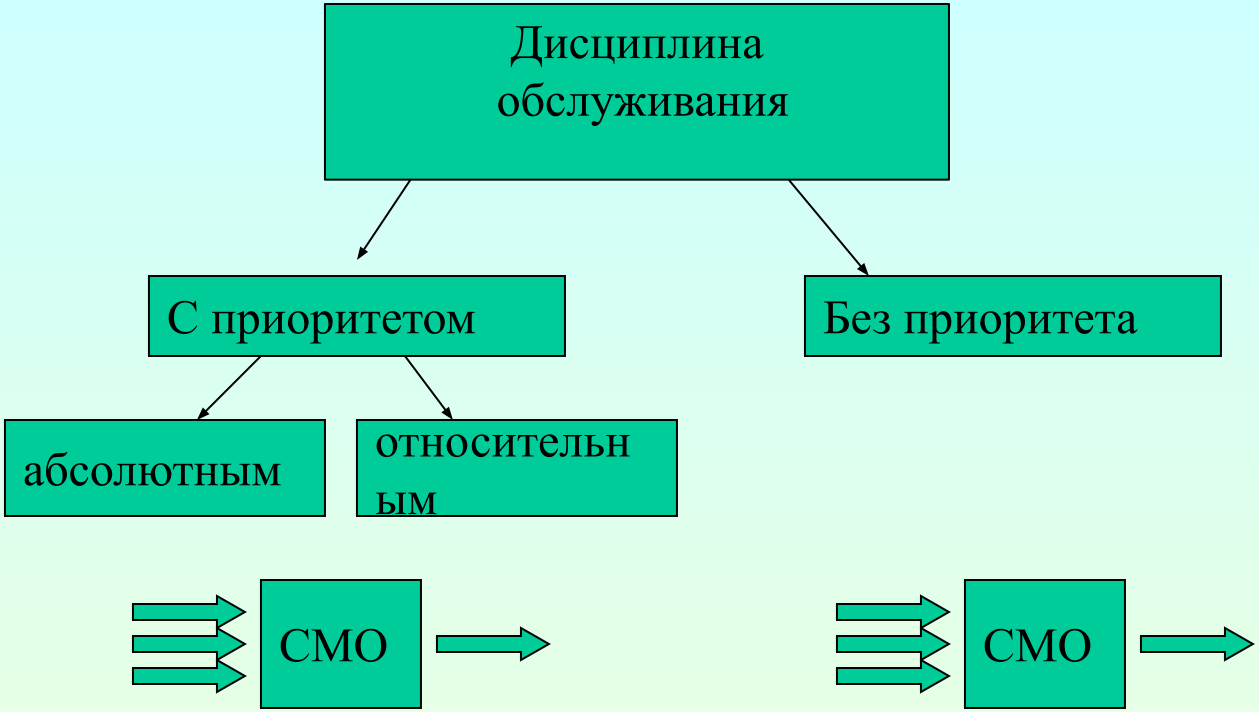
Очередь – множество заявок, ожидающих обслуживания

Каналы обслуживания – совокупность устройств, выполняющих операции по обслуживанию заявок

Выходящий поток требований – поток заявок, покидающих обслуживающую систему

Классификация СМО производится по различным признакам

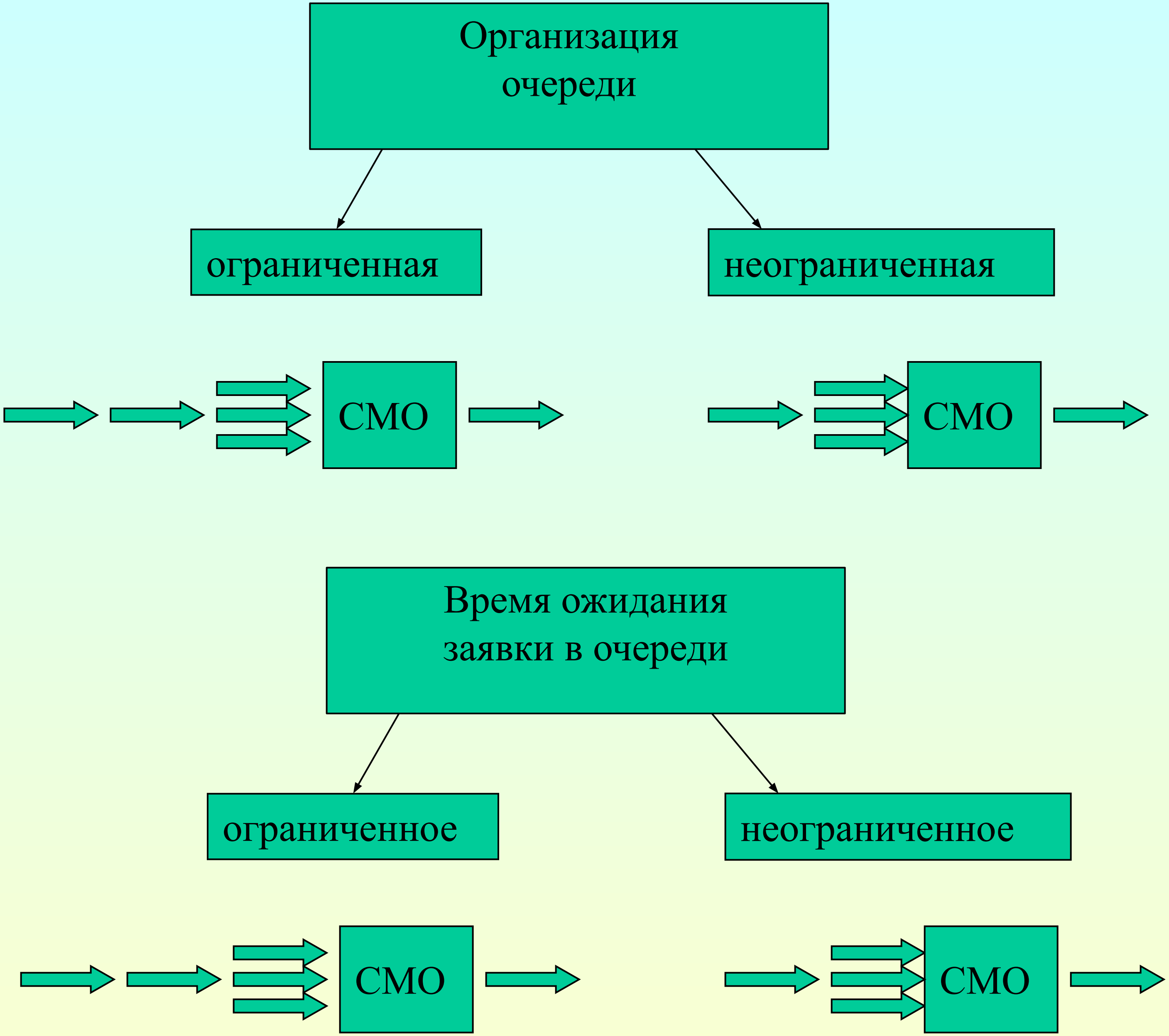




5. ***По приоритетности обслуживания:***

без приоритета - требования обслуживаются в порядке их поступления на СМО;

с приоритетом - требования обслуживаются в зависимости от присвоенного им при поступлении ранга приоритетности (например, заправка автомобилей скорой помощи на АЗС; первоочередной ремонт на АТП автомобилей, приносящих наибольшую прибыль на перевозках).


Организация очереди
ограниченная
неограниченная
СМО
СМО
Время ожидания заявки в очереди
ограниченное
неограниченное
СМО
СМО

Классификация систем массового обслуживания

- Пвх – характер входящего потока
- Воб – распределение времени обслуживания
- Nпр – число обслуживающих приборов
- Енак – емкость накопителя (длина очереди)

$$П_{вх} \;/\; В_{об} \;/\; N_{пр} / \; Е_{нак}$$

Характер входящего потока

M (Markovian) - входящий поток требований является Пуассоновским, т.е. распределение времени между поступающими заявками подчинено экспоненциальному закону;

E (Erlangian) - входящий поток является Эрланговским;

D (Determenistic) - детерминированный постоянный поток;

G (General) - произвольный рекуррентный поток.

Распределение времени обслуживания

M - распределение по экспоненциальному закону;

E - распределение по закону Эрланга;

D - время обслуживания постоянная величина;

G - произвольное распределение времени обслуживания.

Классификация систем
с Марковскими процессами обслуживания

- M/M/1/0 - одноканальная СМО с отказами;
- M/M/n/0 - многоканальная СМО с отказами;
- M/M/1/m - одноканальная СМО с ожиданием (ёмкость накопителя равна m);
- M/M/n/m - многоканальная СМО с ожиданием, но с возможностью отказа (число каналов - n, ёмкость накопителя равна m);
- M/M/1/∞ - одноканальная СМО с ожиданием без отказа (ёмкость накопителя равна ∞).

Формула Литтла

$$T_{\text{сист}} = \frac{1}{\lambda} \cdot L_{\text{сист}}$$

$T_{\text{оч}} = \frac{1}{\lambda} \cdot L_{\text{оч}}$, где

$T_{\text{сист}}$ – время пребывания заявки в системе.

$T_{\text{оч}}$ – время пребывания заявки в очереди.

$L_{\text{сист}}$ – среднее число заявок в системе.

$L_{\text{оч}}$ – среднее число заявок в очереди.

Формула выведена из принципа, что в стационарном режиме работы СМО среднее число заявок, прибывающих в систему, равно среднему числу заявок, покидающих ее: оба потока заявок имеют одну и ту же интенсивность λ.

Показатели качества обслуживания СМО

- Ротк – вероятность потери заявки (вероятность отказа),
- Ро – вероятность простоя,
- λ – интенсивность поступления заявок,
- μ – интенсивность обслуживания,
- $\rho=\lambda/\mu$ – приведенная интенсивность потока заявок,
- $A=\lambda*q$ – абсолютная пропускная способность,
- Q – среднее число заявок за единицу времени,
- ω – среднее число заявок под обслуживанием

для M/M/n/m $\omega=z$,

для M/M/1/∞, при $\rho>1$ $\omega=\rho$

tож –среднее время ожидания в очереди,

tсист – общее время пребывания в системе

z – среднее число занятых каналов для многоканальных СМО

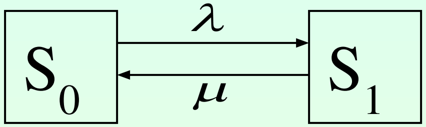
Одноканальная СМО с отказами (M/M/1/0)

Имеет 1 канал на который поступает поток заявок с интенсивностью λ.

μ, - интенсивность потока обслуживания

$\bar{t}_{об}$- среднее время обслуживания;

Граф состояний:



Состояние S_0 – канал свободен;

Состояние S_1 – канал занят;

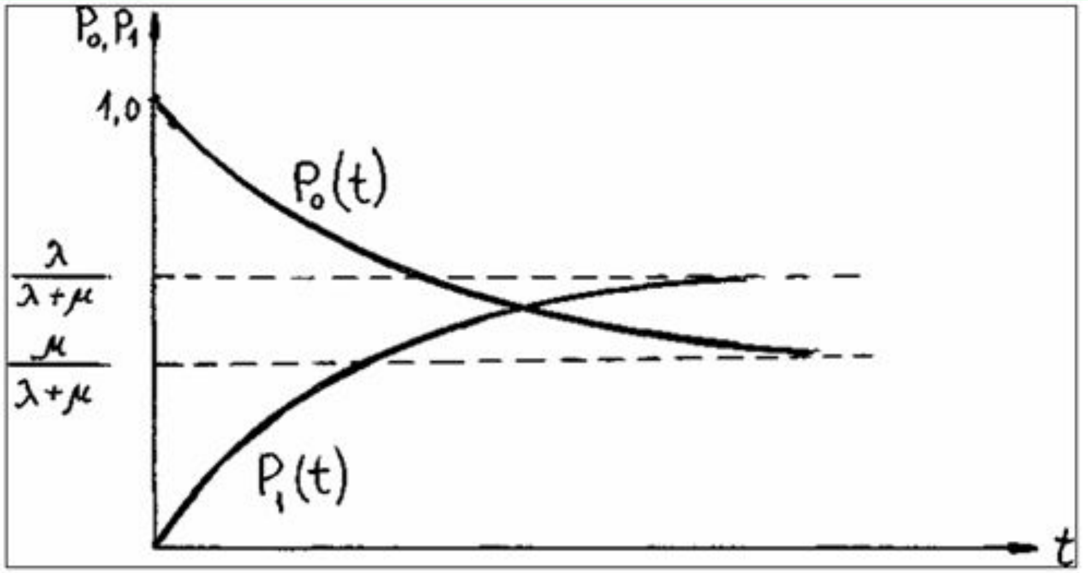
$$\begin{cases} P_1\mu - P_0\lambda = 0 \\ P_0\lambda - P_1\mu = 0 \end{cases}$$

$P_0 = \frac{\mu}{\lambda + \mu}$ q= P_0 (t)= $\frac{\mu}{\lambda + \mu}$

Pотк $= 1 - \frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}$ вероятность отказа в обслуживании заявки

$A = \lambda P_0 = \frac{\lambda\mu}{\lambda+\mu}$ абсолютная пропускная способность (сколько заявок в единицу времени обслуживается)

$A = \frac{\mu}{\lambda+\mu}$ относительная пропускная способность (какая часть заявок будет обслужена)



$\overline{T}_{обсл} = \frac{1}{\mu}$ - среднее время обслуживания одной заявки

Пример: Заявки на телефонные переговоры поступают диспетчеру с интенсивностью λ=90 заявок в час (l/ч). Средняя продолжительность разговора по телефону $\bar{t}_{об}$ = 2 мин. Определить показатели эффективности работы СМО при наличии одного телефонного номера.

<u>Решение:</u> $\lambda = 90(1/час)$ $\mu = \dfrac{1}{\bar{t}_{об}} = \dfrac{1}{2}(1/мин) = 30(1/час)$

Тогда, $Q = \dfrac{30}{90+30} = 0{,}25$, т.е. в среднем диспетчер ответит только на 25% звонков

Вывод: Одного номера <u>недостаточно</u>

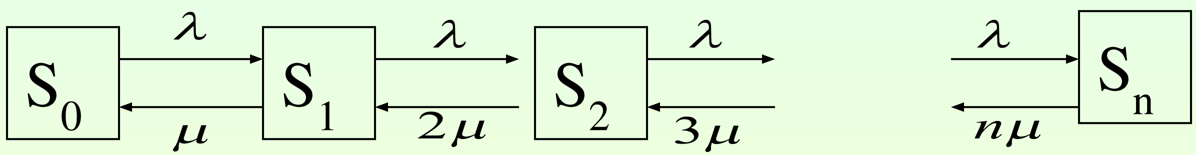
$P_{отк} = 0{,}75$ $A = 90 \cdot 0{,}25 = 22.5$ - среднее число обслуженных заявок

Многоканальная СМО с отказами (M/M/n/0)

Рассмотрим классическую задачу Эрланга:

Имеется *n* каналов, на которые поступает поток заявок ,с интенсивностью λ. Поток обслуживаний имеет интенсивность μ. Найти предельные вероятности состояний системы и показатели ее эффективности.

Граф состояний СМО соответствует процессу гибели и размножения:



Состояние S_k – когда в СМО заняты *k* каналов.

По формулам для процесса гибели и размножения:

$$p_0 = (1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2!\mu^2} + ... + \frac{\lambda^k}{k!\mu^k} + ... + \frac{\lambda^n}{n!\mu^n})^{-1};$$

Обозначим $\rho = \dfrac{\lambda}{\mu}$ – приведенная интенсивность потока заявок;

Тогда, $$p_0 = (1 + \rho + \frac{\rho^2}{2!} + ... + \frac{\rho^k}{k!} + ... + \frac{\rho^n}{n!})^{-1};$$

$\Leftarrow$ **Формулы Эрланга**

$p_1 = \rho p_0;$ $\quad p_2 = \dfrac{\rho^2}{2!} p_0;$

………..

$p_k = \dfrac{\rho^k}{k!} p_0;$ ……… $p_n = \dfrac{\rho^n}{n!} p_0;$

Вероятность отказа СМО есть предельная вероятность того, что все n каналов системы будут заняты:

$$P_{отк} = \frac{\rho^m}{n!} p_0$$

Относительная пропускная способность:

$$Q = 1 - P_{отк} = 1 - \frac{\rho^m}{n!} p_0$$

Абсолютная пропускная способность: $A = \lambda Q = \lambda(1 - \frac{\rho^m}{n!} p_0)$

Среднее число занятых каналов $\bar{k} = \sum_{k=0}^{n} k p_k$

Пример: В условиях предыдущей задачи определить оптимальное число телефонных номеров, если условием оптимальности считать удовлетворение в среднем из каждых 100 заявок не менее 90 заявок на переговоры.

Решение: Рассчитаем интенсивность нагрузки канала:

$\rho = \frac{90}{30} = 3$, т.е. за время среднего (по продолжительности) телефонного разговора поступает в среднем 3 заявки на переговоры.

Будем постепенно увеличивать число каналов (телефонных номеров) $n=2$,3,4,… и определять характеристики СМО.

Например, при n=2:

$$p_0 = (1 + 3 + \frac{3^2}{2!})^{-1} = 0.118 \qquad Q = 1 - (\frac{3^2}{2!})^{-1} \cdot 0.118 = 0.471$$

$$A = 90 \cdot 0.471 = 42.4$$

Характеристики	Число каналов (телефонных номеров)					
	1	2	3	4	5	6
Q	0.25	0.47	0.65	0.79	0.90	0.95
A	22.5	42.4	58.8	71.5	80.1	85.3

По условию оптимальности Q≥0.90 ⇒ необходимо установить <u>5 телефонных номеров.</u>

Пример: В вычислительный центр коллективного пользования с 3 рабочими станциями поступают заказы на вычислительные работы. При загрузке всех ЭВМ вновь поступивший заказ не принимается. Среднее время работы с одним заказом – 3 часа. Интенсивность потока заявок 0.25 (1/ч). Найти предельные вероятности состояний и показатели эффективности.

Решение: По условию n=3 λ=0.25(1/ч) $\bar{t}_{об}$ = 3 часа

Отсюда, $\mu = \frac{1}{\bar{t}_{об}} = \frac{1}{3} = 0.33$

Интенсивность нагрузки: $\rho = \frac{0.25}{0.33} = 0.75$ (1/ч)

Рассчитаем предельные вероятности:

$$p_0 = (1 + 0.75 + \frac{0.75^2}{2!} + \frac{0.75^3}{3!})^{-1} = 0.476$$

$$p_1 = 0.75 \cdot 0.476 = 0.357$$

$$p_2 = \frac{0.75^2}{2!} \cdot 0.476 = 0{,}134$$

$$p_3 = \frac{(0.75)^3}{3!} \cdot 0.476 = 0{,}033$$

<u>Вывод</u>: В стационарном режиме в среднем **47%** времени нет ни одной заявки; **35,7 %** времени – обрабатывается 1 заявка; **13,4%** времени – обрабатываются 2 заявки; **3,3%** времени – обрабатываются 3 заявки;

$$\Rightarrow P_{отк} = p_3 = 0{,}033$$

Q=1-0.033=0.067 $\Rightarrow$ из каждых 100 заявок в среднем будет обслужено 96,7;

A=0.25*0.967=0.242 $\Rightarrow$ в час будет обслужено в среднем 0.242 заявки

$$\bar{k} = \frac{0.242}{0.33} = 0,725$$

$\Rightarrow$ каждая из трех ЭВМ будет занята обслуживанием заявок в среднем на $\frac{72.5}{3} = 29,2\%$

<u>Замечание:</u> При оценке эффективности работы вычислительного центра необходимо сопоставить доходы от выполнения заявок с потерями от простоев ЭВМ. Что важнее – высокая пропускная способность СМО или значительный простой оборудования?