Лекция 5
Математические модели и методы теории систем массового обслуживания, используемые в САПР КЭС

Вопросы лекции

1. Предметная область теории систем массового обслуживания.
2. Классификация систем массового обслуживания.
3. Методы анализа и синтеза систем массового обслуживания. Основные характеристики и методы их расчета

Вопрос 1.
Предметная область теории систем массового обслуживания

Система массового обслуживания (СМО) — система, которая производит обслуживание поступающих в неё требований.

Теория массового обслуживания (теория очередей) — раздел теории вероятностей, целью исследований которого является рациональный выбор структуры системы обслуживания и процесса обслуживания на основе изучения потоков требований на обслуживание, поступающих в систему и выходящие из неё, длительности ожидания и длины очередей.

теория телетрафика - математическая теория, являющаяся одной из ветвей теории массового обслуживания. Применяется, прежде всего, для изучения и проектирования систем телекоммуникаций. Однако, разрабатываемые средства теории телетрафика являются независимыми от конкретной техники, и могут использоваться в области дорожного (авто) и воздушного (авиа) трафика, на производстве, при хранении и распределении готовых товаров, в общем, во всех системах обслуживания

Основы теории телетрафика были заложены в работах **А. К. Эрланга** по исследованию пропускной способности полнодоступного пучка линий, обслуживающего простейший поток вызовов с потерями и с ожиданием. Труды А. К. Эрланга послужили толчком для других работ, которые были связаны с подтверждением, развитием или опровержением его результатов.

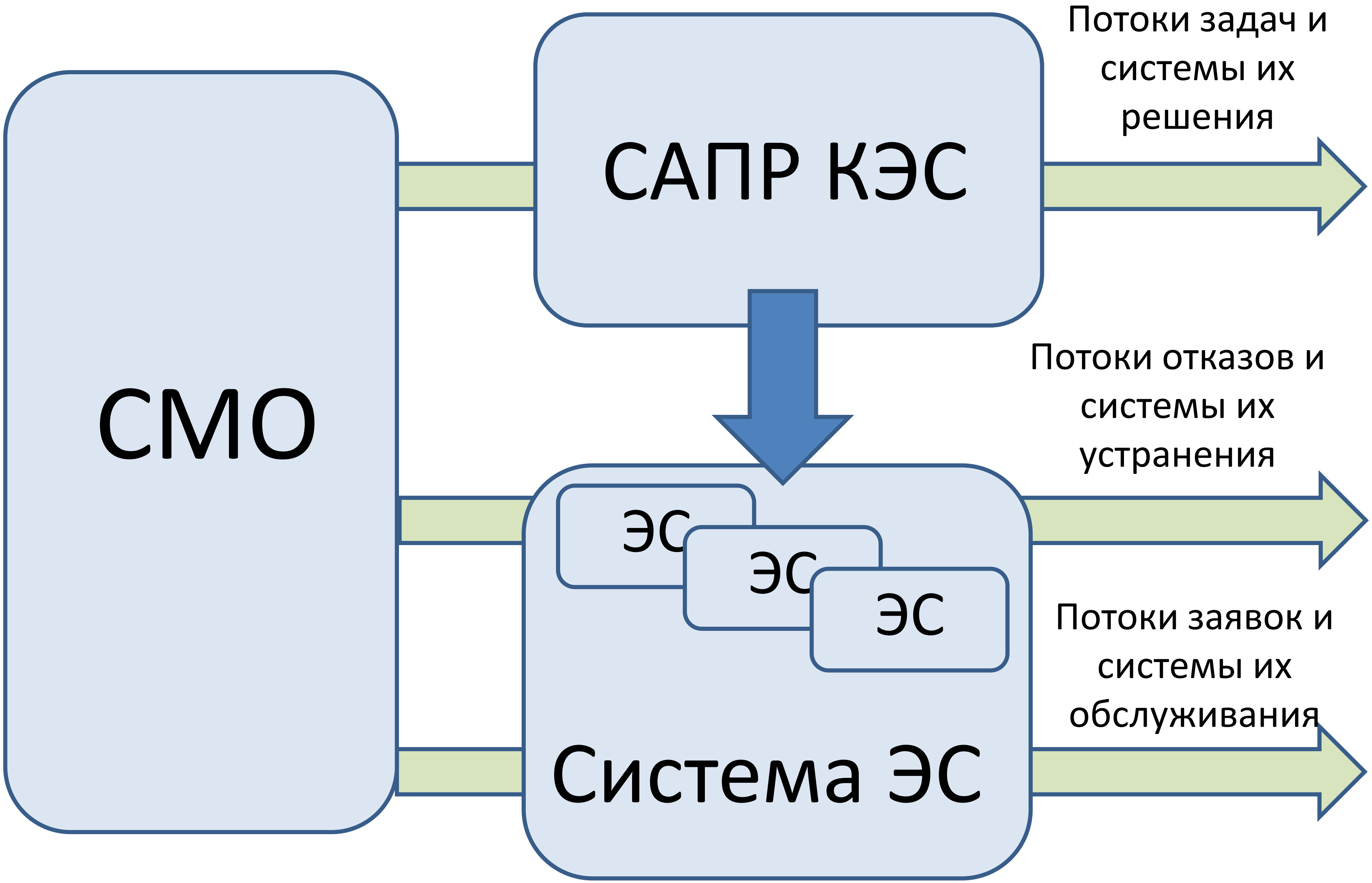
В 1918 году **Т.Энгсет** обобщил результаты А. К. Эрланга на случай обслуживания полнодоступным пучком потока вызовов от конечного числа источников нагрузки.

В 1933 году советский математик **А. Н. Колмогоров** выполнил свою классическую работу по аксиоматическому обоснованию теории вероятностей, в которой идеи А. К. Эрланга были увязаны с марковскими случайными процессами.

В этот же период появились первые работы **А. Я. Хинчина** по исследованию систем массового обслуживания с ожиданием.

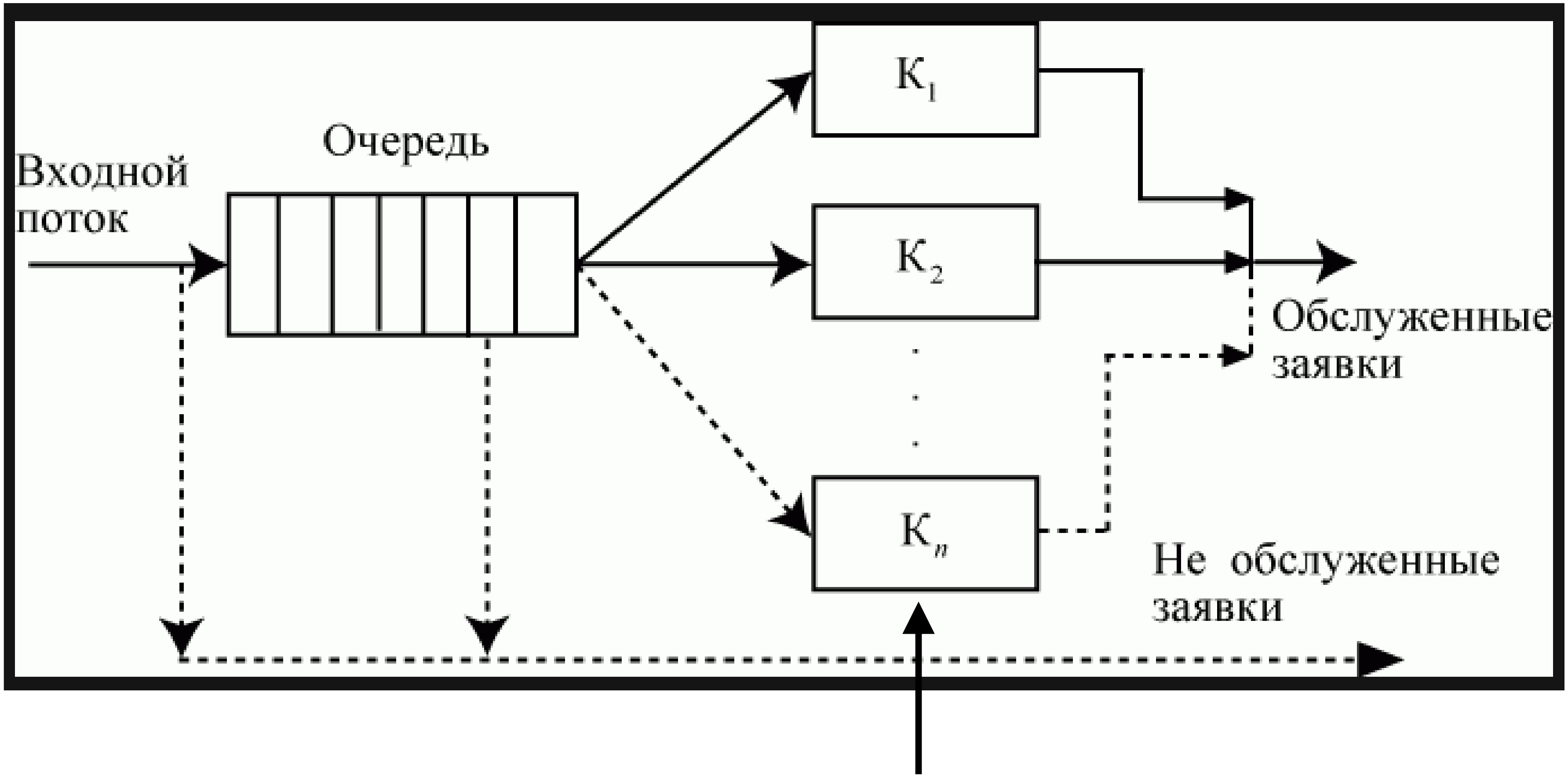
В 1943 году шведский ученый **К. Пальм** обобщил результаты А. К. Эрланга на случай обслуживания потока с ограниченным последействием, и получил важные результаты по изучению колебания телефонной нагрузки.

В 1964 году американский ученый **Л. Клейнрок** разработал основные принципы пакетной коммуникации, которые легли в основу современной технологии Интернета


СМО
САПР КЭС
Потоки задач и системы их решения
ЭС
ЭС
ЭС
Система ЭС
Потоки отказов и системы их устранения
Потоки заявок и системы их обслуживания

Сеть связи – как система (сеть) массового обслуживания

Элементы типовой модели системы массового обслуживания (СМО)



Обслуживающие приборы

Основные понятия

- ***Заявка (требование, запрос, вызов, клиент, сообщение, пакет)*** – объект, поступающий в СМО и требующий обслуживания в обслуживающем приборе. Совокупность заявок, распределенных во времени, образуют ***входной поток заявок***.
- ***Обслуживающий прибор*** или просто ***прибор (устройство, канал, линия)*** – элемент СМО, функцией которого является обслуживание заявок.

Основные понятия

- ***Обслуживание*** – задержка заявки на некоторое время в обслуживающем приборе.
- ***Длительность обслуживания*** – время задержки (обслуживания) заявки в приборе.
- ***Накопитель (буфер)*** – совокупность мест для ожидания заявок перед обслуживающим прибором.
- Количество мест для ожидания определяет ***ёмкость накопителя***.

Основные понятия

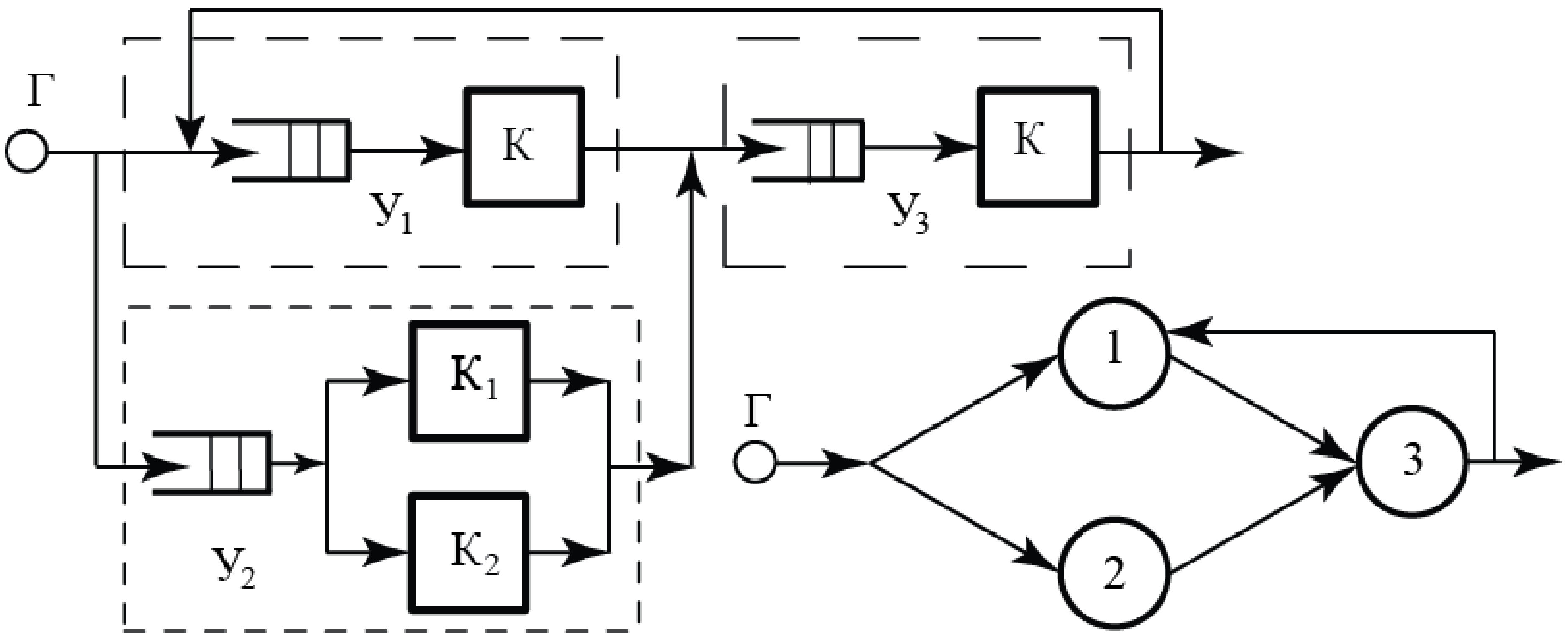
- Заявки, находящиеся в накопителе (буфере) и ожидающие обслуживания, образуют ***очередь*** заявок.
- Количество заявок, ожидающих обслуживания в накопителе (буфере), определяет ***длину очереди***.
- ***Дисциплина буферизации (дисциплина постановки в очередь)*** – правило занесения поступающих заявок в накопитель (буфер).
- ***Дисциплина обслуживания*** – правило выбора заявок из очереди для обслуживания в приборе.
- ***Приоритет*** – преимущественное право на занесение (в накопитель, буфер) или выбор из очереди (для обслуживания в приборе) заявок одного класса по отношению к заявкам других классов.

Основные понятия

При рассмотрении СМО, будем использовать следующие **предположения**:

- поступившая в систему заявка ***мгновенно*** попадает на обслуживание, если прибор свободен;
- в приборе на обслуживании в каждый момент времени может находиться только ***одна*** заявка;
- после завершения обслуживания какой-либо заявки в приборе очередная заявка выбирается на обслуживание из очереди мгновенно, то есть, другими словами, прибор ***не простаивает***, если в очереди есть хотя бы одна заявка.

Сеть массового обслуживания (СеМО) – совокупность взаимосвязанных СМО, в среде которых циркулируют заявки



Основные понятия

Основными элементами СеМО являются узлы (У) и источники (генераторы) заявок (Г).

- ***Узел*** сети представляет собой систему массового обслуживания.
- ***Источник*** – генератор заявок, поступающих в сеть и требующих определенных этапов обслуживания в узлах сети.

Для упрощенного изображения СеМО используется граф СеМО.

- ***Граф СеМО*** – ориентированный граф, вершины которого соответствуют узлам СеМО, а дуги отображают переходы заявок между узлами (рис. 2.2,б).

Модель потока дискретных событий в непрерывном времени

$\lambda = 1 / \tau_{вх}$

Интенсивность входного потока

$\mu = 1 / \tau_{обс}$

Интенсивность обслуживания

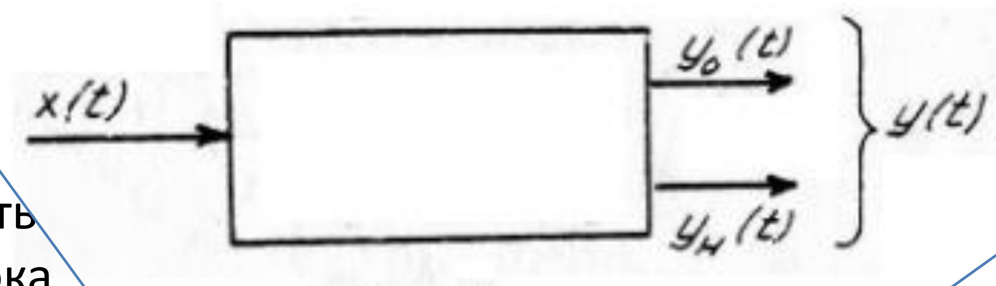


Рис. 1

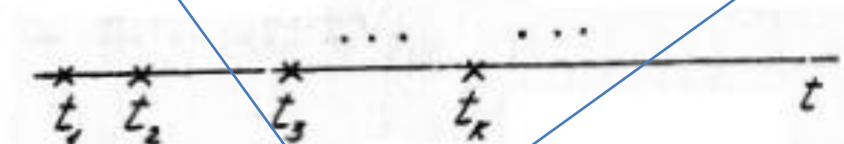


Рис. 2

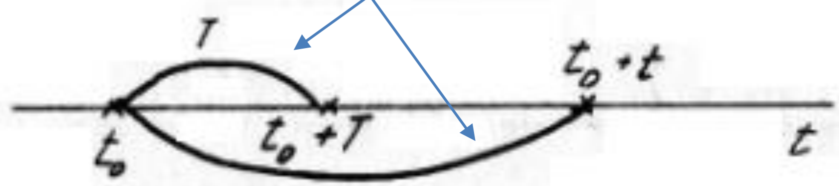


Рис. 3

Входной (входящий) поток заявок

Совокупность событий, распределенных во времени, называется ***потоком***. Если событие заключается в появлении заявок, имеем ***поток заявок***.

Для описания потока заявок, в общем случае, необходимо задать интервалы времени $\tau = t_k - t_{k-1}$ между соседними моментами поступления заявок ($k = \overline{1, 2}, \ldots, t_0 = 0$ – начальный момент времени).

Основной характеристикой потока заявок является его ***интенсивность*** $\boldsymbol{\lambda}$ – среднее число заявок, проходящих через некоторую границу за единицу времени. Величина $\tau = 1/\lambda$ определяет ***средний интервал времени между двумя последовательными заявками***.

Каналы (приборы) обслуживания

Согласно наличию одного или нескольких каналов обслуживания СМО называют одноканальными или многоканальными.

Многоканальные СМО могут состоять из однотипных или разнотипных приборов.

Основная характеристика канала — ***длительность обслуживания*** – время нахождения заявки в приборе – в общем случае величина случайная.

Основной характеристикой канала обслуживания заявок является ***интенсивность обслуживания*** μ – среднее число заявок, обслуживаемых за единицу времени. Величина $\tau = 1/\mu$ определяет ***среднюю длительность обслуживания одной заявки***.

Вопрос 2

Классификация систем массового обслуживания

Классификация систем массового обслуживания

Классификация моделей массового обслуживания

При моделировании реальных систем с дискретным характером функционирования широкое применение находят базовые модели в виде СМО, которые могут быть классифицированы (рис.2.5):

- **по числу мест в накопителе**;
- **по числу обслуживающих приборов**;
- **по количеству классов заявок, поступающих в СМО**.

Классификация моделей массового обслуживания

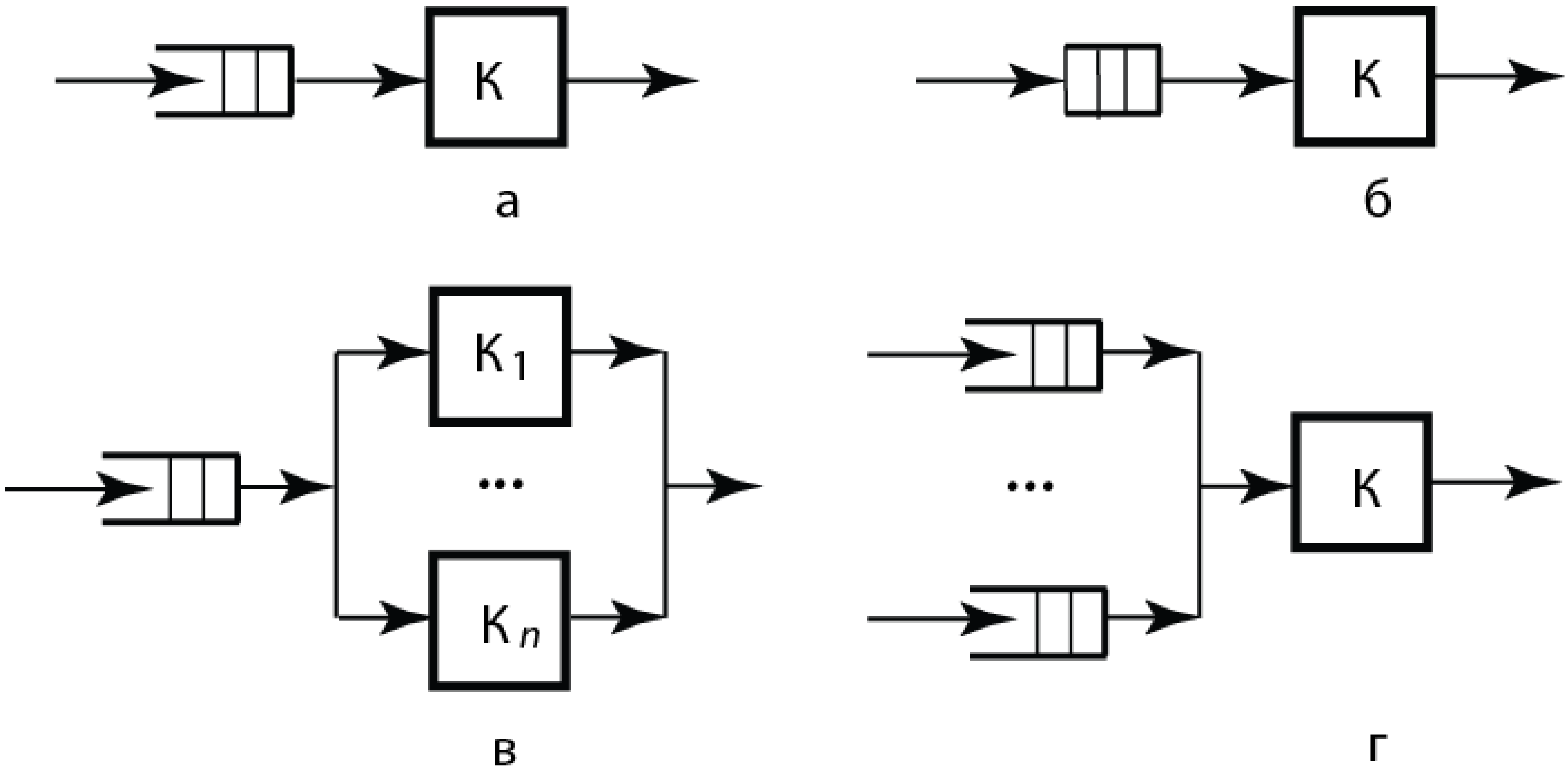


Рис. 2.5. Базовые модели СМО

Классификация моделей массового обслуживания

По ***числу мест в накопителе*** СМО делятся на системы:

- **без накопителя (*СМО с отказами*)**, в которых заявка, поступившая в систему и заставшая все обслуживающие приборы занятыми обслуживанием более высокоприоритетных заявок, получает отказ и теряется;
- **с накопителем ограниченной ёмкости (*СМО с потерями*)** (рис. 2.5, б), в которых поступившая заявка теряется, если она застаёт накопитель заполненным полностью;
- **системы с накопителем неограниченной ёмкости (*СМО без потерь*)** (рис. 2.5, а, в, г).

Классификация моделей массового обслуживания

По ***количеству обслуживающих приборов*** СМО делятся на:

- **одноканальные** (рис. 2.5, а, б, г), содержащие один прибор (канал) К;
- **многоканальные** (рис. 2.5, в), содержащие К – более одного обслуживающих приборов.

В многоканальных СМО обычно предполагается, что все приборы идентичны и равнодоступны для любой заявки.

Классификация моделей массового обслуживания

По ***количеству классов (типов) заявок***, поступающих в СМО, различают системы:

- **с однородным потоком** заявок (рис. 2.5, а, б, в);
- **с неоднородным потоком** заявок (рис. 2.5, г).

В СМО, представляющей собой абстрактную математическую модель, *заявки относятся к разным классам* в том случае, если они в моделируемой реальной системе различаются хотя бы одним из следующих факторов:

- ***длительностью обслуживания;***
- ***приоритетами***.

Классификация моделей массового обслуживания

Рис. 2.6. Вариант классификации СеМО

Классификация моделей массового обслуживания

В зависимости от ***характера процессов поступления и обслуживания заявок*** в сети СеМО делятся на:

- **стохастические**, в которых интервалы времени между поступающими заявками и/или длительности их обслуживания в узлах представляют собой случайные величины, описываемые соответствующими законами распределений;
- **детерминированные**, в которых интервалы времени между поступающими заявками и длительности их обслуживания в узлах являются детерминированными величинами.

Классификация моделей массового обслуживания

По ***виду зависимостей, связывающих интенсивности потоков заявок в разных узлах***, СеМО делятся на:

- **линейные**, если эти зависимости линейные;
- **нелинейные**, если эти зависимости являются нелинейными.

В *линейных* СеМО интенсивность потока заявок в узел j связана с интенсивностью потока заявок в узел i линейной зависимостью: $\lambda_j = \alpha_{ij} \cdot \lambda_i$, где α_{ij} – коэффициент пропорциональности, показывающий, во сколько раз отличаются интенсивности потоков заявок в узел j и в узел i, $i, j = \overline{1, n}$.

Классификация моделей массового обслуживания

В ***нелинейных*** СеМО интенсивности потоков заявок в узлах связаны нелинейными зависимостями, что значительно усложняет их исследование.

Нелинейность СеМО может быть обусловлена:

- ***потерей заявок*** в сети, например из-за ограниченной емкости накопителей в узлах;
- ***размножением заявок*** в сети, заключающимся, например, в формировании нескольких новых заявок после завершения обслуживания некоторой заявки в одном из узлов сети.

Таким образом, **СеМО является линейной, если в ней заявки не размножаются и не теряются**.

Классификация моделей массового обслуживания

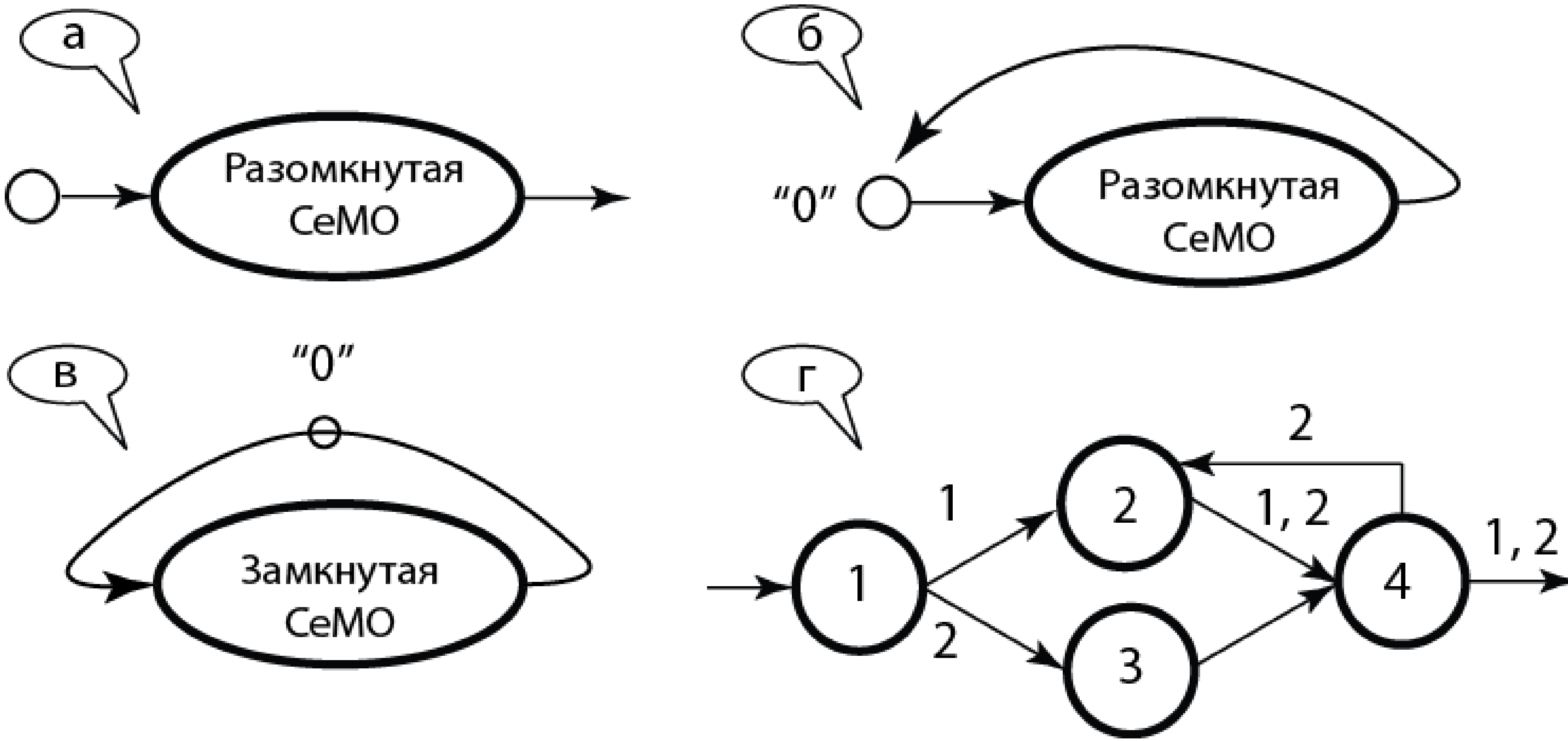


Рис. 2.7. Виды СеМО по числу циркулирующих заявок

Классификация моделей массового обслуживания

Разомкнутая (открытая) СеМО содержит один или несколько ***внешних независимых источников*** заявок, которые генерируют заявки в сеть независимо от числа заявок, находящихся в сети (рис.2.7, а). В разомкнутой СеМО одновременно может находиться ***любое число заявок***, в том числе, и сколь угодно большое, то есть от 0 до бесконечности.

С разомкнутой СеМО связана внешняя среда, из которой поступают заявки в сеть и в которую они возвращаются после обслуживания в сети. Внешняя среда обозначается обычно как нулевой узел «0», рис. 2.7,б.

Классификация моделей массового обслуживания

Замкнутая (закрытая) СеМО не содержит ***независимых внешних*** источников заявок и характеризуется тем, что в ней циркулирует ***постоянное число заявок М*** (рис. 2.7,в).

Замкнуто-разомкнутая СеМО (комбинированная) представляет собой комбинацию замкнутой СеМО и разомкнутой СеМО, в которую, кроме постоянно циркулирующих в сети M^* заявок, из внешнего независимого источника поступают заявки такого же или другого класса, при этом суммарное число заявок в сети $M \geq M^*$.

Классификация моделей массового обслуживания

По ***типу циркулирующих заявок*** различают СеМО:

- **однородные**, в которых циркулирует один класс заявок (однородный поток заявок);
- **неоднородные**, в которых циркулирует несколько классов заявок (неоднородный поток заявок), различающихся хотя бы одним из следующих факторов: ***длительностями обслуживания*** в узлах; ***приоритетами; маршрутами***.

Маршруты заявок разных классов задаются путем указания номеров классов заявок на соответствующих дугах сети (рис. 2.7, г).

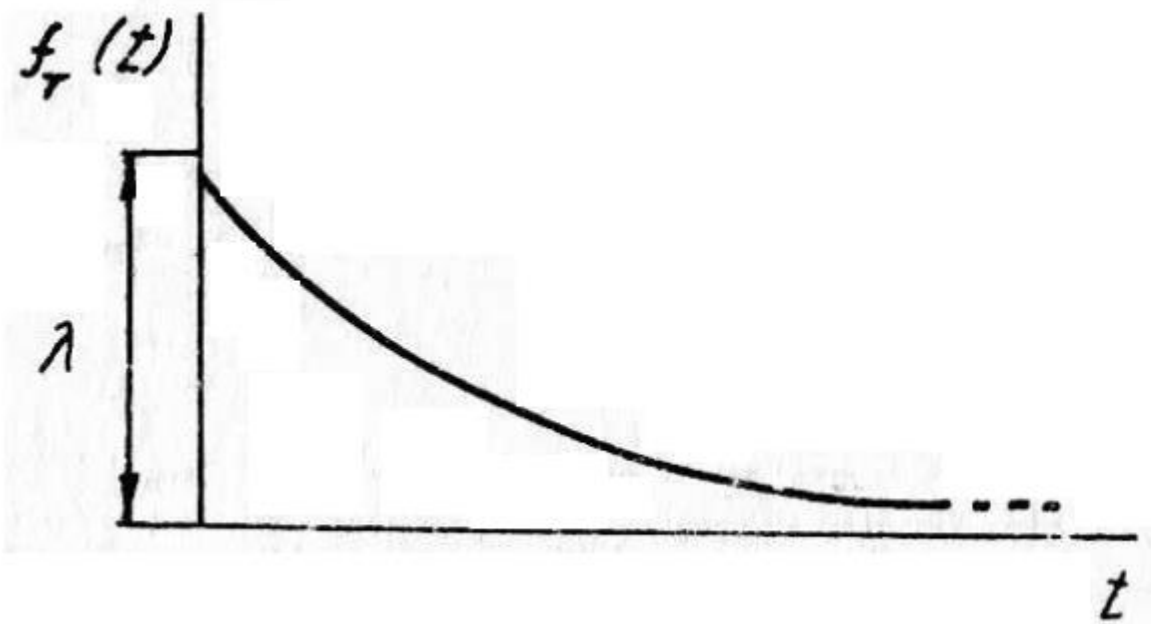
При моделировании СМО к наиболее важным свойствами потоков заявок относят следующие.

1. **Стационарность**. Поток событий считается стационарным, если вероятность попадания определенного числа событий на интервал времени длиной τ зависит только от длины этого интервала и не зависит от того, в каком месте на временной оси расположен этот участок.
2. **Отсутствие последействия**. Поток событий называется потоком без последействия, если для любых неперекрывающихся интервалов времени число событий, попадающих на один временной отрезок, не зависит от числа событий, попадающих на другие отрезки времени.
3. **Ординарность**. Поток событий называется ординарным, если вероятность попадания двух или более событий на элементарный (малый) отрезок времени Δt пренебрежимо мала по сравнению с вероятностью попадания на этот временной интервал одного события.

Поток событий, обладающий тремя перечисленными свойствами, называется **простейшим**, или стационарным **пуассоновским**, поскольку для простейшего потока число событий, попадающих на любой фиксированный интервал времени, распределено по закону Пуассона.

$$p_k(\tau) = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau} \qquad F_T(t) = 1 - e^{-\lambda t} \quad (t > 0)$$

$$f_T(t) = \frac{\mathrm{d}F_T(t)}{\mathrm{d}t} = \lambda\, \mathrm{e}^{-\lambda t}$$



В теории массового обслуживания используют и другие модели потоков Пальма, Эрланга и т. д.

Ординарный поток однородных событий называется потоком с ограниченным последействием (потоком Пальма), если промежутки времени между последовательными событиями независимы. Простейший поток является частным случаем потока Пальма, когда независимые промежутки времени между соседними заявками распределены по показательному закону.

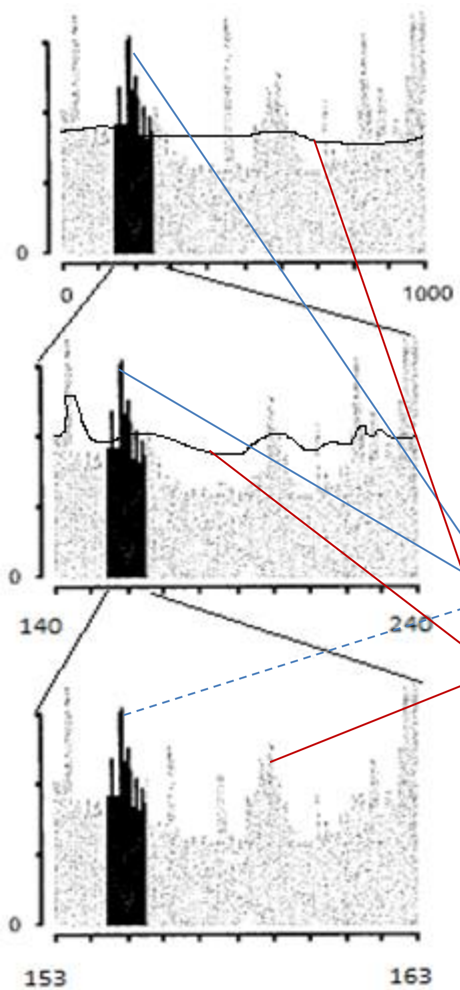
К потокам с ограниченным последействием относятся также потоки Эрланга, образуемые «просеиванием» простейшего потока. Если из простейшего потока исключить каждое второе требование, то оставшиеся требования образуют поток Эрланга первого порядка. Поток Эрланга k-го порядка получается при сохранении в исходном простейшем потоке каждого $(k+1)$-го требования и исключении остальных. Для потока Эрланга k-го порядка плотность распределения интервала между соседними заявками имеет вид

$$f_T(t) = \frac{\lambda(\lambda t)^k}{k!} e^{-\lambda t}, \qquad t > 0.$$

Поток Эрланга любого порядка представляет собой частный случай потока Пальма.

Реальные потоки, образуемые суммированием потоков от многих источников, оказываются близкими к простейшему.

Самоподобный (фрактальный) трафик, возникающий в мультисервисных сетях



Основной параметр самоподобного трафика H (***параметр Хёрста***) определяется из следующего приближенного равенства:

$$\log\left(\frac{R}{S}\right) \approx H \log\left(\frac{N}{2}\right)$$

где S – выборочная дисперсия, N – объем выборки, R - изменчивость:

$R_N = \max\limits_{1<j<N} D_j - \min\limits_{1<j<N} D_j$, где D – интегральное отклонение: $D_j = \sum\limits_{k=1}^{j} X_k - jM$,

где M – выборочное среднее

Взаимосвязь параметра Хёрста H с параметром β убывания дисперсии $S(X^{(n)})$ агрегированного процесса $X^{(n)}$ длиной n:

$$\beta = (1 - H) \cdot 2$$

где $X^{(n)} = \{X_k^{(n)} : k = 1, 2, 3, \ldots\}$; $X_k^{(n)} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_{(k-1)n+i}$.

Самоподобный трафик

Традиционный трафик

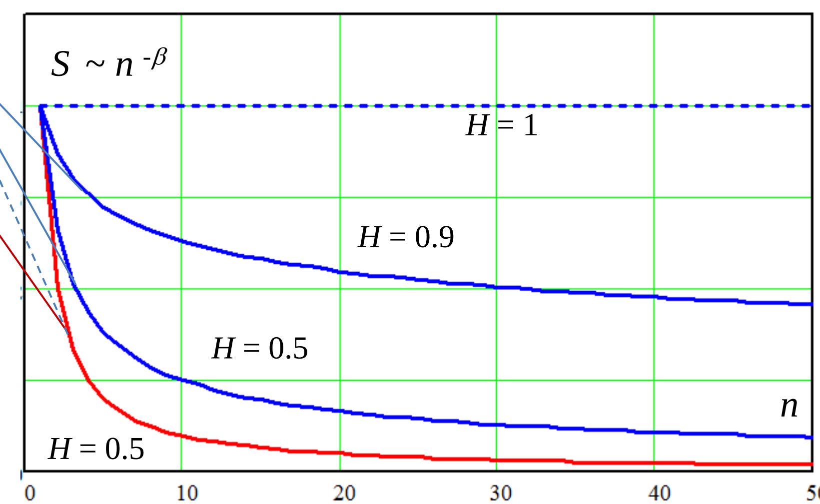


Рисунок 1 – Пояснение самоподобия (фрактальности): характеристики традиционного трафика при усреднении на длительных интервалах изменяются (сглаживаются), а самоподобного – остаются сходными (самоподобными).

Рисунок 2 – Зависимость дисперсии S агрегированного процесса от его длины n

Марковские модели систем массового обслуживания

Случайный процесс называется марковским, если его поведение в будущем определяется только настоящим состоянием и не зависит от того, когда и каким именно образом система перешла в текущее состояние.

Если динамика системы при выполнении ею своих задач может быть представлена в виде случайного процесса смены состояний и при этом вероятностные характеристики перехода из одного состояния в другое не будут зависеть от развития процесса смены состояний в прошлом, то для моделирования такой системы может быть применен математический аппарат марковских случайных процессов

В зависимости от характера изменения состояний и времени все многообразие марковских процессов подразделяется на следующие четыре типа:

- марковские процессы с дискретными состояниями и дискретным временем, называемые дискретными марковскими цепями или цепями Маркова;
- марковские процессы с дискретными состояниями и непрерывным временем, называемые непрерывными марковскими цепями;
- марковские процессы с непрерывными состояниями и дискретным временем;
- марковские процессы с непрерывными состояниями и непрерывным временем.

В практике моделирования сложных систем нашли широкое применение два типа из перечисленных выше марковских процессов, а именно дискретные и непрерывные марковские цепи.

Модели на основе дискретных марковских цепей

Задача разработки модели на основе дискретной марковской цепи формулируется следующим образом

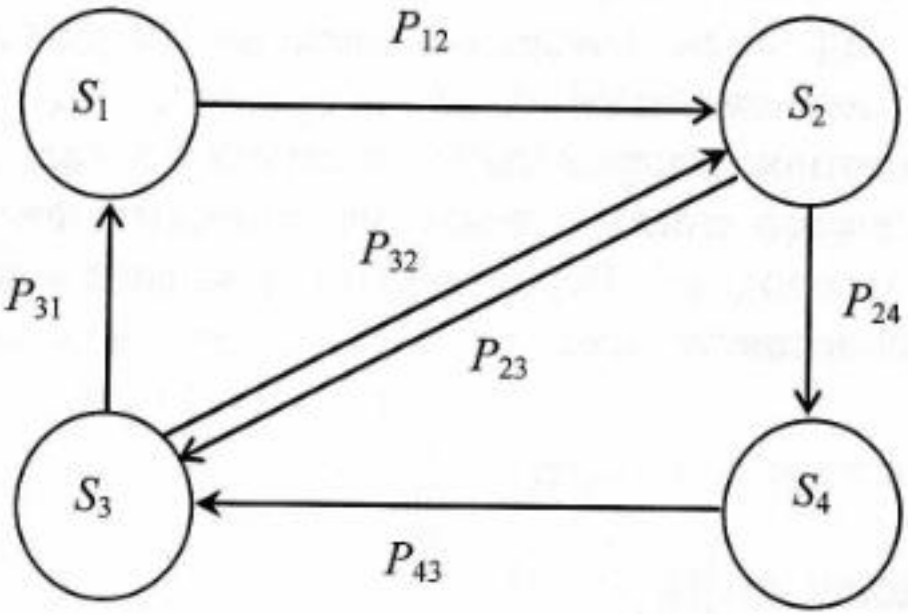
Пусть система S имеет n возможных состояний $S_1, S_2, \ldots, S_i, S_j, \ldots, S_n$, смена которых происходит мгновенно в определенные заранее известные моменты времени 1, 2,…, к,… K. Известны вероятности перехода P_{ij} системы S за один шаг из состояния S_i в состояние S_j, не изменяющиеся от шага к шагу (то есть цепь однородная). Вероятности P_{ij} заданы в виде матрицы переходных вероятностей:

$$\left[P_{ij}\right] = \begin{bmatrix} P_{11} & P_{12} & \ldots & P_{1n} \\ P_{21} & P_{22} & \ldots & P_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ P_{n1} & P_{n2} & \ldots & P_{nn} \end{bmatrix},$$

где для всех i, j справедливы утверждения:

$$P_{ij} = const;$$

$$\sum_{j=1}^{n} P_{ij} = 1.$$



Пример графа состояний и переходов системы

Математическая модель динамики рассматриваемой однородной марковской цепи получается на основе формулы полной вероятности в виде рекуррентного соотношения:

$$P_j(k) = \sum_{i=1}^{n} P_i(k-1) \cdot P_{ij} \,,$$

Модели на основе непрерывных марковских цепей. Уравнения Колмогорова – Чепмена.

В непрерывных марковских цепях использование переходных вероятностей P_{ij} становится невозможным. Это объясняется тем, что вероятность перехода системы из одного состояния в другое в конкретный момент времени t для непрерывной случайной величины равна нулю. Поэтому вместо переходных вероятностей вводятся в рассмотрение плотности распределения вероятностей переходов λ_{ij}, задаваемые равенством :

$$\lambda_{ij} = \lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t}, \tag{1.10}$$

где $P_{ij}(\Delta t)$ – вероятность того, что система, пребывая в момент времени t в состоянии S_i, за время Δt перейдет в S_j.

С точностью до малых второго порядка из приведенной формулы (1.10) можно получить следующее выражение:

$$P_{ij}(\Delta t) \approx \lambda_{ij} \Delta t. \tag{1.11}$$

Пример графа непрерывной марковской цепи с четырьмя состояниями S_1, S_2, S_3, S_4 приведен на рисунке 1.2.

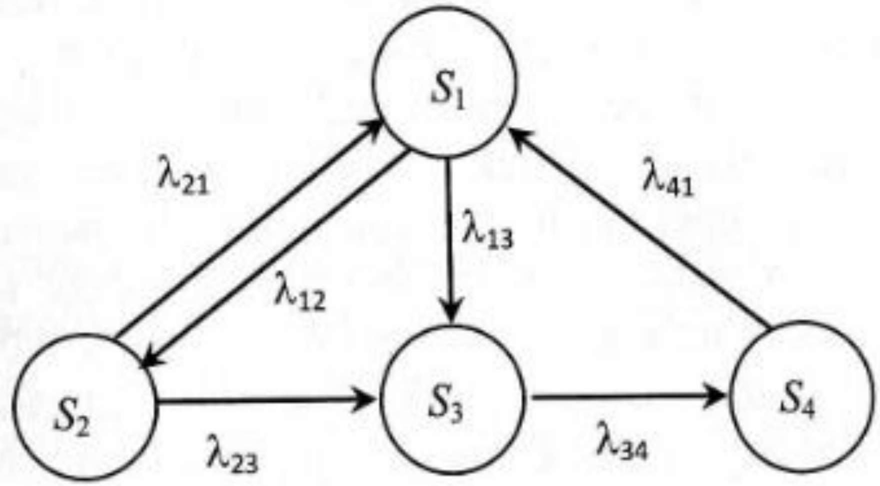


Рисунок 1.2 – Пример графа состояний непрерывной марковской цепи

Вероятности пребывания системы в любом j-м состоянии в момент t удовлетворяют уравнениям Колмогорова-Чепмена, которые являются обыкновенными дифференциальными уравнениями первого порядка с постоянными коэффициентами.

Составление уравнений Колмогорова-Чепмена производится с помощью следующего правила

Левая часть уравнения для каждого состояния представляет собой первую производную вероятности этого состояния по времени, а правая часть образуется сложением отрицательного произведения вероятности этого состояния на сумму интенсивностей всех возможных переходов из него в другие состояния и суммы произведений вероятностей всех предыдущих состояний на интенсивности перехода из них в рассматриваемое состояние. Общее число дифференциальных уравнений равно числу состояний системы.

Общий вид уравнений приведен ниже.

$$\frac{dP_j(t)}{dt} = -P_j(t)\sum_{k=1}^{n}\lambda_{jk} + \sum_{i=1}^{n}P_i(t)\lambda_{ij}, \quad j = 1,2,\ldots,n,$$

где $P_j(t)$ и $P_i(t)$ – вероятность j-го и предшествующего ему i-го состояния;

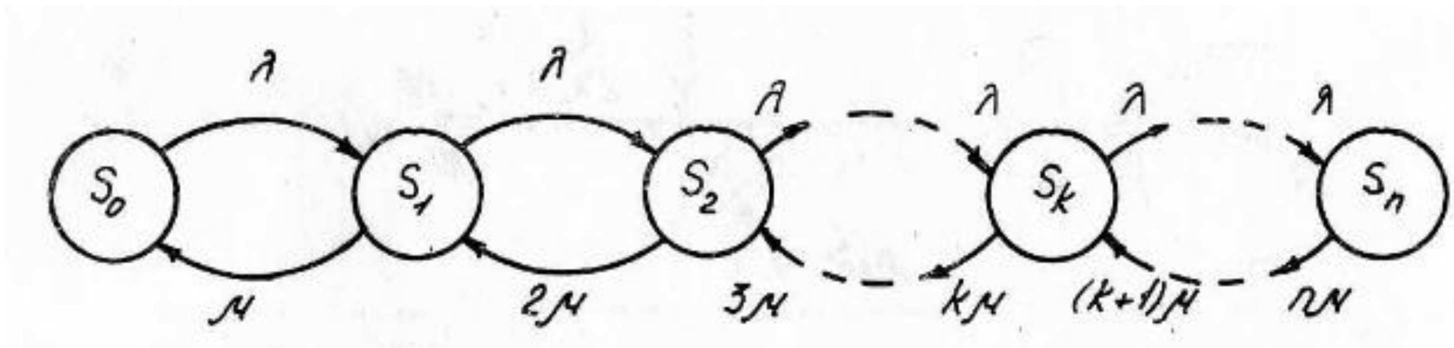
λ_{ij} и λ_{jk} – интенсивности потоков событий, переводящие систему из состояния i в состояние j и из состояния j в k соответственно.

Уравнения для определения предельных вероятностей нетрудно получить из системы уравнений Колмогорова-Чепмена, с учетом того, что в установившемся режиме вероятности P_j не изменяются. Тогда левые части уравнений Колмогорова - Чепмена обратятся в нуль. Модель однородной непрерывной марковской цепи будет представлена системой линейных алгебраических уравнений

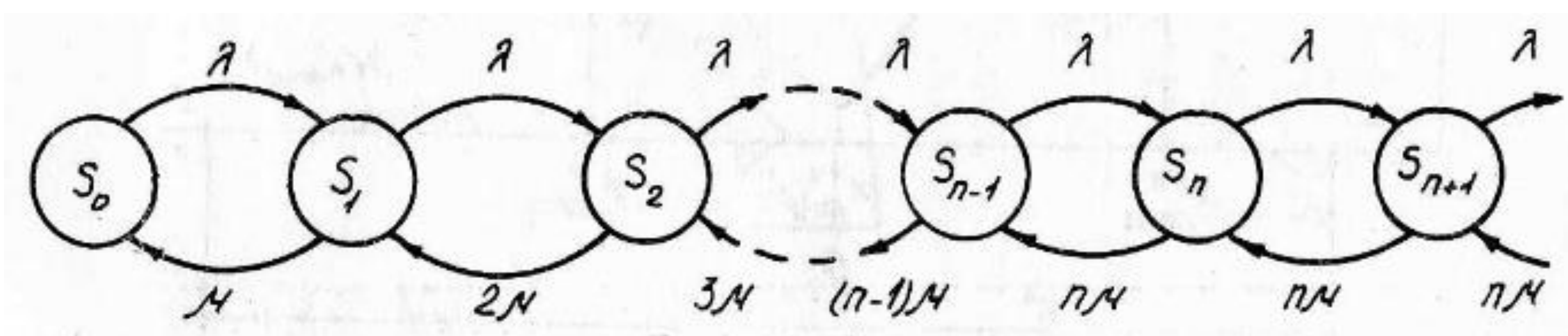
$$0 = -P_j \sum_{k=1}^{n} \lambda_{jk} + \sum_{i=1}^{n} P_i \lambda_{ij}, \quad j = 1,2,\dots,n.$$

По известным значениям предельных вероятностей можно определить необходимые характеристики реальной системы в зависимости от интенсивностей λ_{ij}.

Пример марковской модели многоканальной СМО с отказами



Пример марковской модели многоканальной СМО с ожиданием



Одной из форм классификации систем массового обслуживания является кодовая (символьная) классификация **Д.Кендалла**.

При этой классификации характеристику системы записывают в виде трех, четырех или пяти символов, например A / B / S, где A — тип распределения входящего потока требований, B — тип распределения времени обслуживания, S — число каналов обслуживания.

Для экспоненциального распределения используют символ M (Markovian), для любого (произвольного) распределения — символ G (General). Регулярный поток обозначают буквой D. Распределение Парето — символом P, самоподобный трафик — буквами fbm (фрактальное броуновское движение) и т.д.

Например, запись M / M / 3 означает, что входящий поток требований пуассоновский (простейший), время обслуживания распределено по экспоненциальному закону, в системе имеется три канала обслуживания.

Четвертый символ указывает допустимую длину очереди, а пятый — порядок отбора (приоритета) требований

Вопрос 3

Методы анализа и синтеза систем массового обслуживания. Основные характеристики и методы их расчета

Основная цель теории телетрафика как одной из базовых ветвей теории массового обслуживания заключается **в разработке методов оценки качества** функционирования электронных систем распределения информации, т.е., построение математических моделей, более или менее адекватно отображающих реальные системы распределения и обработки информации, **что позволяет экономично проектировать системы и сети связи, а также их элементы (электронные средства)** при заданном качестве обслуживания.

Задачи теории телетрафика

анализ;
синтез;
оптимизация.

На первом месте стоят **задачи анализа**, т.е., отыскание зависимостей и значений величин, характеризующих качество обслуживания, от характеристик и параметров входящего потока вызовов, схемы и дисциплины обслуживания. Эти задачи в начальный период развития телефонной техники были более актуальными, чем задачи синтеза, и решались, как правило, с помощью теории вероятностей.

Развитие координатной, квазиэлектронной и электронной (цифровой) коммутационной техники поставило перед теорией телетрафика сложные вероятностно-комбинаторные **задачи синтеза**, в которых требуется определить структурные параметры коммутационных систем при заданных потоках, дисциплине и качестве обслуживания.

Близкими к задачам анализа и синтеза являются **задачи оптимизации**. Эти задачи при проектировании систем распределения информации формулируются следующим образом: определить такие значения структурных параметров коммутационной системы (алгоритмы функционирования), для которых:
при заданных потоках, качестве и дисциплине обслуживания стоимость или объем оборудования системы распределения информации минимальны и
при заданных потоках, дисциплине обслуживания и стоимости качественные показатели функционирования системы распределения информации оптимальны.

Характеристики СМО с однородным потоком заявок

- ***нагрузка системы***;
- ***коэффициент простоя системы***;
- ***вероятность потери заявок***;
- ***вероятность обслуживания заявки***, то есть вероятность того, что поступившая в систему заявка будет обслужена;
- ***производительность системы***, представляющая собой интенсивность потока обслуженных заявок, выходящих из системы;
- ***интенсивность потока потерянных (не обслуженных) заявок*** из-за ограниченной ёмкости накопителя;

Характеристики СМО с однородным потоком заявок

- ***среднее время ожидания заявок в очереди***;
- ***среднее время пребывания заявок в системе***, складывающееся из времени ожидания и времени обслуживания;
- ***средняя длина очереди заявок***;
- ***среднее число заявок в системе*** (в очереди и на обслуживании в приборе).

В некоторых системах, например в *СМО с накопителем неограниченной ёмкости*, неустановившийся режим функционирования может быть обусловлен ***перегрузкой системы***, когда интенсивность поступления заявок превышает интенсивность обслуживания, и система не справляется с возлагаемой на неё нагрузкой (**режим перегрузки**). При этом характеристики функционирования СМО с течением времени растут неограниченно. В частности, длина очереди перед прибором с течением времени становится всё больше и в пределе стремится к бесконечности.

Характеристики СМО с неоднородным потоком заявок

Для СМО с неоднородным потоком заявок, в которую поступают H классов заявок с интенсивностями $\mu_1, \ldots, \mu_H$ и средними длительностями обслуживания $t_{1\,обсл}, t_{2\,обсл}, \ldots, t_{H обсл}$, определяются две группы характеристик (показателей) обслуживания заявок:

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

Характеристики по каждому классу заявок $i = \overline{1, H}$ идентичны характеристикам СМО с однородным потоком.

Характеристики СеМО

Для описания ***линейных разомкнутых*** *и* ***замкнутых однородных экспоненциальных*** СеМО используется следующая совокупность параметров:

- ***число узлов*** в сети;
- ***число обслуживающих приборов*** в узлах сети;
- ***матрица вероятностей передач***, где p_{ij} – вероятность передачи заявки из узла i в узел j;
- ***интенсивность*** источника заявок, поступающих в разомкнутую СеМО, или *число заявок* M, циркулирующих в замкнутой СеМО;
- средние ***длительности обслуживания*** заявок в узлах сети.

Примеры формул расчета отдельных параметров различных СМО

$$p_k = \frac{\frac{a^k}{k!}}{\sum_{k=0}^{n} \frac{a^k}{k!}} \quad (0 \le k \le n). \qquad (1)$$

a = λ / μ

Формулы (1) называются формулами Эрланга. Они дают предельный закон распределения числа занятых каналов в зависимости от характеристик потока заявок и производительности системы обслуживания. Полагая в формуле (1') $k = n$, получим вероятность отказа (вероятность того, что поступившая заявка найдет все каналы занятыми):

$$P_{отк} = p_n = \frac{\frac{a^n}{n!}}{\sum_{k=0}^{n} \frac{a^k}{k!}}.$$

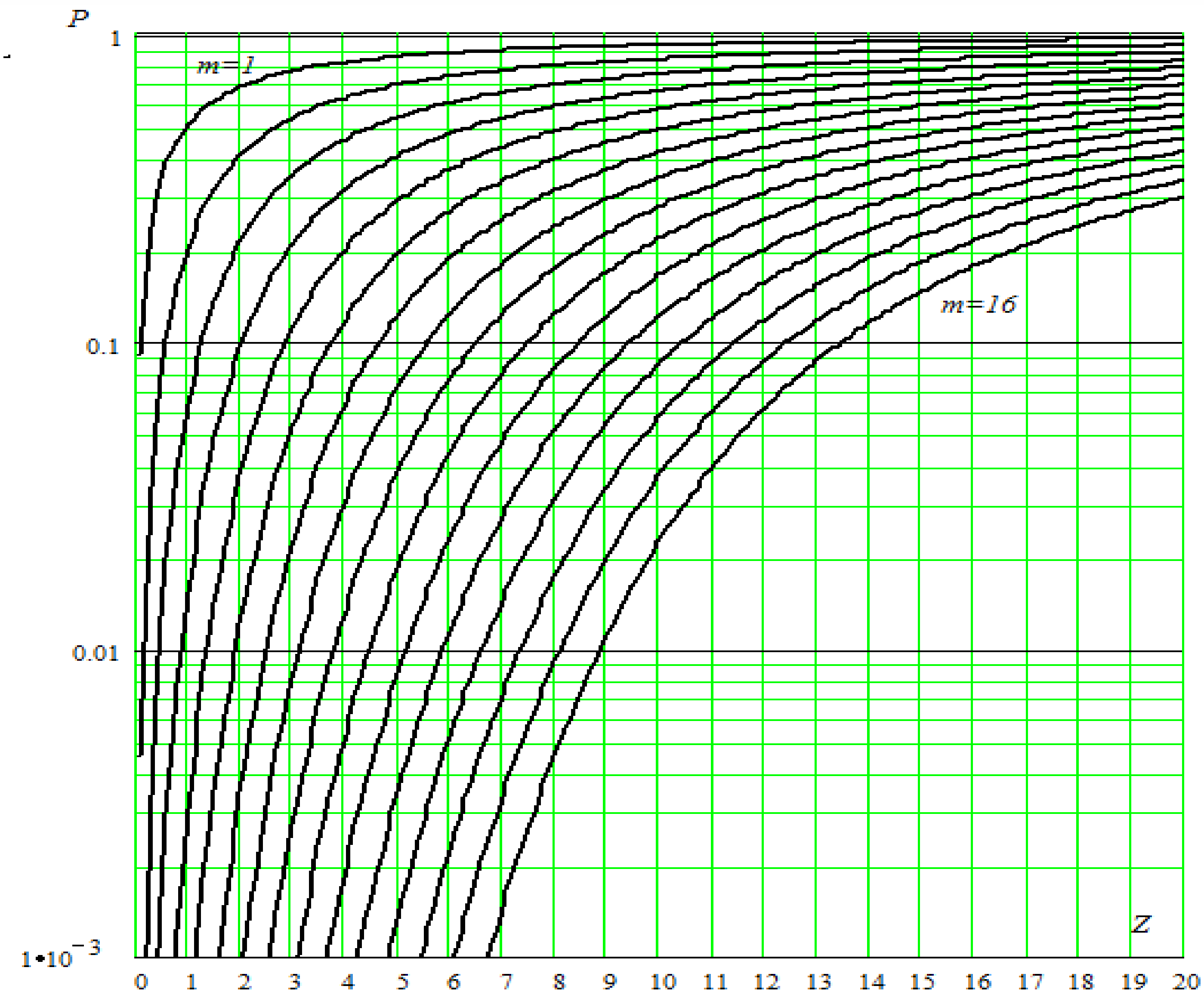
В частности, для одноканальной системы ($n = 1$)

$$P_{отк} = p_1 = \frac{a}{1+a},$$

а относительная пропускная способность

$$q = 1 - P_{отк} = \frac{1}{1+a}.$$

Пример графика, рассчитанного по формуле Эрланга



Задержка передачи отдельных пакетов τ_l по ЦК пути l определяется суммой четырех основных слагаемых:

$$\tau_l = \tau_{l.\text{расп}} + \tau_{l.\text{ком}} + \tau_{l.\text{кан}} + \tau_{l.\text{ож}} \ ,$$

где $\tau_{l.\text{расп}}$ – задержка распространения сигнала,

$\tau_{l.\text{ком}}$ – задержка коммутации,

$\tau_{l.\text{кан}}$ – задержка в канале передачи,

$\tau_{l.\text{ож}}$ – задержка из-за ожидания в очереди на передачу.

Задержка распространения сигнала $\tau_{l.\text{расп}}$ зависит в основном от физической длины пути $R_{l.\text{расп}}$ и от скорости $C_{l.\text{эмв}}$ распространения электромагнитных волн (ЭМВ) по используемой физической среде передачи: $\tau_{l.\text{расп}} = R_{l.\text{расп}} / C_{\text{эмв}}$.

Задержка коммутации $\tau_{l.\text{ком}}$ определяется производительностью устройства пакетной коммутации S_l, приведенной к входу ЦК рассматриваемого пути l. Если производительность S_l задана в пак/с с размером пакета, соответствующим размеру транспортного ПБД, то задержка коммутации будет равна $\tau_{l.\text{ком}} = 1/S_l$.

Задержка в канале передачи $\tau_{l.\text{кан}}$ определяется объемом (размером) V передаваемого ПБД (кадра, пакета) и скоростью передачи g_l в ЦК пути l.

$$\tau_{l.\text{кан}} = \frac{V}{g_l} .$$

Задержка из-за ожидания в очереди на передачу $\tau_{l.\text{ож}}$ существенно зависит от соотношения интенсивности протокольных блоков данных (ПБД) λ_l, поступающих на вход ЦК пути l, и интенсивности обслуживания ПБД в данном ЦК μ_l, равной обратной величине задержки в канале передачи $\tau_{l.\text{кан}}$ и прямо пропорциональной скорости передачи (канальному ресурсу) g_l в ЦК пути l:

$$\mu_l = \frac{1}{\tau_{l.\text{кан}}} = \frac{g_l}{V}.$$

Отношение интенсивности ПБД λ_l, на входе ЦК к интенсивности обслуживания ПБД в ЦК μ_l принято называть нагрузкой на данный ЦК $\rho_l = \lambda_l / \mu_l$.

Из-за случайности моментов прихода очередных ПБД время ожидания в очереди на передачу является случайной величиной, из-за чего в расчетах обычно используют различные вероятностные и усредненные характеристики указанного времени. Методы расчета вероятностных характеристик подобных систем с ожиданием обслуживания являются предметом теории телетрафика и систем массового обслуживания.

Формула Литтла

$\boldsymbol{N}$ (среднее число заявок в СМО) =

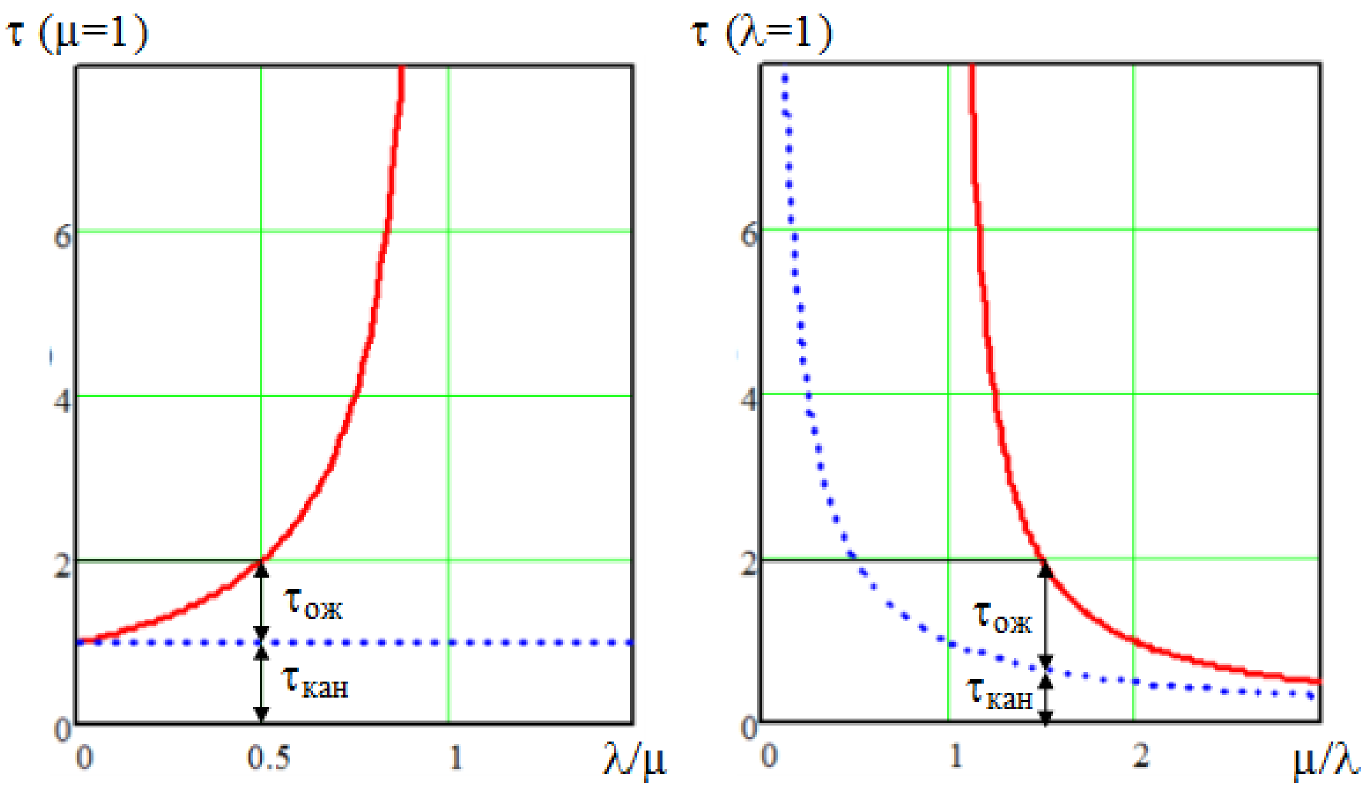
= $\boldsymbol{\lambda}$ (интенсивность заявок) * $\boldsymbol{T}$ (среднее время нахождения заявки в СМО)

Отсюда:

$$\boldsymbol{T = N / \lambda}$$

$$\tau_{кан} = \frac{1}{\mu}, \quad \tau_{ож} = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{1}{1-\rho},$$

$$\tau = \tau_{кан} + \tau_{ож} = \frac{1}{\mu-\lambda}$$



Зависимость относительной средней задержки τ в СМО М/М/1/∞ от соотношения входной интенсивности λ и интенсивности обслуживания μ

Расчет среднего времени задержки и вероятности потери пакетов

Модель СМО	Время задержки, τ	Вероятность потери пакетов, $p_{пот}$
M/M/1/w	$\tau = \frac{1}{\mu - \lambda}$	$p_{пот} = \frac{(1-\rho)\rho^{w}}{1-\rho^{w+1}}$, $\rho=\lambda/\mu$.
G/G/n/w	$\tau = \frac{1}{\lambda} \cdot \frac{2 \cdot \rho(1-\rho) + C_a^2 + \left(\frac{\rho}{n}\right)^2 \cdot C_b^2}{2 \cdot (1-\rho)}$ $\rho < 1$.	$p_{пот} \cong p_{i>n-1} \cdot \frac{(1-\frac{\rho}{n})}{1-\left(\frac{\rho}{n}\right)^{\frac{2 \cdot w}{C_a^2+C_b^2}+1}} \cdot \left(\frac{\rho}{n}\right)^{\frac{2 \cdot w}{C_a^2+C_b^2}}$.
fbm/M(D)/1/w	$\tau = \frac{1}{\lambda}\left(\frac{\rho}{m} \cdot \frac{\rho^{\frac{1}{2(1-H)}}}{(1-\rho)^{\frac{-H}{H(1-H)}}} + \rho\right)$, $\rho < 1$.	$p_{пот} = \bar{\Phi}\left(\frac{1}{\sqrt{a \cdot \rho}} \cdot \left(\frac{1-\rho}{H}\right)^{H} \cdot \left(\frac{m}{2} \cdot \frac{w+1}{1-H}\right)^{1-H}\right)$.
P/P/1/w	$\tau = \frac{\rho}{\mu \cdot (1-\rho)} \cdot \frac{C_a^2 + C_b^2}{2}$, $\rho < 1$.	$p_{пот} = \frac{(1-\rho) \cdot \rho}{1-\rho^{\frac{2 \cdot w}{C_a^2+C_b^2}+1}} \cdot \rho^{\frac{2 \cdot w}{C_a^2+C_b^2}}$.

Для учета влияния случайных всплесков задержки на несвоевременность доставки ПБД $p_{нсв}$ необходимо знать функцию распределения $F(x)=\Pr(\Delta t<x)$ случайных задержек Δt. При этом $p_{нсв}= 1-\Pr(\Delta t<\tau_{доп}) = 1- F(\tau_{доп})$. Однако точный вид функции распределения, как правило, неизвестен. Наиболее популярной и простой является экспоненциальная функция распределения $F_{exp}(x)$:

$$F_{\exp}(x) = 1 - \exp(-\frac{x}{\tau}) \qquad (1)$$

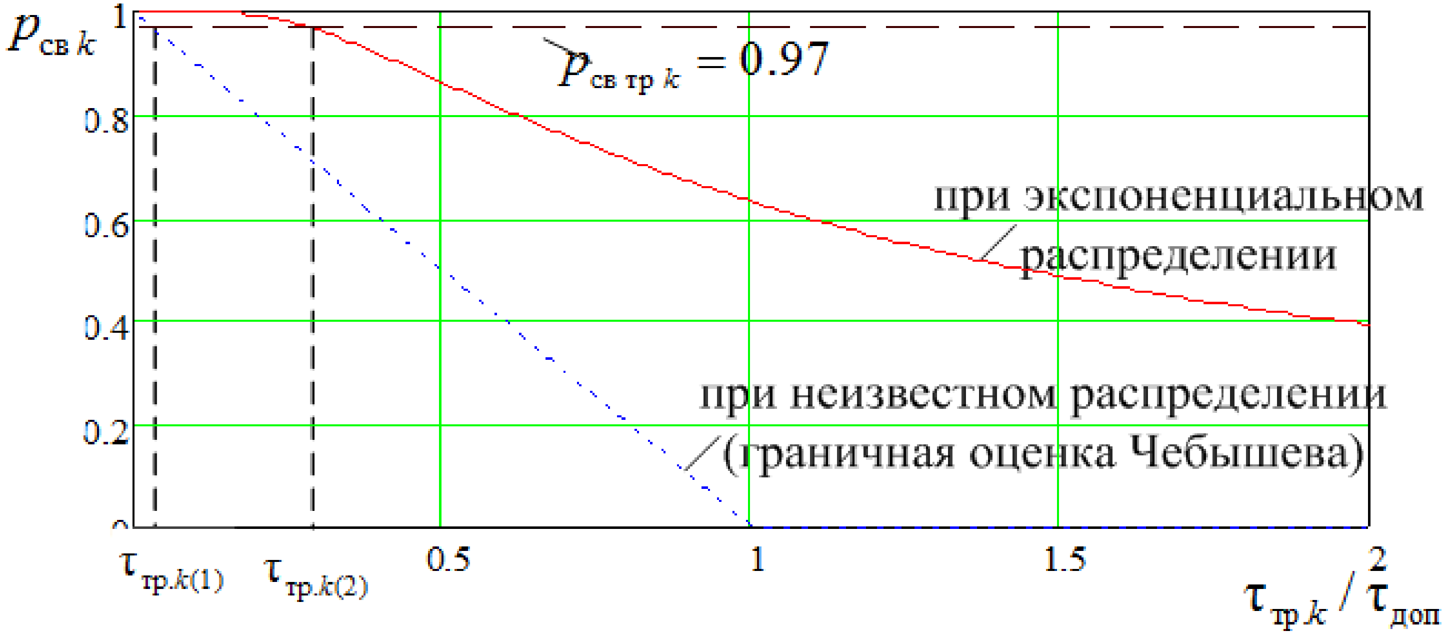
где τ – среднее значение (матожидание) случайной величины Δt.

Допуская справедливость экспоненциального распределения (1) можно рассчитать вероятность своевременной доставки пакетов $p_{св.k}$ при известной средней задержке τ_k по формуле

$$p_{св.k} = 1 - p_{нсв.k} = 1 - \exp(-\frac{\tau_{доп.k}}{\tau}) \qquad (2)$$

В случае, когда справедливость экспоненциального распределения (1) вызывает сомнение, можно оценить вероятность своевременной доставки пакетов $p_{св.k}$ по ее нижней границе Чебышева:

$$p_{св.k} = \begin{cases} 1 - \dfrac{\tau_k}{\tau_{доп.k}}, & \tau_k < \tau_{доп.k} \\ 0, & \tau_k \geq \tau_{доп.k} \end{cases} \qquad (3)$$



Пример пересчета требований к вероятности своевременной доставки за допустимое время в требования к средней задержке