ТЕОРИЯ ТЕЛЕТРАФИКА

Курс лекций для СДО

(Три части и три теста, итоговый зачет и курсовая работа)

Текст лекций

Часть 1

ВВЕДЕНИЕ

Связь развивается в соответствии с законами, по которым происходит передача и обработка информации. Передача и обработка информации включает ее декодирование, запоминание, хранение, извлечение, тиражирование и другие преобразования. Для выбора оптимальных вариантов в области связи важное значение имеют вопросы, которыми занимаются три теории, связанные с передачей и обработкой информации: «теория передачи информации» (ТПИ), «теория обработки информации» (ТОИ), «теория распределения информации» (ТРИ). Теория телетрафика (ТТТ) является важной частью ТРИ, которая разрабатывает методы построения, расчета и функционирования систем и сетей связи. Можно сказать, что ТРИ распределяет ресурсы между пользователями. В качестве ресурсов обычно рассматриваются каналы связи, средства коммутации, средства управления (память) и другие.

ТРИ имеет три основных раздела:

1) теория структур, которая занимается конструированием системы распределения информации;

2) теория управления, которая обеспечивает управление сетями, в том числе и распределение информационных потоков в сетях;

3) теория телетрафика, которая исследует вероятностные процессы функционирования цепей и узлов связи.

Понятие «телетрафик» - составное: «теле» - греч. «далеко», «трафик» - англ. «передача, нагрузка». Таком образом, телетрафик соответствует понятию, связанному с сообщениями в информационной сети, которые передаются на определенное расстояние. Пример понятия, с которым оперирует телетрафик: передача определенного объема данных.

ОСНОВНЫЕ ЗАДАЧИ ТЕОРИИ ТЕЛЕТРАФИКА

ТТТ решает 4 группы задач (1-3 — анализ, 4 — синтез). Рассмотрим их.

1) Полезный эффект в работе информационных сетей заключается в доставке информации. Всякое сообщение передается в информационную сеть в случайный или детерминированный (определенный) момент времени. Если рассматривать только моменты поступления сообщения, то можно говорить о потоках (пачках) вызовов (вызовы или требования, заявки). Потоки вызовов (ПВ) могут иметь различную структуру, различную интенсивность (частоту) и другие параметры. Большинство ПВ подвержены случайным колебаниям (колебаниям сложного характера). Так, например, потоки телефонных вызовов колеблются по дням недели, часам суток, имеют место случайные колебания по сезонам, и так далее.

2) Доставка информации неизбежно влечет за собой дополнительные процессы. Например, не каждый телефонный вызов завершается разговором, так как абонент может не ответить, или быть занят, или произошла ошибка в наборе номера, технические причины, отсутствие свободных ресурсов, и так далее. Однако, независимо от исхода, каждая попытка обслуживания приводит к занятию определенных ресурсов системы, при этом создается так называемая нагрузка на коммутационные устройства и устройства управления. Нагрузка характеризует интенсивность обмена информацией между потребителями. Нагрузка определяется интенсивностью поступления вызовов, а также длительностью занятия вызова элементов соединительного тракта. Таким образом, вторая задача ТТТ состоит в изучении нагрузки, создаваемой потоком вызовов (ПВ).

3) По нагрузке, создаваемой поступающими вызовами, по характеру ее распределения во времени и с расчетом допустимого качества обслуживания, определяется необходимый объем оборудования как для коммутации, так и для управления. В этом состоит еще одна важная задача ТТТ. При этом под качеством обслуживания обычно понимается некоторая функция потерь. Также при решении задачи определения объема оборудования, учитывается порядок приема вызовов на обслуживание (таких алгоритмов достаточно много, например, FIFO, LIFO). Также в технике связи учитываются дисциплины обслуживания (ДО): с потерями, с ожиданием, с повторными вызовами, и так далее.

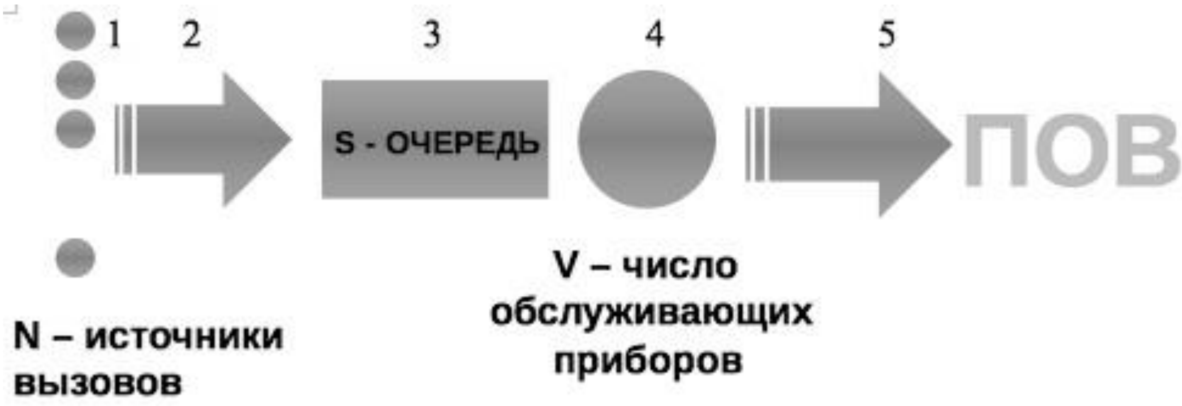
4) Задача установления закономерностей между ПВ, нагрузкой и объемом оборудования, с тем, чтобы обеспечить удовлетворительное качество обслуживания при наименьшем объеме оборудования, для оптимального построения новых систем и сетей связи (задача синтеза).

ТТТ как дисциплина имеет международный статус и тесно связана с такими теориями, как теория массового обслуживания, теория вероятностей, математическая статистика, теория случайных процессов, комбинаторика, и так далее.

В развитие ТТТ большой вклад внесли российские ученые, такие как А.Н. Колмогоров, А.Я. Хинчин, А.Д. Харкевич, Б.С. Лившиц. Среди зарубежных ученых следует выделить Эрланга, Пальма, Энгсета, Кроммелина. Основные результаты ТТ были получены в начале 20 века.

МОДЕЛЬ ОБСЛУЖИВАНИЯ ОБЩЕГО ВИДА

В теории массового обслуживания рассматривается следующее представление любой системы, которая принимает и обрабатывает информацию. Такая модель также рассматривается и в ТТТ. Рассмотрим ее элементы.



1. N — источники вызовов.
2. Источники вызовов создают ПВ, поступающих в обслуживающую систему.
3. Вызовы поступают в очередь, где S — число мест для ожидания.
4. Обслуживающий прибор(ы), число которых равно V (например, число операторов системы или число линий в пучке).
5. Поток обслуженных вызовов — ПОВ.

Для описания таких систем широко распространено пятисимвольное обозначение, которое применимо к любой системе массового обслуживания. Такое обозначение носит название классификации Кендалла:

A/B/V/K/N

A — процесс, характеризующий поступление вызовов. Этот процесс записывается в виде функции распределения промежутков между вызовами *A(x)*, где *x* — длина промежутка (длительность).

B — процесс, характеризующий процесс обслуживания вызова. Описывается через функцию распределения длительности обслуживания *B(x)*, где *x* — длительность обслуживания.

V — количество обслуживающих приборов:

$Min=1$ — однолинейная система обслуживания;

$Max=\infty$ — немедленная система обслуживания.

$K=S+V$ — емкость накопителя системы.

N — число источников (заявок, вызовов, требований).

Некоторые позиции в классификации могут принимать значения бесконечность, либо вообще отсутствовать, если этот символ последний.

Примеры: $A/B/V/\infty/N$ — система с бесконечной очередью;
$A/B/V$ — бесконечное число источников с бесконечной очередью.
Процессы A и B могут относиться к классу случайных или марковских процессов. В этом случае в классификации это отображается символом «M»: $M/M/V/K/N$.

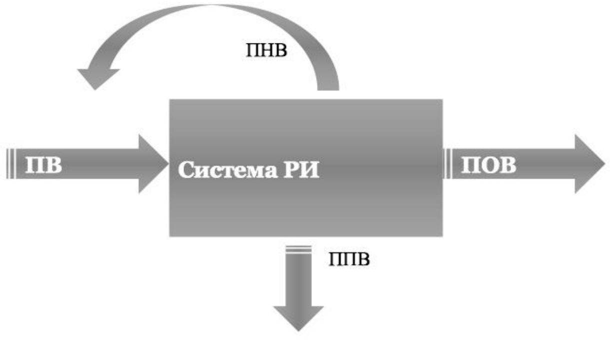
Марковский процесс характеризуется свойством: будущее не зависит от прошлого процесса и определяется только его настоящим. Если функция распределения обладает этим свойством, то она имеет вид показательного (экспоненциального) распределения: $e^{\mu t}$.
Еще один вид закона распределения — детерминированная функция распределения: $D/D/V$.

С точки зрения обслуживания, детерминированное распределение вычислительно менее удобно, поэтому такие модели чаще рассчитываются с помощью численных методов. Если процесс определить трудно, то он считается процессом общего вида и обозначается: $G/G/V/K/N$ — наихудший случай для расчета (самый сложный).

ДИСЦИПЛИНЫ ОБСЛУЖИВАНИЯ (ДО) В ИНФОРМАЦИОННЫХ СЕТЯХ

В телекоммуникационных системах наибольшее распространение получили три модели обслуживания:

1. ДО с отказами и повторными вызовами



РИ — распределение информации

ПВ — поток вызовов

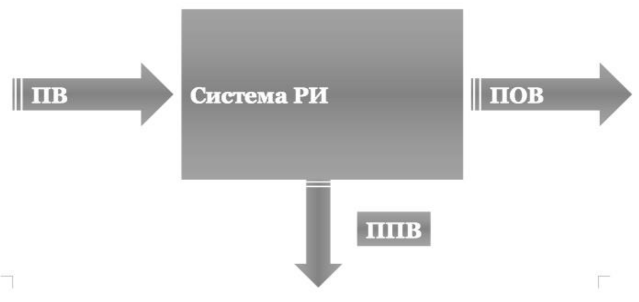
ПНВ — поток необслуженных вызовов

ПОВ — поток обслуженных вызовов

ППВ — поток потерянных вызовов.

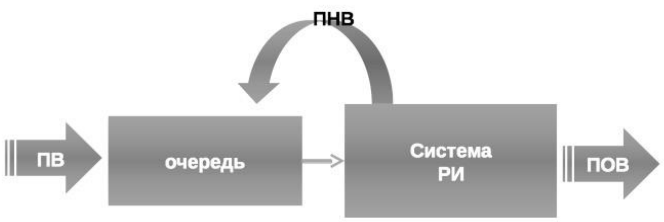
Система РИ представляет вызовам свои ресурсы (места в очереди, обслуживающие приборы). Часть вызовов успешно обслуживается и образует ПОВ. Другая часть образует ППВ, а еще одна часть формирует ПНВ, который возвращается на вход системы. Источник вызова в случае неуспешного обслуживания может послать повторный вызов с целью установки этой самой связи. Например, телефонные абоненты, в случае занятости вызываемого абонента, обычно проявляют настойчивость и организуют серию повторных вызовов. Данная модель является наиболее реальной, то есть хорошо отражает действительную ситуацию. На одно успешное соединение приходится 2-4 повторных вызова. Эту модель можно использовать для расчетов параметров системы, которая обеспечивает передачу как речевой информации, так и данных. На практике используется достаточно редко из-за математической сложности.

2. ДО с потерями



Данная модель характеризуется тем, что вызов, поступивший в момент занятости ресурсов, получает отказ. В такой системе каждый вызов, в том числе и повторный, воспринимается как новый. Модель удобна для расчета оборудования системы с ограниченными ресурсами. Многие реальные модели вполне соответствуют данной, а математический аппарат значительно проще.

3. ДО с ожиданиями

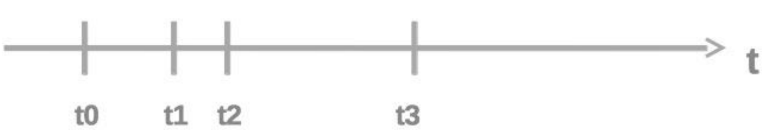


В данной модели нет явных потерь вызовов, так как любой вызов при отсутствии ресурсов будет ожидать обслуживания. Процесс обработки информации начинается с того, что вызов, не найдя свободных ресурсов, становится в очередь; по мере освобождения ресурсов вызов получит обслуживание. В случае неуспешного обслуживания, вызов возвращается на вход в систему, то есть может снова встать в очередь. Данная дисциплина широко используется, в частности, для расчета управляющих устройств в системах с коммутацией пакетов и коммутацией сообщений.

ПОТОКИ ВЫЗОВОВ (ПВ)

ПВ связаны с процессом доставки информации в систему и характеризуются в классификации Кендалла функцией распределения длительности промежутков между вызовами $A(x)$.

ПВ — это последовательность вызовов, поступающих в детерминируемые или случайные моменты времени при непрерывном течении этого времени.



Удобно графически представлять ПВ в виде графического изображения с последовательно расположенными точками t. t_0 — момент начала отсчета.

Через $z_1,z_2,z_3,\ldots$ обозначим промежутки между моментами поступления вызовов. Момент времени — число, соответствующее промежутку времени от начала отсчета t_0 до рассматриваемой точки на оси времени. Моментов времени бесконечно много.

Вызывающий момент времени — момент времени, в который поступает то или иное число вызовов.

Существует три эквивалентных способа задания ПВ:

1) При помощи моментов поступления вызовов t_i. $t_0,t_1,t_2,\ldots$.

2) При помощи промежутков между вызовами $z_1,z_2,\ldots$ (z_0 не существует).

3) При помощи последовательности чисел $k_1,k_2,\ldots$, характеризующих количество вызовов, поступающих на отрезки времени $[0, t_1)$, $[0, t_2)$, $[0, t_3)$.

Эквивалентность способов задания ПВ состоит в том, что если ПВ задан при помощи одного из них, то эти описания позволят описать его и двумя другими способами.

ОСНОВНЫЕ СВОЙСТВА ПВ

Реальные ПВ обладают свойствами, которые помогают упростить описание этих потоков. К их числу относятся: стационарность, ординарность, наличие или отсутствие последействий.

1. Стационарность.

Стационарный ПВ — это такой поток, в котором характер поступления вызовов на некотором промежутке времени зависит от длины этого промежутка, но не зависит от того, где на оси времени этот промежуток находится.

Таким образом, ПВ можно назвать стационарным, если вероятность поступления того или иного числа вызовов на некотором промежутке времени для случайного потока, или число вызовов на некотором промежутке времени для детерминированного потока, зависит от длины промежутка и не зависит от того, где на оси времени этот промежуток находится. Таким образом, стационарность ПВ означает независимость его вероятностного режима от времени.

Любой стационарный случайный ПВ можно задать семейством условных вероятностей $F_k(\tau)$, где $k=0,1,2,...$ поступления k вызовов на некотором отрезке времени τ.

ПВ называется нестационарным, если для случайного ПВ вероятность поступления некоторого числа вызовов за заданный промежуток времени зависит как от длины этого промежутка, так и от того, где на оси времени этот промежуток расположен.

Рассмотрим примеры потоков, которые обладают или не обладают свойством стационарности.

1. ПВ, поступающих на телефонную станцию в течение суток.

Очевидно, что в дневные часы суток интенсивность поступаемого ПВ больше, чем в ночное время. Следовательно, такой ПВ является нестационарным.

2. Но если ограничить наблюдения за данным ПВ меньшим временем, то с приемлемой для практики точностью такой поток можно считать стационарным. Для этого принято рассматривать час наибольшей нагрузки (ЧНН). ЧНН — это непрерывный интервал времени в 60 минут, в течение которого интенсивность нагрузки, оцениваемая как время занятия приборов системы обслуживания, является наибольшей.

2. Последействие.

ПВ называется потоком без последействия, если вероятность поступления того или иного числа вызовов на некотором отрезке врем до ени не зависит от процесса поступления вызовов до момента наблюдения. То есть, не зависит от того, в какие моменты времени и сколько вызовов поступило до начала момента наблюдения. Таким образом, для потока без последействия прошлая история потока не играет никакой роли для предсказания его будущего.

Поток без последействия можно задать с помощью функции распределения $F_k(t_1,t_2)$, где: $k=0,1,2,...$; где $F_k(t_1,t_2)$ - условная вероятность поступления k-вызовов на промежутке времени от t_1 до t_2, вычисленная при любом предположении о течении процесса обслуживания до момента t_1 (точки начала наблюдения). Это значит, что такая вероятность равна безусловной вероятности, и подобный поток можно задать семейством вероятностей $P_k(t_1,t_2) = F_k(t_1,t_2)$.

Если поток без последействия стационарен, очевидно, что его можно задать семейством безусловных вероятностей $P_k(t)$, то есть вероятность поступления вызовов для стационарного потока не зависит от начала отсчета, следовательно, момент начала отсчета не учитывается; где t — длина промежутка наблюдения.

На практике существуют потоки с последействием, среди которых выделяют потоки с частичным последействием. В классе потоков с частичным последействием наиболее значимыми являются потоки с простым последействием. Простое последействие означает, что вероятность за бесконечно малый промежуток времени $[t,t+\tau)$, где $\tau\to 0,$ определяется состоянием системы в момент времени t. При этом под состоянием системы понимается полная информация о том, какие входы и выходы системы заняты.

Примеры:

1) ПВ на телефонную станцию при большой емкости узла можно считать ПВ без последствия. Если емкость узла составляет порядка 10000 абонентов, то при нормальном уровне нагрузки в соединении участвуют от 10 до 20% абонентов. Это статистические данные. Это означает, что поведение абонентов можно рассматривать как независимое. Причем не только от поведения других абонентов, но и от того, каким был процесс поступления вызовов до начала

момента наблюдения. И этот процесс не зависит от того, в каком состоянии находится обслуживающая система.

2) ПВ от спаренного телефонного аппарата. В подобных случаях поступление вызова от одного абонента зависит, в каком состоянии находится другой абонент. Такой ПВ обладает последействием.

3) ПВ от абонентов офисной станции для выхода в город. Так как данный поток является потоком от ограниченного числа источников, то состояние системы до момента наблюдения является существенным для предоставления обслуживания других абонентов. Это ПВ с последействием.

4) Вызовы на перегруженное направление. Это также ПВ с последействием, так как ситуация зависит от того, в какой момент происходит обращение к системе.

3. Ординарность.

Возьмем некий отрезок $[t,t+\tau)$ и обозначим через $\pi_k[t,t+\tau)$ вероятность поступления k и более вызовов на отрезке времени $[t,t+\tau)$. ПВ называют ординарным, если вероятность поступления двух и более вызовов на отрезке $[t,t+\tau)$ при $\tau\to 0$ есть бесконечно малая величина большего порядка малости, чем τ. То есть: $\pi_2(t,t+\tau)=o(\tau)$, при $\tau\to 0$. Такая запись эквивалентна следующему пределу: $\lim\limits_{\tau\to 0}\dfrac{\pi_2(t;t+\tau)}{\tau}=0$.

Практически ординарность потока означает невозможность поступления двух и более вызовов в любой момент времени t.

ПВ называют неординарным, если в каждый вызывающий момент времени возможно поступление более одного вызова. Например, телефонный вызов можно считать ординарным, однако встречаются неординарные вызовы, например, при конференц-связи, групповой рассылке и пр.

ХАРАКТЕРИСТИКИ ПВ

Основными характеристиками ПВ являются:

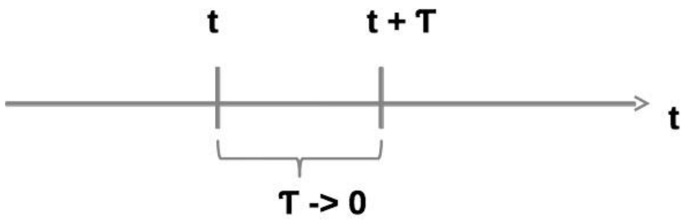
1. Ведущая функция потока
2. Интенсивность потока.
3. Параметр потока.

Рассмотрим их подробнее.

1. Ведущая функция потока *Λ(t)* — математическое ожидание числа вызовов в интервале времени *[0;t)*. Очевидно, что функция *Λ(t)* - неотрицательная, неубывающая, в практических задачах имеет конечное значение.

2. Интенсивность потока. Различается для стационарного и нестационарного потоков. Нестационарный поток характеризуется мгновенной интенсивностью $\mu(t)$ в момент времени t; она определяется так:

$$\mu(t) = \lim_{\tau \to 0} \frac{\Lambda(t+\tau) - \Lambda(t)}{\tau}.$$



Это скорость изменения величины в момент времени *t*.

Стационарный поток характеризуется средней интенсивностью μ, то есть математическим ожиданием числа вызовов в единицу времени.

Размерность для измерения интенсивности обратно пропорциональна времени: $[\mu] = \left[\frac{1}{t}\right]$. За единицу времени можно выбрать час, минуту, секунду. Часто в качестве единицы времени выбирается средняя длительность одного обслуживания, что приводит к упрощению вычислений. Очевидно, для стационарного потока справедливо равенство: $\Lambda(t) = \mu t$.

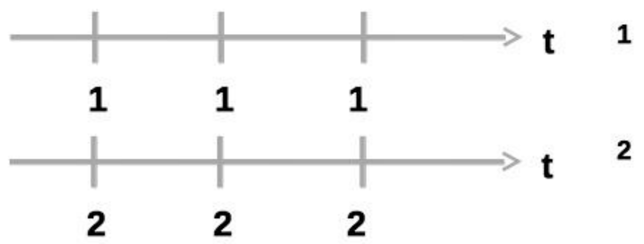
3. Параметр потока *λ(t)* для некоторого случайного потока в момент времени *t* определяется как предел отношения вероятности поступления хотя бы одного вызова на интервале времени *[t,t+τ)* к длине этого интервала, при стремлении этой длины к нулю: $\lambda(t) = \lim_{\tau \to 0} \frac{\pi_1(t;t+\tau)}{\tau}$. **(*)**

Параметр случайного потока *λ(t)* определяет в момент времени *t* плотность вероятности поступления вызывающего момента, то есть выражение **(*)** эквивалентно следующему равенству: $\pi_1(t;t+\tau) = \lambda(t) \times \tau + o(\tau)$, при $\tau \to 0$. **(**)**

Согласно определению стационарного потока, вероятность поступления определенного числа вызовов за некоторый промежуток времени одна и та же, независимо от места, где расположен этот промежуток. Следовательно, плотность вероятности для стационарного потока есть величина постоянная, не зависящая от времени. То есть для случайных стационарных потоков $\lambda(t)=\lambda$ и **(**)** приобретает вид: $\pi_1(t;t+\tau) = \lambda \times \tau + o(\tau)$ при $\tau \to 0$.

Чтобы лучше понять различия между характеристиками стационарного потока μ и λ, рассмотрим следующий пример:

Пусть есть два ПВ, моменты вызовов которых совпадают: 1 — ординарный ПВ, 2 — неординарный ПВ. Для 1-го ПВ в каждый вызывающий момент поступает один вызов, для 2-го — по два.



Так как вызывающие моменты потоков совпадают, то $\lambda_1=\lambda_2$. Однако очевидно, что интенсивности $\mu_1<\mu_2$. Таким образом, параметр потока характеризует не ПВ в целом, а только поток вызывающих моментов.

ОСНОВНЫЕ ХАРАКТЕРСТИКИ ПВ

Информацию по характеристикам ПВ сведем в итоговую таблицу.

Нестационарный поток	$\Lambda(t)$	$\mu(t)$	$\lambda(t)$
Стационарный поток	μt	μ	λ

ПРОСТЕЙШИЙ ПВ

Простейший ПВ — это основная математическая модель для потока, действующего в информационных системах. В качестве примера рассмотрим ПВ, поступающий на узел коммутации подвижных станций. Такой поток создается большим источником вызовов, на практике, это когда $N>100$, благодаря этому действие источников вызовов можно рассматривать как независимое, а это значит, что такой реальный поток, с достаточной для практики точностью, можно рассматривать как поток без последействия. Если предположить, что каждый источник может генерировать только один вызов, то такой поток обладает свойством ординарности. Если ограничить наблюдение за потоком рамками ЧНН, то такой поток можно рассматривать как стационарный. Совокупность рассмотренных свойств определяет математическую модель простейшего ПВ.

Простейший поток — это стационарный ординарный поток без последействия.

Простейший ПВ можно описать двумя способами:

1. При помощи функции распределения промежутков между вызовами $A(x)$.
2. При помощи функции распределения числа событий потока, которые происходят на отрезке от 0 до t. То есть для описания потока можно

использовать вероятность поступления вызовов на данном промежутке t: $P_k(t)$, $[0;t)$, где $k=0,1,2,..$.

Рассмотрим сначала второй способ описания потока. Для этого рассмотрим вероятность $P_k(a;a+t)$ — вероятность поступления точно k вызовов на промежутке времени $[a;a+t)$. Представим промежуток $[a;a+t)$ состоящим из двух частей и добавим $a+t+\tau$. Образуется промежуток $[a;a+t+\tau)=[a;a+t)+[a+t;a+t+\tau)$. Для определения вероятности $P_k(t)$ будем последовательно использоваться свойства простейшего ПВ:

1. Воспользуемся свойством стационарности. Согласно этому свойству, вероятность поступления некоторого числа вызовов на промежутке времени зависит от длины этого промежутка и не зависит от того, где на оси времени расположен этот промежуток. Следовательно, можно упростить введенные обозначения и записывать их следующим образом:

$[a;a+t+\tau)\rightarrow[t+\tau)$

$[a;a+t)\rightarrow[t)$

$[a+t;a+t+\tau)\rightarrow[\tau)$

Рассмотрим все возможные события, которые приводят к тому, что на большом промежутке $[t+\tau)$ поступит точно k вызовов. Чтобы это произошло, необходимо, чтобы на промежутке t поступило: k, или k-1, или k-2, или 0 вызовов, а на промежутке τ соответственно: 0, или 1, или 2, …, или k вызовов. То есть:

$t+\tau$	t	τ
k	k	0
k	$k-1$	1
k	$k-2$	2
.	.	.
.	.	.
.	.	.
k	0	k

На рассматриваемых промежутках происходит только появление (рождение) вызовов, в терминах марковских процессов такой процесс называют процессом чистого размножения.

2. Рассмотрим свойство отсутствия последействия. Оно указывает на то, что события, происходящие на отрезках t и τ, являются независимыми. Эти события являются несовместимыми (соединяются через «или»). Следовательно, можно за каждой реализацией записать следующее выражение: $P_k(t+\tau)=P_{k-i}(t)\times P_i(\tau)$, где $i=0,1,2,...,k$. Здесь, $P_k(t+\tau)$ — вероятность поступления точно k вызовов за промежуток времени $t+\tau$. Учитывая, что рассмотренные события являются независимыми, можно воспользоваться формулой полной вероятности ФПВ (Формула Байеса): $P_k(t+\tau)=\sum_{i=0}^{k} P_{k-i}(t)\times P_i(\tau)$, где $k=0,1,2,...$ **(*)**

3. Воспользуемся свойством ординарности. С этой целью устремим τ к нулю. В соответствии с определением для ординарного ПВ вероятность поступления двух и более вызовов на промежутке $\pi_2(\tau)$, есть бесконечно малая величина большего порядка малости, чем τ, то есть: $\pi_2(\tau)=o(\tau)$. Следовательно, тем более бесконечно малым является поступление точно двух вызовов $P_2(\tau)=o(\tau)$, точно трех вызовов $P_3(\tau)=o(\tau)$, и так далее.

Таким образом, при $\tau\to 0$ в выражении **(*)** появляются бесконечно малые величины. Просуммировав эти бесконечно малые величины, вынесем их за знак суммы. При этом выражение примет вид: $P_k(t+\tau)=\sum_{i=0}^{1} P_{k-i}(t)\times P_i(\tau)+o(\tau)$ **(**)**, где $\tau\to 0$; $k=0,1,2,...$. Сделаем дальнейшие преобразования:

$P_k(t+\tau)=P_{k-1}(t)P_1(\tau)+P_k(t)P_0(\tau)+o(\tau)$, где $\tau\to 0$; $k=0,1,2,...$.

В полученном выражении определим величины $P_1(\tau)$ и $P_0(\tau)$. Можно записать, что $P_1(\tau)=\pi_1(\tau)-\pi_2(\tau)$, при $\tau\to 0$.

Вероятность $\pi_1(\tau)$ определим, исходя из значения параметра простейшего потока. $\lambda=\lim_{\tau\to 0}\frac{\pi_1(\tau)}{\tau}$, заданное выражение эквивалентно следующей записи: $\pi_1(\tau)=\lambda\tau+o(\tau)$. Исходя из свойства ординарности, $\pi_2(\tau)=o(\tau)$, при $\tau\to 0$. Таким образом, $P_1(\tau)=\lambda\tau+o(\tau)$.

Определим $P_0(\tau)$, рассуждая аналогично, запишем: $P_0(\tau)=\pi_0(\tau)-\pi_1(\tau)$, при $\tau\to 0$. $\pi_0(\tau)$ — вероятность достоверного события, которая равна единице, то есть $\pi_0(\tau)=1$. Соответственно, выражения для $P_0(\tau)$ приобретают следующую форму:

$P_0(\tau)=1-(\lambda\tau+o(\tau))=1-\lambda\tau+o(\tau)$, при $\tau\to 0$.

Подставим выражения $P_1(\tau)$ и $P_0(\tau)$ в преобразованное выражение **(**)**:

$P_k(t+\tau)=P_{k-1}(t)(\lambda\tau+o(\tau))+P_k(t)(1-\lambda\tau+o(\tau))+o(\tau)$, где: $\tau\to 0$; $k=0,1,2,\ldots$.

Перенесем в левую часть уравнения вероятность $P_k(t)$ и поделим обе части на τ:

$$\frac{P_k(t+\tau)-P_k(\tau)}{\tau}=P_{k-1}(t)\times\left(\lambda+\frac{o(\tau)}{\tau}\right)-P_k(t)\times\left(\lambda+\frac{o(\tau)}{\tau}\right)+\frac{o(\tau)}{\tau},$$ где: $\tau\to 0$; $k=0,1,2,\ldots$.

Перейдем к пределу, при этом в левой части выражения для каждого выражения получим производную от t, а выражение $\frac{o(\tau)}{\tau}$ обратится в ноль. Получим окончательное выражение: $\frac{d}{dt}P_k(t)=\lambda\times P_{k-1}(t)-\lambda\times P_k(t)$, где: $k=0,1,2,\ldots$.

Получилась система из бесконечно большого числа дифференциальных уравнений. Существуют различные методы решения этой системы. Одним из удобных способов является метод производящих функций. Результатом решения данной системы является следующая формула:

$P_k(t)=\frac{(\lambda\times t)^k}{k!}\times e^{-\lambda t}$ — формула Пуассона, где $k=0,1,2,\ldots$.

Итак, мы рассмотрели процесс чистого размножения, который описывается формулой Пуассона. Его также называют стационарным Пуассоновским процессом. Следовательно, простейший ПВ называют стационарным Пуассоновским потоком.

ХАРАКТЕРИСТИКИ ПРОСТЕЙШЕГО ПВ

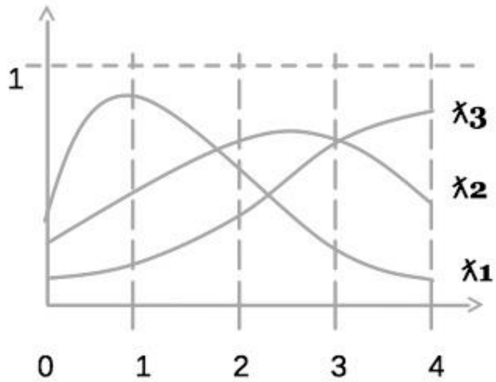
1. При объединении n независимых простейших ПВ с параметрами $\lambda_1,\lambda_2,\ldots,\lambda_n$ получается общий суммарный, также простейший поток вызовов с параметрами $\lambda_{\Sigma}=\lambda_1,\lambda_2,\ldots,\lambda_n$.

Объединение большого числа стационарных ординарных потоков с практически любым последействием, при малых значениях параметров этих потоков, создает общий поток, близкий к простейшему.

2. Оценим сумму вероятностей событий простейшего потока

$\sum_{k=0}^{\infty} P_k(t) = \sum_{k=0}^{\infty} \frac{(\lambda \times k)^k}{k!} \times e^{-\lambda t} = \underbrace{\sum_{k=0}^{\infty} \frac{(\lambda \times k)^k}{k!}}_{=e^{\lambda t}} \times e^{-\lambda t} = e^{\lambda t} \times e^{-\lambda t} = 1$. Следовательно, $\sum_{k=0}^{\infty} P_k(t) = 1$.

3. Исследуем функцию Пуассона графически. Функция Пуассона является функцией дискретной случайной величины числа вызовов. Рассмотрим характер изменения данной функции при изменении параметра λ.



$P_k(f)=f(k)$

Пусть $t=1$

λ - *vor*

Будем для простоты рассматривать огибающие вероятностей дискретной случайной величины. С ростом λ огибающая значений вероятностей перемещается вправо и принимает вид все более симметричной относительно оси координат. Следовательно, $\lambda_1<\lambda_2<\lambda_3$. На практике при $\lambda t>10$ имеет место хорошее совпадение между огибающей закона Пуассона и нормальным законом распределения, который является законом распределения непрерывной случайной величины.

4. Установим связь между параметром λ и интенсивностью μ для простейшего ПВ. Пусть $t=1$, следовательно, $P_k(1) = \frac{\lambda^k}{k!} \times e^{-\lambda}$. По определению

$\mu = \sum_{k=1}^{\infty} k \times P_k(1) = \sum_{k=1}^{\infty} k \times \frac{\lambda^k}{k!} \times e^{-\lambda} = e^{-\lambda} \times \lambda \times \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} =$ [пусть $n=k-1$]

$= e^{-\lambda} \times \lambda \times \underbrace{\sum_{0}^{\infty} \frac{\lambda^n}{n!}}_{=e^{\lambda}} = e^{-\lambda} \times \lambda \times e^{\lambda} = \lambda$. Следовательно, $\lambda=\mu$. Данное равенство верно для любого стационарного потока, в том числе и для потока с последействием. Это равенство указывает на то, что простейший ПВ полностью описывается интенсивностью потока, то есть μ. На практике достаточно знать среднее число вызовов в единицу времени, чтобы определить все остальные характеристики простейшего потока, в том числе и вероятность P_k.

5. Определим математическое ожидание числа вызовов, поступающих за промежуток времени t. Обозначим интересующую нас величину k, k$=\Lambda(t)=\sum_{k=1}^{\infty} k \times P_k(t)=\lambda \times t=\mu \times t$. Следовательно, среднее число вызовов определяется выражением k$=\lambda t$. Этому же значению соответствует и дисперсия числа поступающих вызовов: $D_k=\lambda t$. Математическое ожидание характеризует среднее число вызовов, а дисперсия — меру разброса относительно этого среднего значения.

Среднеквадратическое отклонение (СКО): $\delta=\sqrt{D_k}$.

Таким образом, первый (k) и второй (D_k) моменты распределения дискретной случайной величины равны между собой.

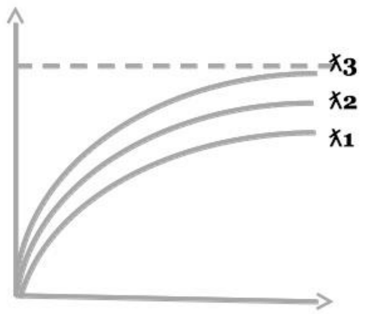
ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ ПРОМЕЖУТКОВ ВРЕМЕНИ МЕЖДУ ВЫЗОВАМИ

Рассмотрим другой способ описания простейшего потока при помощи функции распределения промежутков между вызовами $A(x)$. Обозначим $F(x)$ — вероятность того, что некоторая случайная величина z не превзойдет некоторое заданное значение x, где x — промежуток времени. $F(x)=P\{z \leq x\}$ — вероятность того, что в промежуток времени между двумя последовательными моментами поступления вызовов не превзойдет x. Эта вероятность равносильна вероятности поступления хотя бы одного вызова на промежутке $[0;x)$. Следовательно, $F(x)=\pi_1(x)=\pi_0(x)-P_0(x)=$ [где: $\pi_0=0,1,2,...$; $P_0=0$; π_0 - достоверное событие. Следовательно,] $=1-e^{-\lambda x}$ [из формулы Пуассона]. Получаем: $F(x)=1-e^{-\lambda x}$, $A(x)=F(x)$. Таким образом, $F(x)$ подчиняется экспоненциальному (показательному) закону.

Только такая показательная функция обладает свойством отсутствия последействия (из дискретных распределений отсутствием последействия обладает геометрическое распределение).

Марковский процесс, обладающий свойством отсутствия последействия, характеризуется упрощенной математической моделью.

Исследуем функцию *F(x)* графически. Зафиксируем некоторую точку x_1 и покажем три графика подобного вида.



Чем больше λ, тем больше доля промежутков между вызовами будет иметь длительность меньше, чем x_1. То есть, тем выше вероятность того, что за промежуток $[0;x_1)$ поступит хотя бы один вызов. Следовательно, $\lambda_1<\lambda_2<\lambda_3$.

Выберем некоторую точку $F^*(x)$, опустим перпендикуляры из точек пересечения этих значений на ось x. Чем больше λ, тем меньше длина промежутка, который не будет превышен. Из анализа этих графиков виден физический смысл формулы для математического ожидания длины промежутка между вызовами: $\dot{z}=\frac{1}{\lambda}$.

Таким образом, чем больше λ, тем меньше математическое ожидание длины промежутка между вызовами.

Дисперсия для промежутков между вызовами имеет следующее значение: $D_z=\frac{1}{\lambda^2}$, а среднеквадратическое отклонение $\delta_z=\sqrt{D_z}=\frac{1}{\lambda}$.

ПОТОКИ С ПРОСТЫМ ПОСЛЕДЕЙСТВИЕМ

Пусть существует система РИ, на которую поступает ПВ. Пусть из системы выходит поток обслуженных вызовов, и система предоставляет поступающим вызовам свои ресурсы, например, соединительные пути.

Состояние системы $S(t)$ — это такое состояние, при котором фиксированы каждый занятый и свободный ресурс системы в момент t, при этом известно, вызовы каких именно источников обслуживаются данной системой. Если в процессе исследования системы необходимо учитывать состояние, которое принимает система, то параметр поступающего ПВ определяется следующим образом: $\lambda_{s(t)} = \lim\limits_{\tau \to 0} \frac{\pi_1(t;t+\tau)}{\tau}$, при условии, что в момент времени t, система находилась в состоянии $S(t)$; где $\pi_1(t;t+\tau)$ при $S(t)$ — вероятность поступления за промежуток $[t;t+\tau)$ одного и более вызовов, если в момент времени t, система находится в состоянии $S(t)$.

Рассмотрим среди потоков с последействием такие потоки, у которых это последействие является простым. Поток с простым последействием — это ординарный поток, для которого в любой момент времени существует конечный параметр потока состояния $\lambda_{s(t)}$. Это параметр зависит только от состояния системы в момент времени t и не зависит от процесса обслуживания вызовов до момента времени t. Данный поток для определения $\lambda_{s(t)}$ требует знания значения системы $S(t)$ в этот момент. Также такой поток является нестационарным, так как в различные моменты времени $t_1,t_2,...$ имеют место различные состояния $S(t_1),S(t_2),...$ соответственно. Однако при $t\to\infty$ поток с простым последействием стремится к стационарному потоку.

Рассмотрим несколько частных случаев с простым последействием.

Симметричный поток — это такой поток с простым последействием, параметр которого $\lambda_{s(t)}$ в любой момент времени t зависит только от числа установленных соединений i, то есть зависит от числа вызовов, которые обслуживаются системой в рассматриваемый момент t. Следовательно, для симметричного потока $\lambda_{s(t)}=\lambda_i$. Зависимость λ_i может быть различной.

Например, примитивный поток — это такой симметричный поток, параметр которого λ_i зависит от числа обслуживаемых системой вызовов, причем параметр λ_i прямо пропорционален числу свободных источников. $\lambda_i = \alpha(N\text{-}i)$, где α — коэффициент пропорциональности, численно равный параметру одного источника; N — число источников вызовов; i — число занятых источников; $N\text{-}i$ — число свободных источников.

Частным случаем примитивного потока является простейший ПВ. Простейший поток — это такой примитивный поток, параметр которого λ_i не зависит от состояния системы. $\lambda_i=\lambda=const$. Действительно, с увеличением N и с уменьшением α, последействие убывает. В предельном случае, когда $N\to\infty$ и $\alpha\to\infty$, причем так, чтобы $\alpha N=const$, параметр $\lambda=\alpha N$ не зависит от состояния системы и, при этих условиях, модель примитивного потока переходит в модель простейшего потока.

Среди потоков с простым последействием, интерес представляет также поток с повторными вызовами. Рассмотрим такие потоки более подробно. Действительно, система обслуживает не все поступающие вызовы. Часть из них не обслуживается по целому ряду причин. Например, занятость абонента, отсутствие абонента, неисправности, ошибки в наборе номера, и тому подобное. Если вызов теряется, то нередко источники необслуженных вызовов совершают повторные вызовы. То есть поток с повторными вызовами состоит как из первичных, так и из повторных вызовов. Оценим параметр повторных вызовов. Пусть поток первичных вызовов будет простейшим, с параметром λ. Параметр потока повторных вызовов определим через произведение $j \times \beta$, где j — число источников повторных вызовов, и β — интенсивность повторений, то есть параметр одного источника повторных вызовов. Тогда суммарный поток $\lambda_{повт.в.}=\lambda+j\beta$. Если поток первичных вызовов является примитивным, то для потока с повторными вызовами верно: $\lambda_{повт.в.}=\alpha(N\text{-}i\text{-}j)+j\beta$.

Часть 2

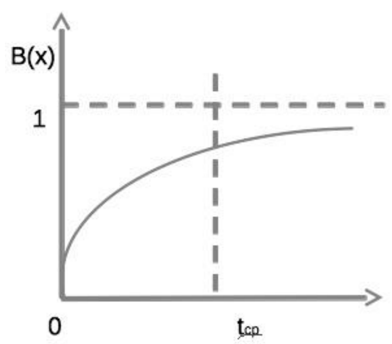
ВРЕМЯ ОБСЛУЖИВАНИЯ

Процесс обслуживания описывается символом *B(x)* на второй позиции в классификации Кендалла. Процесс обслуживания характеризуется определенной длительностью. Эта длительность может быть как случайной, так и детерминированной. В соответствии с этим, рассмотрим наиболее распространенные функции распределения длительности обслуживания.

1) Показательное распределение длительности обслуживания. Многие реальные функции распределения подчиняются показательному закону. В частности, показательное распределение хорошо описывает длительность телефонного разговора и другие подобные процессы. Показательное распределение обладает таким свойством, как полное отсутствие последействия. Математически это дает значительные преимущества, так как позволяет пользоваться теорией марковских процессов, в которых будущее не зависит от прошлого и определяется известным настоящим. Рассмотрим выражение для функции распределения длительности обслуживания: $B(x)=P\{t\leq x\}=1-e^{-\beta x}$, где β — интенсивность освобождения вызовов (параметр функции распределения *B(x)*); *B(x)* — вероятность того, что длительность обслуживания *t* будет меньше, или равна некоторому, наперед заданному, времени *x*. Математическое ожидание (среднее время обслуживания): $\bar{t}=\frac{1}{\beta}$.

Дисперсия (мера разброса, относительно среднего значения): $D_t=\frac{1}{\beta^2}$.

Среднеквадратическое отклонение: $\delta_t=\frac{1}{\beta}$.

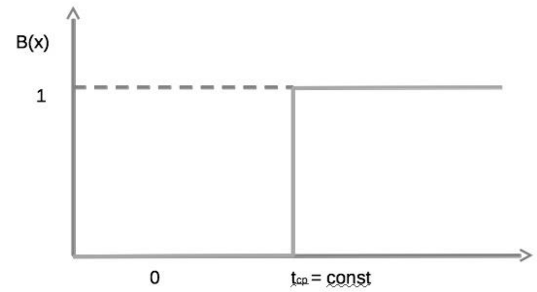


Рассмотрим график *B(x)*:

На этом графике на оси абсцисс выделим некоторую точку, назовем ее t_{cp}. Построим перпендикуляр из этой точки. Площади, которые образуются слева и справа, будут равны друг другу. С целью упрощения математических расчетов, обычно за единицу времени принимают среднюю длительность обслуживания $t=1$. В этом случае, $\beta=1$, а $B(x)=1-e^{-x}$.

2) Детерминированное распределение. Данное распределение также широко используется, например, для описания процессов обслуживания пакетов, или сообщений фиксированной длины. В этом случае $B(x) = P\{t \le x\} = \{\begin{matrix} 0;0 \le x \le t \\ 1; x \ge t \end{matrix}$.

Графически это можно изобразить так:



Детерминированное распределение значительно усложняет математический аппарат, так как свойство марковости здесь не наблюдается. Обе функции распределения: и показательная, и детерминированная, являются предельными случаями распределения и относятся к так называемым *k*-распределениям Эрланга.

НАГРУЗКА

Доставка и обработка информации в системе характеризуется определенной работой этой системы.

Можно рассмотреть три вида этой работы:

1. Работа, вносимая в систему (поступающая нагрузка).

2. Работа, исполненная системой (обслуженная нагрузка).

3. Работа, незавершенная системой (избыточная, или остаточная нагрузка).

Так как каждый вызов вносит в систему работу, измеряемую временем, определим разные виды работ математически.

1) Поступающая нагрузка.

Рассмотрим математическое ожидание работы, вносимой в систему за промежуток времени $[t_1;t_2)$.

$MA(t_1;t_2)=\mu(t_2-t_1)t$,

где M — математическое ожидание, A — поступающая нагрузка, μ — интенсивность нагрузки, t — среднее время обслуживания. (t_2-t_1) — длительность интервала.

Таким образом, поступающая нагрузка оценивается не только через количество вызовов $\mu(t_2-t_1)$, которые поступили на промежутке вызовов $(t_1;t_2)$, но зависит и от длительности обслуживания вызовов t.

2) Обслуженная нагрузка.

Для обслуженной нагрузки математическое ожидание значения нагрузки

$MУ(t_1;t_2)=\varkappa(t_2-t_1)$,

где $\varkappa$ — среднее число вызовов; (t_2-t_1) — длина рассматриваемого промежутка вызовов.

3) Избыточная или остаточная нагрузка.

Следовательно, незавершенная системой нагрузка

$U(t_1;t_2)=A(t_1;t_2)-Y(t_1;t_2)$,

где A — поступившая нагрузка; Y — обслуженная нагрузка.

Нагрузка измеряется в часозанятиях. Эта единица характерна для европейских стран, в том числе и для России, а в Америке используется секундозанятие. Обычно рассматривают нагрузку в единицу времени, придавая ей значение интенсивности нагрузки. Следовательно, интенсивность любого вида нагрузки можно записать следующим образом:

1) $y=\mu t$, где y — интенсивность поступающей нагрузки;

2) $y_0=x$, где y_0 — обслуженная нагрузка;

3) $u=y\text{-}y_0$.

Единица измерения интенсивности нагрузки - часозанятие, разделенное на час, для неё вводится название Эрл (произносится «Эрланг»).

На практике, часто удобно оценивать интенсивность поступающей нагрузки по следующей формуле: $y=N\hat{c}t$, где N — источник вызовов; $\hat{c}$ — среднее число вызовов от одного источника; t — среднее время обслуживания одного вызова.

ХАРАКТЕРИСТИКИ КАЧЕСТВА ОБСЛУЖИВАНИЯ

Характеристики качества обслуживания и эффективности систем распределения информации различаются в зависимости от конкретных задач по расчетам таких систем, от структуры самой системы, от дисциплины обслуживания и других факторов.

Рассмотрим важнейшие из этих характеристик:

1. Вероятность потерь.

В зависимости от дисциплины обслуживания различаются явные (когда вызов получает отказ в обслуживании) и условные (или неявные) (когда вызов должен ожидать обслуживания) потери. Рассмотрим виды потерь в системах с повторными вызовами, с потерями и с ожиданиями.

1) В системах с повторными вызовами различают следующие виды потерь:

а) потери по времени P_t — вероятность потери первичного вызова, или доля времени, в течение которого заняты все ресурсы системы, способные обслужить вызов, если он поступит. Иногда говорят, что это опасное время, в течение которого вызовы теряются.

б) потери по вызовам $P_в$ — вероятность безуспешного вызова, оцениваемая как отношение всех потерянных вызовов (как первичных, так и повторных) ко всем поступившим вызовам: $P_в = \frac{C_{пот}}{C}$. Для стационарных потоков $P_в$ можно рассматривать как отношение интенсивностей потерянных потоков к поступающим $P_в^{стац} = \frac{\mu_{пот}}{\mu}$.

в) потери сообщения P_c — учитывает функцию настойчивости α, которая расценивается, как вероятность того, что абонент после очередного неуспешного вызова, через случайное время создаст следующий повторный вызов. Статистика показывает, что величина α составляет, приблизительно 0,8-0.9. Потеря сообщений является наиболее объективной характеристикой в системе с повторными вызовами.

$P_c=LP_в(1\text{-}\alpha)$, где L — число вызовов, которые необходимо послать, чтобы появился успешный вызов. Математическое ожидание геометрического распределения, оценивает $L = \frac{1}{1-\alpha \times P_в}$. Очевидно, что величина (L-1) - это среднее число неуспешных вызовов, после которого появится успешный вызов. Следовательно, $P_c = \frac{P_в \times (1-\alpha)}{1-\alpha \times P_в}$.

Существуют и другие виды потерь, например, потери по нагрузке *($P_н$)*, потери по линиям *(P_v)*, и так далее. Независимо от вида потерь, любые потери, как вероятностный показатель, принимают значения в диапазоне от *0* до *1* и измеряются в процентах (*%*), или промиллях (*‰*).

Рассмотрим пример. Пусть поступило *1000* вызовов, *5* из которых были потеряны. Тогда можно определить $P_в$ как: $P_в = \frac{5}{1000} = 0{,}5\% = 5‰$.

2) В системах с потерями рассматриваются такие же характеристики качества, так как ранее было показано, что система с потерями — это частный случай системы с повторными вызовами. Оценим величину потерь сообщений в системе с потерями. Для таких систем, функция настойчивости $\alpha=0$, следовательно, $P_c=P_в$.

3) В системах с ожиданием широко распространена такая характеристика качества, как вероятность ожидания $P\{\gamma>0\}$. Это вероятность того, что поступающий вызов должен ждать, или, иначе говоря, вероятность задержки вызова, или вероятность того, что длительность ожидания начала обслуживания γ будет больше нуля. Однако данная характеристика дает неполную картину качества обслуживания, так как не позволяет судить о законе распределения γ. По этой причине, более общей характеристикой является $P\{\gamma>t\}$ — это вероятность того, что γ будет больше некоторой, заранее заданной, величины t. Обычно и γ и t измеряются в относительных единицах времени, а именно, в единицах длительности обслуживания.

При учете характеристик качества, в системе с ожиданиями также используются средние величины. К их числу относятся: среднее время ожидания начала обслуживания $\bar{\gamma}$ и среднее число вызовов, находящихся на ожидании $\bar{j}$. Для $\bar{\gamma}$ обычно рассматриваются две величины: $\bar{\gamma}$ и $\bar{\gamma}_з$ — среднее время ожидания, относящиеся только к тем вызовам, которые фактически ожидают. В общем случае, $\bar{\gamma}$ будет меньше, или равно, чем $\bar{\gamma}_в$ $(\bar{\gamma} \leq \bar{\gamma}_в)$.

2. Пропускная способность.

Пропускная способность системы — это свойство системы, заключающееся в ее возможности выполнять работу по доставке информации. Пропускная способность характеризуется объемом работы, производимым в единицу времени в соответствии с техническими характеристиками системы и режимом функционирования. Пропускную способность принято рассматривать с вероятностной и технической сторон. Вероятностная оценка связана с проблемой доставки информации в условиях обслуживания случайных потоков вызовов. Техническая сторона связана с техническими характеристиками оборудования системы. Каждая из сторон характеризуется своими показателями. Основными для вероятностной стороны является показатель пропускной способности, а для технического аспекта — показатель производительности системы.

Пропускная способность C эквивалентна средней мощности, развиваемой системой при заданном качестве, то есть это интенсивность обслуживания нагрузки (y_0) при заданном качестве: $C=y_0$. Техническая сторона характеризуется производительностью $П$ и эквивалентна максимальной интенсивности нагрузки: $П=y_0^{max}$. Таким образом, $П$ характеризует технические возможности оборудования при непрерывной нагрузке системы без интервалов простоя. В качестве показателя пропускной способности нередко рассматривается и интенсивность поступающей нагрузки при заданном качестве. Это объяснимо по двум причинам:

1) поступающая нагрузка находится в явной зависимости от вероятности потерь. Для обслуженной нагрузки явной зависимости обычно нет, поэтому работать с поступающей нагрузкой математически удобнее.

2) так как системы распределения информации работают обычно с высоким качеством обслуживания, что соответствует малому значению вероятности потерь, следовательно, различия между поступающей у и обслуженной y_0 нагрузками будет незначительны, и в инженерных расчетах ими можно пренебречь: $y \approx y_0$.

Итак, пропускная способность y является сложной функцией от многих параметров ($A(x)$ — закон, по которому в систему поступают вызовы; S — структура системы; ДО — дисциплина обслуживания; $B(x)$ — функция распределения длительности обслуживания; P — потери): $y=f(A(x),S,ДО,B(x),P)$. Часто указанная зависимость рассматривается относительно величины потерь: $p=g(y,A(x),S,ДО,B(x),P)$.

3. Время доставки сообщений.

В настоящее время все большую значимость приобретает такая характеристика качества обслуживания, как время доставки сообщений. Данное время — это сумма времен ожидания, обработки и передачи сообщения, начиная с момента поступления этого сообщения в систему и заканчивая моментом его выхода из системы. Различают среднее, наибольшее и наименьшее время доставки. Среднее время доставки принято интерпретировать, как среднее время пребывания в системе T. Оно определяется следующим соотношением: $T=\bar{\gamma}+\bar{t}$, где $\bar{\gamma}$ — среднее время ожидания начала обслуживания. Наряду с T интерес представляет такая характеристика, как среднее число требований в системе $\bar{k}$. Эти величины (T и $\bar{k}$) связаны между собой формулой Литтла:

$$\bar{k}=\lambda T,$$

где λ — параметр потока вызовов (для стационарного потока равный интенсивности поступлений вызовов в систему).

Математическое доказательство формулы Литтла сложное, поэтому воспользуемся распространенным интуитивным пояснением. Оно сводится к утверждению, что требования, входящие в систему, находят в ней то же самое число требований $\bar{k}$, которое будет в системе, когда данный вызов покидает систему. Формула Литтла не зависит ни от каких частных ограничений (ни на характер поступающего потока, ни на функцию длительности обслуживания, ни на дисциплину обслуживания, ни на саму систему, и так далее). Эта формула может рассматриваться по отношению к отдельным частям системы обслуживания:

$$\bar{k}_s=\lambda\bar{\gamma};\ \bar{k}_v=\lambda\bar{t},$$

где $\bar{k}_s$ — средняя длина очереди, а $\bar{k}_v$ — среднее число вызовов в обслуживающих приборах.

Часть 3

КЛАССИЧЕСКИЕ СООТНОШЕНИЯ И ВАЖНЕЙШИЕ ВЕРОЯТНОСТНЫЕ ПРОЦЕССЫ

К важнейшим вероятностным процессам в теории телетрафика относятся, так называемые марковские процессы. В этих процессах устанавливается очень простая форма взаимосвязи между случайными величинами, которые образуют этот процесс: если известно состояние процесса в момент времени t, то изменение состояния этого процесса после момента времени t, в вероятностном смысле не зависит от течения этого процесса до данного момента времени t. Другими словами, для марковского процесса будущее определяется известным настоящим и не зависит от прошлого.

Различают дискретные и непрерывные марковские процессы. Для дискретного марковского процесса переходы из одного состояния в другое осуществляются в дискретные моменты времени, а для непрерывного — в любые моменты времени.

Марковские процессы бывают однородными и неоднородными. Для однородных процессов вероятности перехода из одного состояния в другое не зависят от времени. Соответственно, неоднородные процессы зависят от времени. Мы будем изучать однородные непрерывные марковские процессы.

Рассмотрим модель марковского процесса в виде коммутационной системы, в которой число обслуживаемых источников нагрузки может как совпадать, так и не совпадать с числом занятых линий в этой системе. Несовпадение характеризует достаточно редкие случаи работы коммутационных систем, однако, в рамках классических, или элементарных систем теории телетрафика, рассматривается частный случай, когда число занятых источников нагрузки всегда совпадает с числом линий в этой системе. Этому условию соответствует полнодоступный пучок линий (любому вызову доступна любая свободная линия в этом пучке). Полнодоступный пучок может принимать различное число состояний, в зависимости от дисциплины обслуживания. В системах с потерями, число состояний может быть $0,1,2,...v$ (где v — число линий полнодоступного пучка), всего — $(v+1)$ состояние. Если все линии пучка заняты, система не принимает на обслуживание новые вызовы. В системах с ожиданием при наличии v занятых линий, вызовы не отвергаются системой, а ставятся на ожидание. При этом при v занятых линиях, в очереди могут находиться $0,1,2,...$ вызовов. Следовательно, суммарное число состояний в этой системе бесконечно. Итак, классические системы с чистым ожиданием имеют бесконечное число состояний.

Сформулируем задачу обслуживания вызовов полнодоступным пучком с общих позиций. Пусть *v(t)* — число линий в полнодоступном пучке в момент времени *t*. И пусть в этот момент занятыми окажутся *k* линий (*k=0,1,2,...*). Следовательно, такую систему можно рассматривать как систему, в которой существует бесконечно большое число состояний, и в этой системе может быть занято любое число линий. Такую ситуацию можно рассматривать как бесконечно-линейный пучок ($v=\infty$). Введем следующее обозначение: если в бесконечно-линейном пучке занято *k* линий, то будем считать, что он находится в состоянии *k*. Проследим за состоянием этого пучка на некотором промежутке времени: $[0;t+\tau)$. Пусть в момент времени $t=0$ система находится в состоянии *j*, и к моменту времени $t+\tau$, перешла в состояние *k*. Следовательно, в некоторый промежуточный момент времени *t*, система перешла через некоторое промежуточное состояние *r* ($0 \le r \le v$), а за промежуток времени от *t* до $t+\tau$, перешла в состояние *k*. Оценим вероятность перехода системы из состояния *j* в состояние *k*, для чего воспользуемся ФПВ для независимых событий. Согласно этой формуле, $P_{jk}(0;t+\tau)=\sum_{r=0}^{v} P_{j\tau}(0;t)\times P_{rk}(t;t+\tau)$. **(*)**

В теории случайных процессов, данное выражение называется уравнением Колмогорова-Чэпмена. Оно справедливо не только для непрерывных однородных марковских процессов, но и для дискретных неоднородных. Это уравнение справедливо, при следующих условиях: $0 \le j \le v$; $0 \le k \le v$; $0 \le r \le v$.

Важнейшим случаем марковского процесса является процесс размножения (поступление вызовов в систему) и гибели (уход вызовов из системы). Процесс размножения и гибели — это марковский процесс с непрерывным параметром *t*, имеющий конечное ($0,1,2,...,v$) или счетное количество состояний, в котором за промежуток времени $[t;t+\tau)$, при $\tau \to 0$, с вероятностью более *0* (то есть с конечной вероятностью) возможен непосредственный переход системы только в соседнее состояние (то есть ($k-1$), k, ($k+1$)). Во все другие состояния система переходит с вероятностью $o(\tau)$. Ранее рассмотренный пуассоновский процесс — это частный случай процесса размножения и гибели, в котором гибели вызовов не происходит (его называют процессом чистого размножения).

ПРОЦЕСС ОБСЛУЖИВАНИЯ СИММЕТРИЧНОГО ПОТОКА ПОЛНОДОСТУПНЫМ ПУЧКОМ (ОСНОВНЫЕ УРАВНЕНИЯ ДВИЖЕНИЯ ПРОЦЕССА РАЗМНОЖЕНИЯ И ГИБЕЛИ)

Постановка задачи. Рассмотрим полнодоступный пучок из бесконечно большого числа линий ($v=\infty$). Пусть на этот пучок поступает симметричный поток вызовов с параметром λ_k. λ_k — параметр поступления вызовов потока в состоянии k (в большинстве случаев можно считать, что это интенсивность размножения в состоянии k). Интенсивность обслуживания вызовов характеризуется параметром β_k, что можно считать интенсивностью (скоростью) гибели в состоянии k. λ_k и β_k не зависят от времени, а зависят только от состояния k. Следовательно, в силу свойств марковских процессов, события размножения и гибели являются независимыми. Рассмотрим динамику процесса, описываемым уравнением **(*)**. Определим вероятность $P_k(t)$, где $P_k(t)$ — вероятность того, что в полнодоступном пучке на момент времени t занято k линий. Общий подход к решению данной задачи рассмотрим с учетом того, что возможны переходы только в соседние состояния. Покажем все возможные события, которые приводят к тому, что в момент времени $t+\tau$, в системе будет точно k вызовов, в виде следующей таблицы:

$t+\tau$	t	τ	
k	$k-1$	1 вызов	$P_{k-1}(t)$ $P_{k-1,k}(\tau)=P_в(\tau)$ $P_k(t+\tau)=P_{k-1}(t)P_в(\tau)$
	$k+1$	1 освобождение	$P_{k+1}(t)$ $P_{k+1,k}(\tau)=P_{осв}(\tau)$ $P_k(t+\tau)=P_{k+1}(t)P_{осв}(\tau)$

	k	0 вызовов 0 освобождений	$P_k(t)$ $P_{k,k}(\tau)=[1-P_в(\tau)-P_{осв}(\tau)]$ $P_k(t+\tau)=P_k(t)[1-P_в(\tau)-P_{осв}(\tau)]$
	$k=-2,-3,...$ $k=2,3,...$	2 вызова, 3 вызова, … 2 осв., 3 осв., ...	$o(\tau)$ — указывает, что в процессе размножения и гибели, кратные рождения и кратные гибели запрещены

По формуле полной вероятности (ФПВ) получаем:

$P_k(t+\tau)=P_{k-1}(t)P_в(\tau)+P_{k+1}(t)P_{осв}(\tau)+P_k(t)[1-P_в(\tau)-P_{осв}(\tau)]$ — это выражение справедливо для случаев $k\geq1$.

$P_0(t+\tau)=P_1(t)P_{осв}(\tau)+P_0(t)[1-P_{осв}(\tau)-P_в(\tau)]$ — для $k=0$.

$$\sum_{k=0}^{\infty} P_k(t+\tau)=1$$

Необходимо решить данную систему из бесконечно большого числа уравнений для $P_k(t)$. Определим вероятность поступления вызова на отрезке τ и вероятность освобождения вызова на τ. Для определения $P_в(\tau)$, воспользуемся определением параметра потока с простым последействием: $\lambda_{s(t)}=\lim_{\tau\to 0}\frac{\pi_1(t;t+\tau)}{\tau}|_{s(t)}$, при условии, что система в момент времени t находится в состоянии $S(t)$; этой записи эквивалентна следующая: $\pi_1(t;t+\tau)=\lambda_{S(t)}\tau+o(\tau)$, при $\tau\to0$.

Нами рассматривается симметричный ПВ, в котором $\lambda_{S(t)}=\lambda_k$. Следовательно, $\pi_1(t;t+\tau)=\lambda_k\tau+o(\tau)$, при $\tau\to0$. Так как симметричный поток является ординарным потоком, а для ординарного потока, вероятность $\pi_2(t;t+\tau)=o(\tau)$, при $\tau\to0$. $P_в(\tau)=\lambda_k\tau+o(\tau)$, при $\tau\to0$.

Рассуждая аналогичным образом, для $P_{осв}(\tau)=\beta_k\tau+o(\tau)$, при $\tau\to0$. Подставим значения $P_в(\tau)$ и $P_{осв}(\tau)$ в полученную систему:

$P_k(t+\tau)=P_{k-1}(t)[\lambda_k\tau+o(\tau)]+P_{k+1}(t)[\beta_k\tau+o(\tau)]+P_k(t)[1-(\beta_k\tau+o(\tau))-(\lambda_k\tau+o(\tau))]$

$P_0(t+\tau)=P_1(t)[\beta_1\tau+o(\tau)]+P_0(t)[1-(\beta_k\tau+o(\tau)-(\lambda_k\tau+o(\tau))]$

Перенесем в левую часть вероятность $P_k(t)$, и поделим обе части на τ:

$$\frac{P_k(t+\tau)-P_k(t)}{\tau}=(\lambda_{k-1}\times P_{k-1}(t)+\beta_{k+1}\times P_{k+1}(t)-\lambda_k+\beta_k)\times P_k(t)+\frac{o(\tau)}{\tau}$$, где $k\geq 1$.

$$\frac{P_0(t+\tau)-P_0(t)}{\tau}=\beta_1\times P_1(t)-\lambda_0\times P_0(t)+\frac{o(\tau)}{\tau}$$, где $k=0$.

$$\sum_{k=0}^{\infty}P_k(t+\tau)=1$$

Перейдем к пределу, при этом, в левой части получим $\frac{d}{dt}P_k(t)$ и $\frac{o(\tau)}{\tau}\to 0$:

$$\frac{d}{dt}P_k(t)=\lambda_{k-1}P_{k-1}(t)+\beta_{k+1}P_{k+1}(t)-\lambda_k+\beta_k P_k(t)$$

$$\frac{d}{dt}P_0(t)=\beta_1 P_1(t)-\lambda_0 P_0(t) \qquad \textbf{(**)}$$

$$\sum_{k=0}^{\infty}P_k(t+\tau)=1$$

Система **(**)** - это система дифференциально-разностных уравнений, решением которых является характеристика $P_k(t)$. Эта характеристика является основной для решения многих задач ТТ.

ОБЩЕЕ РЕШЕНИЕ ДЛЯ СТАЦИОНАРНОГО РЕЖИМА

Будем рассматривать полнодоступный пучок в состоянии статистического равновесия. Перейдем от вероятности $P_k(t)$ к аналогичной вероятности для стационарного режима. Для этого определим, устанавливается ли вероятность $P_k(t)$ независимо от времени.

Введем следующее предположение: $\lim\limits_{k\to\infty}P_k(t)=P_k$. Для симметричного потока известно, что при $t\to\infty$, он стремится к стационарному потоку. Тогда P_k — это вероятность того, что в произвольный момент достаточно отдаленного будущего будет равно k вызовов.

Преобразуем **(**)** в соответствии с принятым предположением. Очевидно, что для стационарного режима: $\frac{d}{dt}P_k(t)=\frac{d}{dt}P_k(t)=0$. Тогда:

$0=\lambda_{k-1}P_{k-1}+\beta_{k+1}P_{k+1}-(\lambda_k+\beta_k)P_k$, при $k\geq 1$.

$0=\beta_1 P_1-\lambda_0 P_k$, при $k=0$. **(***)**

$$\sum_{k=0}^{\infty} P_k = 1$$

Эти уравнения известны, как уравнения движения Колмогорова для стационарного режима (уравнения статистического равновесия).

Решим систему **(***)** методом последовательной подстановки:

$$P_1 = \left(\frac{\lambda_0}{\beta_1}\right) \times P_0$$

$k=1$, $P_2 = \left(\frac{\lambda_0 \times \lambda_1}{\beta_1 \times \beta_2}\right) \times P_0$

Можно убедиться в том, что общее решение можно искать в виде:

$$P_k = \left(\frac{\lambda_0 \times \lambda_1 \times \ldots \times \lambda_{k-1}}{\beta_1 \times \beta_2 \times \ldots \times \beta_k}\right) \times P_0$$

$P_k = \prod_{i=0}^{k-1} \frac{\lambda_i}{\beta_{i+2}} \times P_0$, при $k=0,1,\ldots$

Это предположение можно проверить методом математической индукции, подставляя P_k в выражение для P_{k+1}, для $k=0$, получаем: $\prod_{i=0}^{-1} \frac{\lambda_i}{\beta_{i+1}} = 1$.

Справедливость этого была доказана в теории множеств.

Найдем P_i, воспользовавшись нормирующим уравнением:

$$\sum_{k=0}^{\infty} \prod_{i=0}^{k-1} \frac{\lambda_i}{\beta_{i+1}} \times P_0 = 1;$$

следовательно, $P_0 = \dfrac{1}{\sum_{k=0}^{\infty} \prod_{i=0}^{k-1} \frac{\lambda_i}{\beta_{i+1}}}$.

В окончательном виде получим основополагающее равенство классической (элементарной) теории телетрафика, которое является исходной моделью для анализа всех схем.

$$P_k = \frac{\prod_{i=0}^{k-1} \frac{\lambda_i}{\beta_{i+1}}}{\sum_{k=0}^{\infty} \prod_{i=0}^{k-1} \frac{\lambda_i}{\beta_{i+1}}}, \text{ при } k=0,1,2,\ldots$$

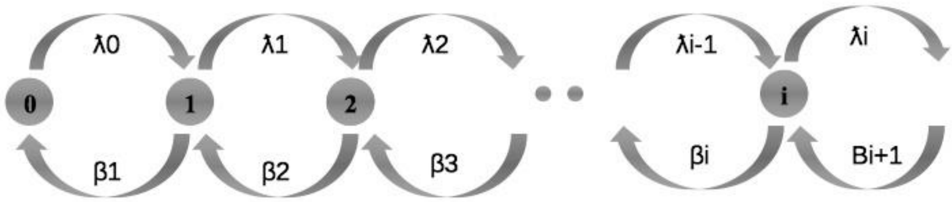
Рассмотрим условие устойчивости для этой формулы: $\frac{\lambda_k}{\beta_{k+1}} < 1$, так как требуется, чтобы P_0 было больше *1*; данное утверждение может быть доказано математически.

Физическая интерпретация этого условия позволяет получить частные математические соотношения, которые характеризуют модели обслуживания ПВ, укладывающиеся в рамки марковских процессов. Эти частные модели могут иметь различные характеристики поступающего потока (простой или примитивный). Но при этом всегда функция распределения промежутков между вызовами *A(x)* должна подчиняться показательному закону.

Это могут быть модели с потерями, или с ожиданием, но длительность обслуживания *B(x)* также должна всегда подчиняться показательному закону.

В 1910 году Эрлангом была выдвинута идея статического равновесия, которая позволила иным путем подойти к схеме **(***)**. Им было предложено описывать движение процесса (размножения и гибели) при помощи диаграмм состояний и переходов. Состояния *k* на диаграмме обозначались овалом (кругом) *(k* – число вызовов в системе*)*. Овалы соединялись ребрами со стрелками. Ребра указывают возможные переходы из одного состояния в другие. Ребрам приписываются числа, характеризующие интенсивности размножения и гибели.

Эрланг предполагал, что стационарное состояние процесса должно удовлетворять, так называемому, условию сохранения.



Для каждого состояния схемы, входящий поток события должен быть равен исходящему. Для состояния *(i)*, входящий поток образуется индексами λ_{i-1} и β_{i-1}, исходящий поток образуется дугами с индексами λ_i и β_i. Тогда интенсивность потока в состоянии *i*: $\lambda_{k-1}P_{k-1}+\beta_{k+1}P_{k+1}$

$\lambda_k P_k+\beta_k P_k=(\lambda_k+\beta_k)P_k$

В условиях статистического равновесия интенсивности равны:

$\lambda_{k-1}P_{k-1}+\beta_{k+1}P_{k+1}=(\lambda_k+\beta_k)P_k$

ЭЛЕМЕНТАРНЫЕ СИСТЕМЫ ТЕОРИИ ТЕЛЕТРАФИКА

СИСТЕМА *M/M/1*

Обслуживание простейшего потока вызовов однолинейным пучком при показательном законе распределения длительности обслуживания и бесконечно-большом числе мест для ожидания.

Данная модель используется для расчета характеристик процессорных устройств в управлении, а также при расчете характеристик коммутирующих каналов.

Исходные условия:

$\lambda_k=\lambda=const$, при $k=0,1,2,...$

$\beta_k=\beta=const$, при $k=1,2,...$ $\quad(\beta_0=0)$

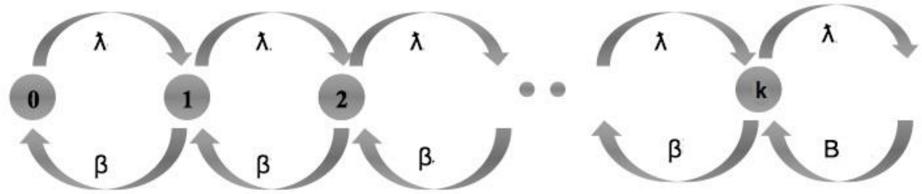
$A(x)=1-e^{-\lambda x}$

$B(x)=1-e^{-\beta x}$

$\frac{\lambda}{\beta}<1$, следовательно, $\lambda<\beta$ — условие устойчивости.

Обслуживание вызовов осуществляется в порядке поступления (FIFO).

Изобразим диаграмму для данной системы:



Определим характеристику P_k, воспользуемся основным равенством теории телетрафика:

$$P_k = \frac{\prod_{i=0}^{k-1} \frac{\lambda}{\beta}}{1 + \sum_{k=1}^{\infty} \prod_{k=1}^{\infty} \frac{\lambda}{\beta}} =$$

Рассмотрим отдельно сумму в знаменателе:

$\sum_{k=1}^{\infty} \left(\frac{\lambda}{\beta} \right)^x$ — это сумма членов бесконечного ряда, ряд сходится, так как $\lambda<\beta$

$$\frac{\frac{\lambda}{\beta}}{1 - \frac{\lambda}{\beta}}$$

Тогда, $= \dfrac{\left(\frac{\lambda}{\beta} \right)^x}{1 + \dfrac{\frac{\lambda}{\beta}}{1 - \frac{\lambda}{\beta}}} = \left(1 - \frac{\lambda}{\beta} \right) \times \left(\frac{\lambda}{\beta} \right)^k$

$P_k = \left(1 - \frac{\lambda}{\beta} \right) \times \left(\frac{\lambda}{\beta} \right)^k$, где k=0,1,2,...

Известно, что интенсивность поступающей нагрузки: $y = \frac{\lambda}{\beta} = \lambda \times \frac{1}{\beta} = \lambda \times t = \mu \times t.$

В рассматриваемом случае $v=1$ (то есть одна линия).

$C = \frac{y}{v}$ (удельная нагрузка на линию).

В данной модели, $C = y = \frac{\lambda}{\beta} = \rho$ — коэффициент использования однолинейной системы.

$\rho \le 1$ — коэффициент использования в системе.

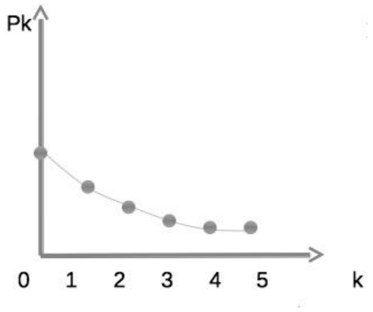
Следовательно, $P_k=(1-\rho)P^k$, где $k=0,1,2,...$

Так как, $P_0=1-\rho$, следовательно, $\rho=1-P_0.$

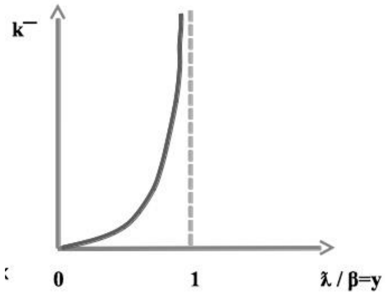
В устойчивых системах $P_0>0$, следовательно, $0 \le \rho < 1$.

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СИСТЕМЫ *M/M/1*

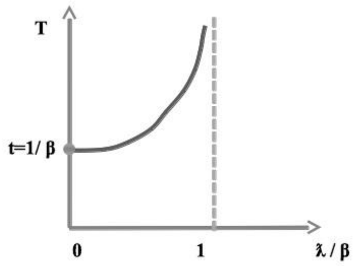
1. Рассмотрим графически функцию $P_k = f\left(k, \frac{\lambda}{\beta}\right)$. Важно, что P_k зависит именно от соотношения $\frac{\lambda}{\beta}$ и не зависит отдельно от каждого из значений λ и β.



2. Среднее число вызовов в системе: $\bar{k} = \sum_{0}^{\infty} k \times P_k = \ldots = \frac{\frac{\lambda}{\beta}}{1-\frac{\lambda}{\beta}}$. Графически, это выражение выглядит следующим образом:



3. Аналогичный характер зависимости будет и для среднего времени пребывания в системе T; применим для получения T формулу Литтла: $T = \frac{k}{\lambda} = \frac{\frac{1}{\beta}}{1-\frac{\lambda}{\beta}}$, при $\bar{k}=\lambda T$.



Если $\frac{\lambda}{\beta}=0$, то $T=\frac{1}{\beta}$. Физический смысл этого состоит в том, что при $\frac{\lambda}{\beta}$, вызовы не ожидают в очереди и длительность обслуживания вызова (t) фактически совпадает со временем его пребывания в системе T. При $\frac{\lambda}{\beta} \to 1$, возрастают $\bar{k}$ и T. Такой вид зависимости характерен почти для всех систем телетрафика. Когда $\frac{\lambda}{\beta}$ близко к единице, характеристики системы становятся неустойчивыми, они резко возрастают. Интуитивно объяснение этого явления состоит в том, что при случайном характере поступающего потока, возникают всплески нагрузки, которые перегружают систему, то есть за использование системы на грани ее пропускной способности, приходится дорого платить.

СИСТЕМА $M/M/\infty$

Обслуживание простейшего потока вызовов бесконечно-линейным пучком при показательном законе распределения длительности обслуживания; без очереди. Данная модель называется моделью с немедленным обслуживанием. Действительно, в ней всегда найдется новый обслуживающий прибор, доступный поступившему вызову.

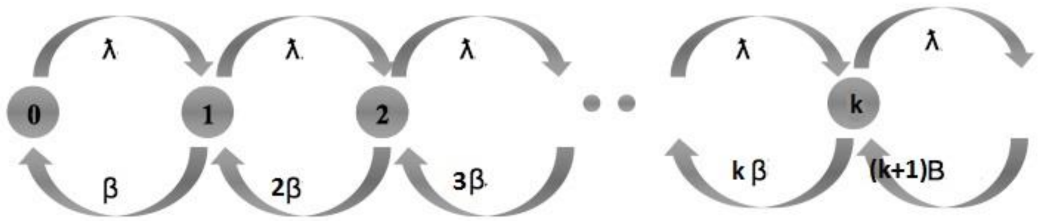
Определим исходные данные для такой системы:

$\lambda_k = \lambda$, где $k = 0,1,2,...$

$\beta_k = k\beta$, где $k = 1,2,...$

Условия устойчивости: $\frac{\lambda_k}{\beta_{k+1}} < 1$, следовательно, $\frac{\lambda}{(k+1)\times\beta} < 1$, следовательно, $\frac{\lambda}{\beta} < \infty$

Диаграмма состояний перехода:



Рассчитаем из основополагающего равенства теории телетрафика

$$P_k = \frac{\prod_{i=0}^{k-1}\frac{\lambda_i}{\beta_{i+1}}}{\sum_{x=0}^{\infty}\prod_{i=0}^{x-1}\frac{\lambda_i}{\beta_{i+1}}} = \frac{\prod_{i=0}^{k-1}\frac{\lambda}{\beta\times(i+1)}}{\sum_{x=0}^{\infty}\prod_{i=0}^{x-1}\frac{\lambda}{(i+1)\times\beta}} = \frac{\frac{\lambda}{1\beta}\times\frac{\lambda}{2\beta}\times\ldots\times\frac{\lambda}{k\times\beta}}{\sum_{x=0}^{\infty}\left(\frac{\lambda}{1\beta}\times\frac{\lambda}{2\beta}\times\ldots\times\frac{\lambda}{x\times\beta}\right)} = \frac{\left(\frac{\lambda}{\beta}\right)^k\times\frac{1}{k!}}{\sum_{x=0}^{\infty}\left(\frac{\lambda}{\beta}\right)^x\times\frac{1}{x!}} = \frac{\frac{1}{k!}\times\left(\frac{\lambda}{\beta}\right)^k}{e^{\frac{\lambda}{\beta}}} = \frac{\left(\frac{\lambda}{\beta}\right)^k\times e^{-\frac{\lambda}{\beta}}}{k!}$$

То есть, $P_k = \frac{\left(\frac{\lambda}{\beta}\right)^k\times e^{-\frac{\lambda}{\beta}}}{k!}$, где $k=0,1,2,\ldots$

Мы получили формулу Пуассона, но здесь это распределение оценивает вероятность того, что бесконечно-линейный пучок в произвольный момент времени находится в состоянии с k занятыми линиями.

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СИСТЕМЫ *M/M/∞*

1) Графическая зависимость рассматривалась в теме простейшего потока вызовов.

2) $\bar{k}=\frac{\lambda}{\beta}$ (берется без вывода).

3) $T=\frac{k}{\lambda}=\frac{1}{\beta}=\bar{t}$.

Действительно, время пребывания вызова в системе T будет равно среднему времени обслуживания $\bar{t}$, так как данная система является системой немедленного обслуживания.

Практическая значимость модели состоит в следующем: модель можно использовать для получения данных о нагрузке, которая поступает в систему. Действительно, вся поступающая нагрузка будет обслуженной, а, следовательно, может оцениваться через характеристики обслуженной нагрузки как среднее число вызовов, находящихся в системе в произвольный момент времени.

СИСТЕМА *M/M/V*

Обслуживание простейшего ПВ v-линейным полнодоступным пучком, при показательном законе распределения длительности обслуживания, с очередью и неограниченном числе мест для ожидания.

Практическим применением данной модели является расчет систем коммутации пакетов и сообщений. Также по системе с ожиданием работают многие службы сервиса в системах с коммутацией каналов.

Итак, имеется пучок линий, где v может меняться от одного до бесконечности. На систему поступает поток с параметром λ, а освобождение, связанное с обслуживанием вызовов, с интенсивностью β. Обслуживание вызовов осуществляется с ожиданием, причем вызовы обслуживаются в порядке их поступления. В системе может быть от 0 до v линий, причем при занятых v линиях на ожидании может быть $0,1,2,\dots,\infty$ вызовов в очереди. Изобразим такую ситуацию на плоскости:

Вызов на обслуживании	0	1	2	...	v
Вызов на ожидании					0
					1
					2
					.
					.
					.
					j
					.
					.
					.
					∞

Суммарное число состояний в такой системе будет бесконечным.

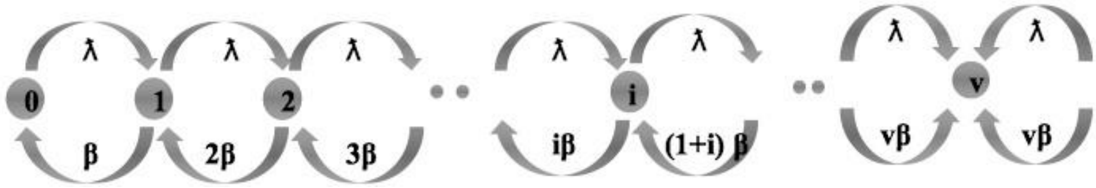
Рассмотрим условие устойчивости для такой системы:

$\frac{\lambda_k}{\beta_{k+1}}$ *<1*

$\frac{\lambda}{v\times\beta}$ <1, следовательно, $\frac{\lambda}{\beta}$ <*v*.

Исследуем данную систему при помощи процесса размножения и гибели, в котором будут следующие исходные условия: λ_k=λ; $\beta_k = \{\begin{matrix} k\times\beta; 0\le k< v \\ v\times\beta; k\ge v\end{matrix}$.

Обозначим, что для верхнего выражения β_k=$k\beta$ (i=k), а для β_k=$v\beta$ (j=k-v).



Ищем решение для P_k (вероятность того, что в системе находится k вызовов). Задачу поиска P_k разобьем на две части, так как зависимость β_k также имеет две части. Вновь используем основополагающую формулу теории телетрафика: $P_k=\prod_{i=0}^{k-1}\frac{\lambda_i}{\beta_{i+1}}\times P_0$, где $P_0=\frac{1}{\sum_{x=0}^{\infty}\prod_{i=0}^{x-1}\frac{\lambda_i}{\beta_{i+1}}}$.

а) $k\le v$, пусть k=i.

$$P_k=\prod_{i=0}^{k-1}\frac{\lambda}{(i+1)\times\beta}\times P_0=\frac{\frac{\lambda}{\beta}}{1}\times\frac{\frac{\lambda}{\beta}}{2}\times\ldots\times\frac{\frac{\lambda}{\beta}}{k}\times P_0=\frac{\left(\frac{\lambda}{\beta}\right)^k}{k!}\times P_0$$

В рассмотренной модели *M/M/1* соотношение $\frac{\lambda}{\beta}$ численно равно интенсивности поступающей нагрузки (y), следовательно, запишем $P_i=\frac{y^i}{i!}\times P_0$, где $i\le v$.

б) $k\ge v$, пусть j=k-v. Разобьем произведение на две части:

$$P_k=\prod_{i=0}^{v-1}\frac{\lambda_i}{\beta_{i+1}}\times\prod_{i=v}^{k-1}\frac{\lambda_i}{\beta_{i+1}}\times P_0=\prod_{i=0}^{v-1}\frac{\lambda}{(i+1)\times\beta}\times\prod_{i=v}^{k-1}\frac{\lambda}{v\times\beta}\times P_0=\frac{\left(\frac{\lambda}{\beta}\right)^v}{v!}\times\frac{\left(\frac{\lambda}{\beta}\right)^{k-v}}{v^{k-v}}\times P_0$$

С учетом введенного индекса j, получим:

$$P_k = W_j = \frac{y^v}{v!} \times \left(\frac{y}{v}\right)^j \times P_0,$$ при k-v=j.

W_j — вероятность того, что в полнодоступном пучке из v-линий, все линии заняты и на ожидании находятся j-вызовов.

Определим вероятность P_0. Будем исходить из нормировочного условия, согласно которому сумма состояний всех вероятностей состояний в системе равна единице,

$$\underbrace{\sum_{i=0}^{v-1} P_i + \sum_{j=0}^{\infty} W_j}_{(*)} = \sum_{i=0}^{v} P_i + \sum_{j=1}^{\infty} W_j = 1$$

Воспользуемся нормировочным условием **(*)**.

$$\sum_{i=0}^{v-1} \frac{y^i}{i!} \times P_0 + \sum_{j=0}^{\infty} \frac{y^v}{v!} \times \left(\frac{y}{v}\right)^j \times P_0 = 1.$$

Следовательно, $P_0 = \dfrac{1}{\sum_{i=0}^{v-1} \frac{y^i}{i!} + \sum_{j=0}^{\infty} \frac{y^v}{v!} \times \left(\frac{y}{v}\right)^j}$.

$\sum_{j=0}^{\infty} \left(\frac{y}{v}\right)^j = \frac{v}{v-y}$, следовательно, $P_0 = \dfrac{1}{\sum_{i=0}^{v-1} \frac{y^i}{i!} + \frac{y^v}{v!} \times \frac{v}{v-y}}$.

В окончательном виде, выражение для P_k принимает следующий вид:

$$P_k = \{ \begin{array}{l} P_i = \dfrac{\frac{y^i}{i!}}{\sum_{i=0}^{v-1} \frac{y^i}{i!} + \frac{y^v}{v!} \times \frac{v}{v-y}}; i \le v \\ W_j = \dfrac{\frac{y^v}{v!} \times \left(\frac{y}{v}\right)^j}{\sum_{i=0}^{v-1} \frac{y^i}{i!} + \frac{y^v}{v!} \times \frac{v}{v-y}}; k - v = j \end{array}$$

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СИСТЕМЫ *M/M/V*

1) Вероятность потерь во времени $P_0 = P\{\gamma > 0\}$.

В системах с ожиданием, потери по времени численно равны вероятности ожидания. Другими словами, это вероятность того, что поступивший вызов не будет немедленно обслужен, а поступит на ожидание, и длительность этого ожидания γ будет больше *0*. Эти потери включают те состояния, когда занято *v*-линий и на ожидании находится *j*-вызовов (где *j=0,1,2,...*).

$$P_t = P\{\gamma > 0\} = \sum_{k=v}^{\infty} P_k = \sum_{j=0}^{\infty} W_j = \sum_{j=0}^{\infty} \frac{\frac{y^v}{v!} \times \left(\frac{y}{v}\right)^j}{\sum_{x=0}^{v-1} \frac{y^x}{x!} + \frac{y^v}{v!} \times \frac{v}{v-y}}$$

Рассмотрим числитель:

$$\sum_{j=0}^{\infty} \frac{y^v}{v!} \times \left(\frac{y}{v}\right)^j = \frac{y^v}{v!} \times \sum_{j=0}^{\infty} \left(\frac{y}{v}\right)^j = \frac{y^v}{v!} \times \frac{v}{v-y}.$$

$$P_t = \frac{\frac{y^v}{v!} \times \frac{v}{v-y}}{\sum_{x=0}^{v-1} \frac{y^x}{x!} + \frac{y^v}{v!} \times \frac{v}{v-y}}$$ — это выражение называют 2-я (вторая) формула Эрланга.

Таким образом, P_t — вероятность того, что поступающий вызов не застанет ни одной свободной линии.

Данная формула табулирована (то есть, есть таблицы).

2) Функция распределения времени ожидания начала обслуживания.

Потери по времени не указывают на характер распределения времени ожидания начала обслуживания, поэтому более общей является характеристика $P\{\gamma > t\}$ — это вероятность того, что вызов, поступающий в произвольный момент времени, попадает на ожидание, и время ожидания будет больше, чем *t*. При этом обычно, и γ и *t*, определятся в относительных единицах, а именно — единицах длительности обслуживания.

Для оценки $P\{\gamma > t\}$ можно воспользоваться формулой полной вероятности (ФПВ). Итоговое выражение берем без вывода:

$P\{\gamma > t\} = P\{\gamma > 0\} e^{-\beta(v-y)t}$. **(*)**

Данное выражение **(*)** также табулировано. Выражение **(*)** справедливо по отношению к любому поступившему вызову, в том числе и такому, которому не придется ждать.

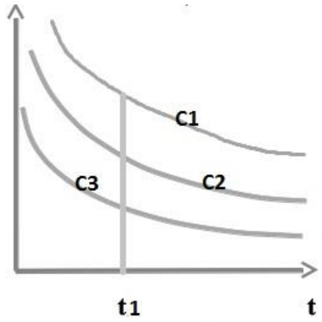
Рассмотрим аналогичную вероятность для вызова, про который известно, что он точно попадает в ожидание. Обозначим эту вероятность $P_з$.

$P_з\{\gamma > t\} = \dfrac{P\{\gamma > t\}}{P\{\gamma > 0\}}$.

В итоге получим:

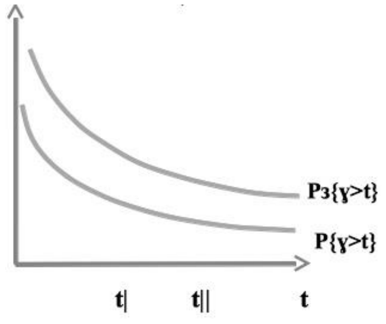
$P_з\{\gamma > t\} = e^{-\beta(v-y)t}$. **(**)**

Рассмотрим графическую интерпретацию формул **(*)** и **(**)**. Изобразим графики сначала в общем виде. За счет экспоненты, они имеют убывающий характер. Пусть $v=const$. Введем величину $c = \dfrac{y}{v}$ (нагрузка на одну линию) и рассмотрим характер этих зависимостей.



Возьмем некоторую точку t_1 и поднимем перпендикуляр до пересечения с этими кривыми. Обозначим c_1, c_2 и c_3. Выше располагается кривая, которая имеет большие потери: $c_1 > c_2 > c_3$. Чем больше c, тем выше доля вызовов, время ожидания начала обслуживания у которых превысит t_1.

Изобразим на другом графике:



При фиксированных значениях v и c, с увеличением t, вероятность того, что длительность ожидания начала обслуживания γ превысит заданную величину t, уменьшается, а кривая подмножества задержанных вызовов располагается выше.

3) $\bar{\gamma}$ — среднее время ожидания начала обслуживания и $\bar{\gamma}_з$ — среднее время

ожидания начала обслуживания для задержанных вызовов.

$\gamma = \int_0^\infty t \frac{d}{dt} P\{\gamma > t\} dt = ...$ — без вывода, через математическое ожидание $\bar{\gamma}$ получаем:

$$\gamma = \frac{P\{\gamma > 0\}}{\beta\left(v - \frac{\lambda}{\beta}\right)},$$

где $\frac{\lambda}{\beta} = y$ — интенсивность обслуженной нагрузки.

$$\gamma_з = \frac{1}{\beta \times \left(v - \frac{\lambda}{\beta}\right)}$$

4) Среднее число вызовов, находящихся на ожидании $\bar{j}$. Здесь речь идет о функции распределения дискретной случайной величины числа вызовов. Поэтому при оценке выражения используется не интеграл, а сумма:

$$j = \sum_{k=v}^{\infty} k \times P_k = \sum_{j=0}^{\infty} j \times W_j ,$$

Где j — число вызовов, находящихся на ожидании; W_j — вероятность того, что в системе есть j вызовов, находящихся на ожидании.

$\mathrm{j} = \frac{\mathrm{P}\{\gamma > 0\}}{\beta\left(v - \frac{\lambda}{\beta}\right)} \times \lambda$ — то же самое выражение можно получить из формулы Литтла.

$j = \bar{\gamma}\lambda$ — частный случай формулы Литтла для очереди.

СИСТЕМА *M/M/V/K*, где $K=V$

Обслуживание простейшего ПВ, v-линейным пучком, при показательном законе распределения длительности обслуживания, без мест для ожидания. Система, которая будет иметь явные потери.

Данная система является системой с потерями, так как вызов, поступивший в момент занятости всех линий в пучке, отвергается системой. Следовательно, такая система может находиться в состояниях с 0 вызовов, 1 вызовом, … , v вызовами (суммарное число состояний конечно и равно $(v+1)$).

Исходные условия. Простейший поток, следовательно: $\lambda_k = \{\begin{matrix} \lambda, k < v \\ 0, k \geq v \end{matrix}$.

Параметр освобождения: $\beta_k = k \times \beta$, при $k = 1,.2,...$ (так как $\beta_0 = 0$).

Условие устойчивости: $\frac{\lambda_k}{\beta_{k+1}} < 1$, следовательно, $\frac{\lambda}{(k+1) \times \beta} < 1$, при $\frac{\lambda}{\beta} < \infty$.

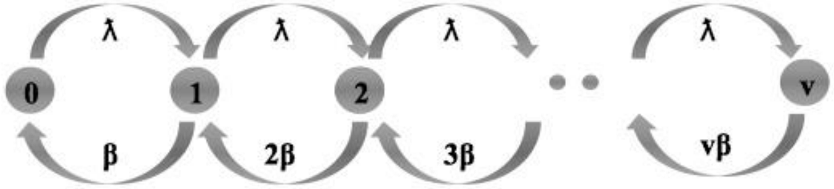


Диаграмма состояний перехода для системы *M/M/V/K*, где *K*=*V*:

Оценим $P_k = \frac{\prod_{i=0}^{k-1} \frac{\lambda_i}{\beta_{i+1}}}{\sum_{x=0}^{\infty} \prod_{i=0}^{k-1} \frac{\lambda_i}{\beta_{i+1}}}$. Подставим в это выражение исходные условия:

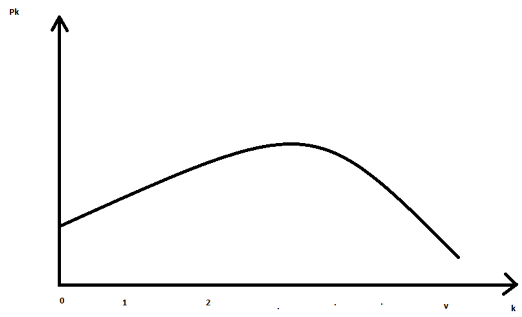
$$P_k = \frac{\prod_{i=0}^{k-1} \frac{\lambda}{(i+1) \times \beta}}{\sum_{x=0}^{v} \prod_{i=0}^{x-1} \frac{\lambda}{(i+1) \times \beta}} = \frac{\frac{\left(\frac{\lambda}{\beta}\right)^k}{k!}}{\sum_{x=0}^{v} \frac{\left(\frac{\lambda}{\beta}\right)^x}{x!}}$$, при $\frac{\lambda}{\beta} = y$. Следовательно:

$P_k = \frac{\frac{y^k}{k!}}{\sum_{x=0}^{v} \frac{y^x}{x!}}$ — это формула носит название 1-ой (первой) формулы Эрланга,

где: P_k — вероятность того, что в полнодоступном пучке из v-линий занято точно k линий, если на этот пучок поступает простейший поток вызовов; при этом длительность обслуживания подчиняется показательному закону, а дисциплина обслуживания в такой системе - с потерями.

ОСНОВНЫЕ ХАРАКТЕРИСИКИ СИСТЕМЫ *M/M/V/K*, где *K*=*V*

1. Исследуем P_k графически.



Таким образом, P_k — это функция распределения дискретной случайной величины k случайных вызовов. Огибающие этой функции близки к огибающим распределения Пуассона.

2. P_t — потери по времени (доля времени, в течение которого вызов, если он поступит, будет потерян). Это время образуется, когда заняты все линии. Для этого приравняем P_t к P_v, следовательно, $P_t = P_v = E_{v,v}(y) = \dfrac{\frac{y^v}{v!}}{\sum_{x=0}^{v} \frac{y^x}{x!}}$, где y — интенсивность поступающей нагрузки.

Потери по времени табулированы в таблицах Пальма и Башарина.

3. $P_в$ — потери по вызовам, которые могут быть оценены через соотношение интенсивности потерянного и поступающего потоков вызовов (так как речь идет о простейшем потоке вызовов, который обладает свойством стационарности).

$P_в = \dfrac{\mu_n}{\mu} = \dfrac{\lambda \times P_v}{\lambda} = P_v$. Следовательно, потери по вызовам численно равны потерям по времени.

Можно доказать, что и другие виды потерь, такие как потери по нагрузке ($P_н$), также равны P_v: $P_н = \dfrac{y_n}{y} = P_v$

Следовательно, $P_t = P_в = P_н = P_v$.

4. Рекуррентные соотношения первой формулы Эрланга. Рассмотрим некоторые

значения P_i и P_{i-1}, где $P_i = \dfrac{\dfrac{y^i}{i!}}{\sum\limits_{x=0}^{v} \dfrac{y^x}{x!}}$, а $P_{i-1} = \dfrac{\dfrac{y^{i-1}}{(i-1)!}}{\sum\limits_{x=0}^{v} \dfrac{y^x}{x!}}$.

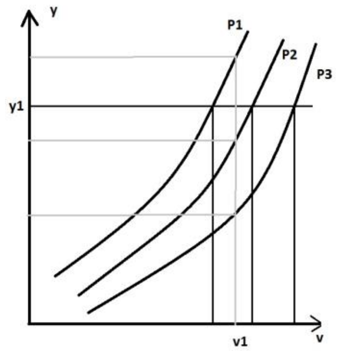
Если $\dfrac{P_i}{P_{i-1}} = \dfrac{y}{i}$, следовательно, можем считать (i-1), через i и наоборот. То есть, $P_i = \dfrac{y}{i} \times P_{i-1}$ и $P_{i-1} = \dfrac{i}{y} \times P_i$.

5. Исследуем первую формулу Эрланга, устремив v к бесконечности.

$$P_k = \frac{\dfrac{y^k}{k!}}{\sum\limits_{x=0}^{\infty} \dfrac{y^x}{x!}} = \frac{\dfrac{y^k}{k!}}{\underbrace{e^y}_{при \sum\limits_{x=0}^{\infty} \frac{y^x}{x!} = e^y}} = \frac{y^k}{k!} \times e^{-y}$$

Таким образом, при $v \to \infty$, первая формула Эрланга переходит в формулу Пуассона. Здесь P_k — вероятность того, что в произвольный момент времени бесконечно-линейный пучок находится в состоянии k.

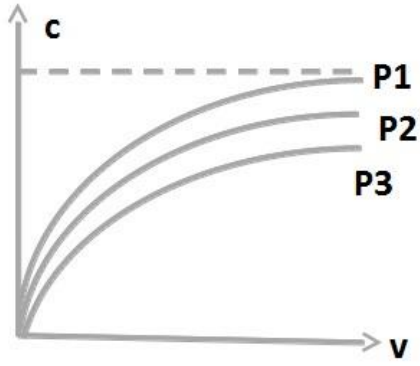
6. Рассмотрим графические зависимости:



Пусть: P — потери; y — интенсивность поступающей нагрузки; v — емкость пучка; $P=const$.

Зафиксируем качество обслуживания P ($P=const$). Очевидно, что чем больше емкость пучка, тем большая нагрузка может быть обслужена этим пучком при заданных потерях. Зафиксируем некоторую нагрузку y_1, проведем горизонтальную прямую и опустим перпендикуляры в точках пересечения с P. Чем больше потери, тем ниже качество обслуживания, следовательно, тем меньшим пучком линий может быть обслужена эта нагрузка. Можно сделать вывод, что $p_1 > p_2 > p_3$.

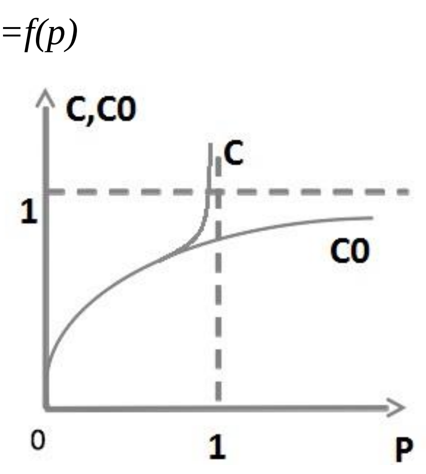
Рассмотрим обратную зависимость. Зафиксируем значение *v*. Чем меньше потери, тем меньшая нагрузка может быть обслужена таким пучком линий.
Рассмотрим еще одну зависимость. Введем удельную нагрузку. Более наглядная взаимосвязь между нагрузкой, числом линий и качеством обслуживания наблюдается при рассмотрении графика зависимости удельной интенсивности нагрузки на линию *c* ($C = \frac{y}{v}$). Рассмотрим *c=f(v)*, при фиксированных потерях для разных значений этих потерь.



Из характера этих графиков видно, что в области малых пучков (обычно не больше *10-15* линий), существует низкое использование линий в пучке. Однако, с ростом *v*, происходит рост удельной пропускной способности. При больших значениях *v* (свыше *100* линий), наступает насыщение, то есть с увеличением *v* пропускная способность линии практически не меняется.
Рассмотрим график, показывающий различия между удельной нагрузкой на линии поступающей С и обслуженной Со в зоне насыщения.

$c_0=f(p)$

$c=f(p)$



В области больших потерь, значения поступающей нагрузки не дает наглядной характеристики качества обслуживания (так как в области больших потерь, основная часть потока не обслуживается, а теряется). Из определения нагрузки, c_0 (удельная нагрузка обслуживания на одну линию), не может быть больше *1*, так как одна линия в единицу времени не может обслужить нагрузку более *1* Эрл.

7. Сопоставим две модели: *M/M/V* и *M/M/V/K*, где $K=V$ (то есть вторая и первая формулы Эрланга).

Сравним вероятности P_i и потери по времени:

$$\underbrace{P_i}_{ожид} = \frac{\frac{y^i}{i!}}{\sum_{x=0}^{v-1} \frac{y^x}{x!} + \frac{y^v}{v!} \times \frac{v}{v-y}}$$ — вторая формула Эрланга

$$\underbrace{P_i}_{потери} = \frac{\frac{y^i}{i!}}{\sum_{x=0}^{v} \frac{y^x}{x!}}$$ — первая формула Эрланга

Сопоставление формул показывает, что вероятность P_i в системе с ожиданием будет меньше, чем P_i в системе с потерями. $P_{i(ожид)}<P_{i(пот)}$.

Действительно, при сравнении формул видно, что знаменатель больше у y_i в системе с ожиданием, следовательно, данная вероятность будет меньше. Этому есть и объяснение следующего рода: система с ожиданием примет бесконечно-большое число состояний, а система с потерями — только одно $v+1$ состояние. Так как сумма всех вероятностей равна единице для обеих моделей, то вероятность каждого состояния в системе с ожиданием будет меньше.

Сравним P_t в системе с ожиданием:

$$\underbrace{P_t}_{ожид} = \frac{\frac{y^v}{v!} \times \frac{v}{v-y}}{\sum_{x=0}^{v-1} \frac{y^x}{x!} + \frac{y^v}{v!} \times \frac{v}{v-y}}$$

$$\underbrace{P_t}_{потери} = \frac{\frac{y^v}{v!}}{\sum_{x=0}^{v} \frac{y^x}{x!}}$$

Без вывода принимаем, что $P_{t(ожид)}>P_{t(пот)}$.

При заданной величине P_t, не только поступающая, но и обслуженная нагрузка в системах с потерями будет больше, чем в системах с ожиданием.

СИСТЕМА *M/M/V/K/N*, где *K*=*V*

Обслуживание примитивного потока вызовов полнодоступным пучком линий при показательном законе распределения длительности обслуживания без мест для ожидания и при ограниченном числе источников.

В данной модели принципиально новым является то, что входящий поток вызовов больше не является простейшим, а является потоком, который создается конечным числом источников. Для такого потока, параметр $\lambda_k=\alpha(N-k)$, где $N-k$ — число свободных источников; α — коэффициент пропорциональности (численно равен параметру одного свободного источника).

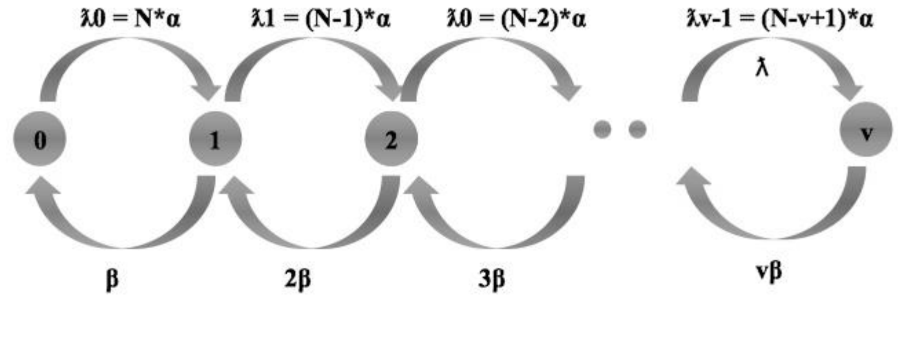
Исходные условия:

$$\lambda_k = \begin{cases} \alpha\times(N-k); k<v, k=i \\ 0; k\ge v \end{cases}$$

$$\beta_k = k\times\beta; k=1,2,...,v$$

Условие устойчивости: $\dfrac{\lambda_k}{\beta_{k+1}}=\dfrac{\alpha\times(N-k)}{(k+1)\times\beta}<1$. Следовательно, $\dfrac{\alpha\times N}{\beta}<\infty$.

Изобразим диаграмму состояний перехода:



Определим P_k из основополагающего равенства теории телетрафика:

$$P_k=\frac{\prod_{i=0}^{k-1}\frac{\lambda_i}{\beta_{i+1}}}{\sum_{x=0}^{\infty}\prod_{i=0}^{x-1}\frac{\lambda_i}{\beta_{i+1}}}=\frac{\prod_{i=0}^{k-1}\frac{\alpha\times(N-i)}{(i+1)\times\beta}}{\sum_{k=0}^{v}\prod_{i=0}^{x-1}\frac{\alpha\times(N-i)}{(i+1)\times\beta}}=\frac{\left(\frac{\alpha}{\beta}\right)^k\times\frac{N\times(N-1)\times...\times(N-k+1)}{1\times2\times...\times k}}{\sum_{x=0}^{v}\left(\frac{\alpha}{\beta}\right)^x\times\frac{N\times(N-1)\times...\times(N-x+1)}{x!}}=$$

Воспользуемся выражением для определения числа сочетаний:

$$C_N^k = \frac{N!}{k! \times (N-k)!} = \frac{N \times (N-1) \times ... \times (N-k+1) \times (N-k)!}{k! \times (N-k)!} = \frac{N \times (N-1) \times ... \times (N-k+1)}{k!}$$

Следовательно, $= \dfrac{C_N^k \left(\dfrac{\alpha}{\beta}\right)^k}{\sum\limits_{x=0}^{v} C_N^x \left(\dfrac{\alpha}{\beta}\right)^x}$.

Окончательно получаем, при $k=i$, $k<v$, следующее выражение:

$$P_i = \frac{C_N^i \left(\dfrac{\alpha}{\beta}\right)^i}{\sum\limits_{x=0}^{v} C_N^x \left(\dfrac{\alpha}{\beta}\right)^x}$$ — это формула Энгсета.

Распределение Энгсета представляет собой функцию от множества параметров: $P_i=f(N,i,\alpha,\beta,\delta)$. Данная формула является более общей, чем первая формула Эрланга. Если N устремить к бесконечности и одновременно параметр α устремить к нулю, так что произведение $N\,\alpha = const = \lambda$, где λ — параметр потока вызовов всех свободных источников; в этих условиях, формула Энгсета переходит в первую формулу Эрланга.

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СИСТЕМЫ *M/M/V/K/N*, где *K=V*

1. Рассмотрим взаимосвязь между параметром потока α (альфа) и нагрузкой, которая поступает от одного свободного источника, которую обозначим a (а). Для этого, рассмотрим систему без потерь, то есть такую, в которой число линий v равно числу источников v. В такой системе каждый источник может обслуживаться независимо от других источников. Для рассмотрения взаимосвязи между α и a, рассмотрим такой случай системы, когда $v=N=1$.

$$P_1 = \frac{C_1^1 \left(\dfrac{\alpha}{\beta}\right)^1}{\sum\limits_{x=0}^{1} C_1^x \left(\dfrac{\alpha}{\beta}\right)^x} = \frac{\dfrac{\alpha}{\beta}}{1+\dfrac{\alpha}{\beta}}.$$

Здесь, P_1 — доля времени, в течение которого источник занят. Это численно соответствует нагрузке, поступающей от одного источника. Отсюда следует, что $P_1 = a$. Следовательно, $a = \dfrac{\frac{\alpha}{\beta}}{1 + \frac{\alpha}{\beta}}$. Следовательно, $\dfrac{\alpha}{\beta} = \dfrac{a}{1-a}$.

Эта связь позволяет представить формулу Энгсета в другом виде:

$$P_i = \frac{C_N^i \left(\frac{a}{1-a} \right)^i}{\sum_{x=0}^{v} C_N^x \left(\frac{a}{1-a} \right)^x}.$$

2. Рассмотрим вероятности P_i в системе без потерь (если число источников совпадает с числом обслуживающих приборов). Для такой системы:

$$P_i = \frac{C_N^i \left(\frac{a}{1-a} \right)^i}{\sum_{x=0}^{v} C_N^x \left(\frac{a}{1-a} \right)^x} =$$

Домножим числитель и знаменатель на $(1-a)^N$. Получим:

$$= \frac{C_N^i a^i (1-a)^{N-1}}{\sum_{x=0}^{v} C_N^x a^x (1-a)^{N-x}} =$$

Рассмотрим знаменатель выражения:

$\sum_{x=0}^{v} C_N^x a^x (1-a)^{N-x} = [a + (1-a)]^x = 1$ (Сумма членов бинома Ньютона).

Следовательно: $= C_N^i a^i (1-a)^{N-i}$.

$P_i = C_N^i a^i (1-a)^{N-i}$ — формула Бернулли.

Здесь, P_i — вероятность того, что в произвольный момент времени, из общего числа N источников, занято i источников.

3. $P_i = P_v = \dfrac{C_N^v \left(\frac{a}{1-a} \right)^v}{\sum_{x=0}^{v} C_N^x \left(\frac{a}{1-a} \right)^x}$. **(*)**

4. Потери по вызовам: $P_в = \dfrac{\mu_п}{\mu}$, где μ — интенсивность потока.

$$P_в = \frac{\mu_n}{\mu} = \frac{C_{N-1}^{v}\left(\frac{a}{1-a}\right)^{v}}{\sum_{x=0}^{v} C_{N-1}^{x}\left(\frac{a}{1-a}\right)^{x}} \qquad (**)$$

Из соотношения формул **(*)** и **(**)** следует, что в полнодоступном пучке из *v* линий, на который поступает поток вызовов, и каждый источник создает одну и ту же нагрузку *a*, $P_в$ при наличии *N* источников, численно равны P_t, при наличии (*N-1*) источника.

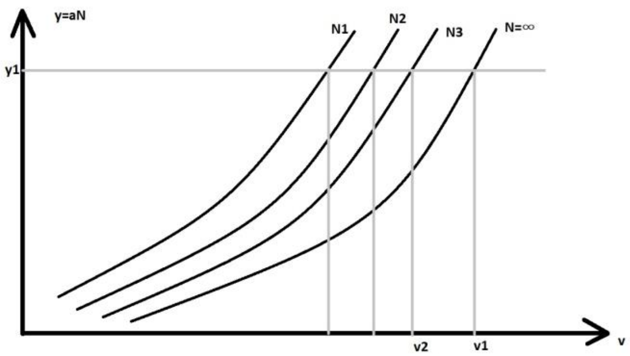
$P_в(N,a,v) = P_t((N-1),a,v)$, следовательно, $P_в < P_t$,

$P_в(N,a,v) < P_t(N,a,v)$.

Потери по вызовам табулированы. Чтобы по таблицам определить P_t, достаточно взять табличное значение $P_в((N+1),a,v)$.

5. Сравним системы *M/M/V/K*, где *K=V* и *M/M/V/K/N*, где *K=V*.

Первая формула Эрланга используется для расчета числа обслуживающих приборов в системах с потерями, при незначительном числе источников вызовов. Формула Энгсета более подходит при анализе систем с ограниченном числом источников вызовов. Выполним графическое исследование:



Из рисунка видно, что пропускная способность пучка *y* при заданных потерях возрастает с ростом емкости этого пучка. Она также возрастает и с ростом потерь (см. рассуждения для простейшего потока вызовов, для системы *M/M/V/K*, где *K=V*). Также на пропускную способность *y* влияет число источников вызовов *N*, причем, с уменьшением *N*, пропускная способность пучка растет. Следовательно, $N_1<N_2<N_3$.

Рассмотрим следующий пример. Пусть необходимо определить ресурс системы, число линий *v*, если известно, что эта система должна обслуживать заданную нагрузку y_1. Рассмотрим две кривые N_3 и $N=\infty$. Возможны два варианта решения задачи:

а) применить наиболее распространенную модель обслуживания с $N=\infty$, тогда для обслуживания нагрузки требуется v_1 линий;

б) выбрать формулу Энгсета, которая учитывает конечность числа источников N_3, и, в этом случае, нам достаточно иметь пучок из v_2 линий. $v_2<v_1$, следовательно, пример показывает, что неточный выбор математической модели может привести к излишнему расходу ресурсов системы.

СИСТЕМЫ ТЕЛЕТРАФИКА БОЛЕЕ ОБЩЕГО ВИДА

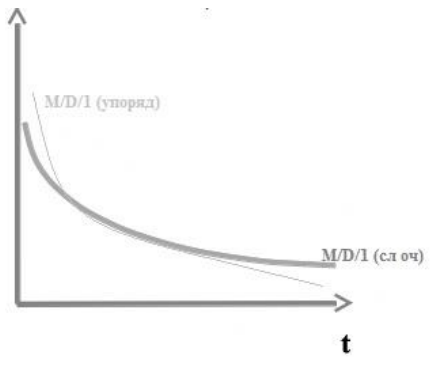
Если система в своем описании выходит за рамки марковского процесса, тогда ее относят к системам телетрафика более общего вида.

СИСТЕМА *M/D/1* СО СЛУЧАЙНОЙ ОЧЕРЕДЬЮ

Обслуживание простейшего ПВ однолинейным пучком при постоянном времени обслуживания.

Исходные условия: $\lambda k=const$

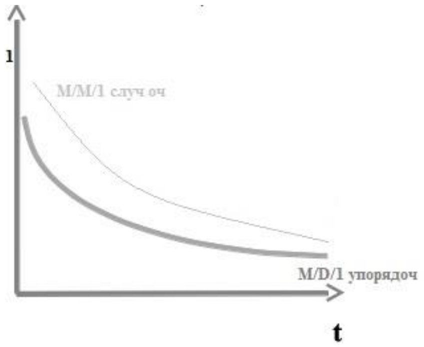
Такая система исследована Берком. В силу сложности математического выражения для практического использования зависимости данной системы выражаются графически. Сравним их с результатом той же системы, но при обслуживании в порядке очереди.



При малых значениях *t*, *M/D/1* со случайной очередью лучше, так как $P\{\gamma>t\}$ меньше. Именно в области малых значений *t*, работа такой системы представляет наибольший интерес.

При больших значениях *t*, характеристики *M/D/1* со случайной очередью значительно хуже, однако если доля таких вызовов мала, то это практически не ухудшает качество обслуживания вызовов.

Выбор из очереди в случайном порядке при малых значениях *t* так же как и выбор в порядке очереди, в целом, обеспечивает более высокое качество обслуживания при постоянной длительности обслуживания по сравнению со случайным выбором для показательного распределения.



СИСТЕМА *M/D/V* С УПОРЯДОЧЕННОЙ ОЧЕРЕДЬЮ

Система с *v*-линейным пучком, при постоянной длительности обслуживания, без ограничения числа мест для ожидания.

Исходные условия: $\lambda \leq v \leq \infty$

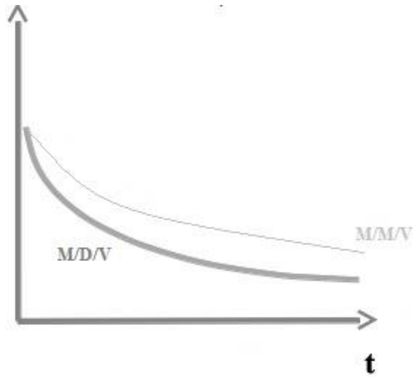
Пусть емкость пуска конечна и на такой поток поступает пучок с параметром λ.

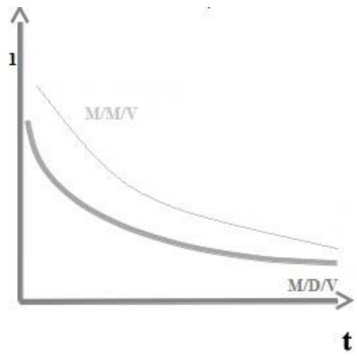
Длительность обслуживания *h* примем за единицу времени.

Условием устойчивости такой системы может служить неравенство $\lambda<u$. Определим вероятность $P\{\gamma>t\}$, где *t* — дополнительное время ожидания.

$T=\frac{t_{доп}}{h}$ — данная задача была решена Кроммелиным, который определил математические зависимости $P\{\gamma>t\}$ как функции от *t*, то есть $P\{\gamma>t\}=f(t)$, при $v=const$; $c=\frac{y}{u}=const$ (нагрузка на одну линию).

Известно, что всякий выход за рамки марковских процессов приводит к значительному усложнению результатов, поэтому результаты Кроммелина были представлены графически, в диапазоне *v =1-20* и *c=(0,002-0,8)* Эрл.
Характер зависимости *P{γ>t}* аналогичен зависимости, полученной во второй формуле Эрланга, однако количественно характеристики различаются.





С ростом *v* характеристики качества обслуживания улучшаются. Данные графические зависимости показывают, что для улучшения качества обслуживания при фиксированном значении *v* лучше стандартизировать длительность обслуживания при прочих равных условиях.

ОБОСНОВАНИЕ ОБЛАСТИ ПРИМЕНЕНИЯ СИСТЕМ С ОЖИДАНИЕМ И С ПОТЕРЯМИ

Самая общая классификация систем РИ предполагает два вида коммутаций:

а) система с коммутацией каналов (КК);

б) система с коммутацией пакетов (КП).

Система с КК может работать как с ожиданием, так и с потерями. Система с КП является системой с ожиданием.

Рассмотрим систему с коммутацией каналов, в которой можно выделить две основные группы устройств:

1) устройства, образующие тракт передачи информации;

2) управляющие устройства.

Указанные группы устройств существенно различаются по закону распределения длительности занятия, количеству обслуживаемых приборов, дисциплине обслуживания, и так далее.

Рассмотрим приборы первой группы.

Обычно, для них можно считать, что длительность занятия распределяется по показательному закону со средним значением от единиц до сотен секунд (например, для среднего времени разговора $t_{разг}=70\text{-}80$ с). Кроме того, обычно используются достаточно большие емкости пучков ($v=10\text{-}100$), а нормы потерь составляют порядка *2-3%*. В этой области потерь сопоставление первой и второй формул Эрланга показывает, что системы с потерями обладают большей пропускной способностью, чем системы с ожиданием при прочих равных условиях, следовательно, в этой группе выгоднее использовать системы с потерями.

Вторая группа приборов (с ожиданием, коммутация пакетов).

Управляющие устройства характеризуются длительностью занятия, близкой к постоянной, со средним значением этой длительности на 2-3 порядка меньше, по сравнению с приборами первой группы. Емкость приборов (процессоров) обычно $v \le 5$. Постоянное время обслуживания и малое значение этой длительности обслуживания позволяет устанавливать большие вероятности: $P\{\gamma>t\}$, при $t=0$. Однако, с увеличением t, $P\{\gamma>t\}$ уменьшается, и, если системы строить с большими значениями t, то при этом мы будем удерживаться в диапазоне малых $P\{\gamma>t\}$, следовательно, управляющие устройства целесообразно строить по системе с ожиданием.

ПЕРСПЕКТИВЫ РАЗВИТИЯ ТЕОРИИ ТЕЛЕТРАФИКА

В связи с развитием различных инфокоммуникационных систем и сетей происходит и дальнейшее развитие моделей и методов теории телетрафика.

Основные направления такого развития связаны со следующими задачами:

1) разработка и проектирование многопроцессорных комплексов систем и

сетей;

2) автоматизация управления технологическими процессами;

3) автоматизация организационного управления;

4) автоматизация разработки информационных технологий, и так далее.

В связи с этим, для теории телетрафика можно выдвинуть следующие группы проблем:

1. Научно-исследовательская группа проблем.

2. Проблемы проектирования.

3. Проблемы эксплуатации и развития.

Научно-исследовательские проблемы связаны с тем, что теория телетрафика является инструментом, обеспечивающим создание систем с совершенными пропорциями, которая подразумевает:

а) наличие внутренней структуры, отвечающей требованиям работоспособности и пропускной способности;

б) соответствие внутренней структуры условиям внешней среды. Теория телетрафика помогает правильно устанавливать характер поступающего потока вызовов, учитывать колебание нагрузки, повторные вызовы, вид передаваемой информации, и так далее;

в) соответствие систем современной электронной базе. Это подразумевает правильное распределение оборудования между ресурсами коммутации и ресурсами управления.

Итак, роль теории телетрафика в разработке структур и систем сводится к решению технико-экономических задач, которые включают в себя некоторые неравенства, то есть условия, смысл которых состоит в обеспечении необходимого качества обслуживания. При этом под качеством обслуживания могут пониматься различные функционалы от параметров системы и внешней среды (потери, задержки и другие).

Проблемы проектирования рассматривают в узком и широком смыслах. В узком смысле, задачи проектирования включают в себя комплектование и привязку оборудования к конкретным условиям. В широком смысле, проблема проектирования связана с выбором единого критерия эффективности системы. Проблемы эксплуатации и развития связаны с тем, что необходимо обеспечить управление по различным направлениям:

а) управление параметрами работы системы;

б) управление устранением неисправностей;

в) управление конфигурацией;

г) управление расчетами;

д) управление защитой от несанкционированного доступа, и так далее.

Все эти задачи решаются параллельно с проблемами измерения и прогнозирования параметров системы для создания новых высокоэффективных сетей и систем связи.